

Centre Armand-Frappier Santé Biotechnologie

CARACTÉRISATION DE LA NUCLÉASE SSNA : UN RÉGULATEUR DES TRANSFERTS HORIZONTALS DE GÈNES ET DÉTERMINANT DE LA VIRULENCE DE *NEISSERIA MENINGITIDIS*

Par
Martin Chenal

Thèse présentée pour l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Biologie

Jury d'évaluation

Président du jury et examineur interne	Jonathan Perreault INRS-Centre Armand-Frappier Santé-Biotechnologie
Examineur externe	Sylvain Moineau Département de biochimie, microbiologie et bio- informatique Université Laval
Examineur externe	Emmanuelle Bille INSERM U1151, CNRS UMR 8253, Institut Necker Enfants-Malades, Paris
Directeur de recherche	Frédéric Veyrier INRS-Centre Armand-Frappier Santé-Biotechnologie

REMERCIEMENTS

En premier lieu, je tiens à remercier mon directeur de thèse, Frédéric Veyrier, pour son implication et son support tout au long de ce projet. Sans son éternel optimisme qui m'a poussé à persévérer malgré les nombreux obstacles, ce projet de recherche n'aurait jamais atteint le point où il en est. Je lui suis également reconnaissant de m'avoir permis de tracer ma propre voie lors de cette longue étape qu'est le doctorat, en me laissant libre de m'impliquer auprès de projets qui me tenaient à cœur.

Je dois également un immense merci à ma partenaire de cœur et de vie, Chloé, qui me supporte depuis plus de cinq ans dans les bons et les mauvais moments. Traverser ensemble cette épreuve qu'est le doctorat a contribué de façon décisive à ma réussite, tant sur l'aspect scientifique que psychologique.

Évidemment, merci à ma famille d'avoir fait de moi qui je suis et de m'avoir épaulé du début à la fin de ce long parcours académique.

Merci à mes collègues de toujours, Ève, Cécilia et Sammy, qui m'ont accompagné tout au long du doctorat et m'ont poussé à être le meilleur de moi-même. À Alex également, qui a contribué au dernier chapitre de cette thèse.

Merci également aux amis et collègues, trop nombreux pour être nommés, dont le support moral et scientifique a été d'une précieuse aide.

Un grand merci aussi à nos collaborateurs ainsi qu'aux membres de mon jury, qui ont su m'aiguiller ces dernières années.

Enfin, merci à l'INRS et aux organismes subventionnaires dont les Fonds de recherche du Québec en santé, d'avoir permis l'accomplissement de ce projet de recherche.

RÉSUMÉ

Les bactéries du genre *Neisseria* sont un exemple remarquable d'évolution et d'adaptation. Composées principalement d'espèces commensales, seules deux espèces sont pathogènes chez l'humain : *N. meningitidis* et *N. gonorrhoeae*. La première est responsable d'infections systémiques souvent mortelles lorsque non traitées, tandis que la seconde cause une infection transmise sexuellement très prévalente au Canada. Plusieurs attributs contribuent à l'évolution rapide et complexe des *Neisseria*. Leur génome est une véritable mosaïque d'éléments répétés, dont beaucoup sont encore méconnus. De plus, ces bactéries sont naturellement compétentes, signifiant qu'elles ont la capacité d'acquérir de l'ADN de leur environnement par transformation, mais la régulation de ce mécanisme est mal comprise.

Cette étude s'intéresse à une petite protéine hypothétique des *Neisseria*, dont plusieurs caractéristiques lui suggèrent des rôles biologiques et évolutifs importants. Nous avons nommé celle-ci SsnA pour *Specific single-strand nuclease*. Son gène est entouré de nombreuses copies d'une séquence répétée aux fonctions jusque-là inconnues, nommées NTS. Dans le but de percer les secrets de ces deux éléments intrigants, nous avons procédé à leur caractérisation exhaustive en utilisant des approches complémentaires incluant des études bio-informatiques, biochimiques, phénotypiques en plus de modèles animaux. Nos résultats démontrent que SsnA est une endonucléase d'ADN simple-brin ciblant spécifiquement les séquences répétées NTS, présentes en centaines de copies dans le génome de *N. meningitidis*. Grâce à cette activité enzymatique unique, SsnA régule l'intégration d'ADN exogène par transformation, un des principaux mécanismes de transfert horizontal de gènes chez les *Neisseria*. Notre étude révèle également que cette nucléase est un déterminant majeur de la virulence de *N. meningitidis*. Enfin, nos analyses suggèrent que SsnA est le premier membre caractérisé d'une vaste famille de nucléases aux spécificités diverses, dont les fonctions biologiques restent à définir.

En conclusion, cette étude marque plusieurs découvertes majeures en recherche fondamentale, mais aussi dans le domaine biotechnologique. En effet, en plus de mettre en lumière un nouveau déterminant de la virulence de *N. meningitidis* et un médiateur important des transferts horizontaux de gènes, cette étude révèle une toute nouvelle famille d'enzymes clivant spécifiquement l'ADN simple-brin, une activité enzymatique particulièrement rare qui pourrait être exploitée en ingénierie génétique.

Mots-clés : Microbiologie; Bactériologie; Biologie moléculaire; *Neisseria*; Nucléase; Virulence; Compétence naturelle; Évolution; Bio-informatique; Biotechnologie

ABSTRACT

Bacteria from the *Neisseria* genus are a remarkable example of host adaptation and intricate evolution. Mainly composed of commensal species, only two species are human pathogens: *N. meningitidis* and *N. gonorrhoeae*. The former is responsible for deadly systemic infections, while the latter causes a sexually transmissible disease highly prevalent in Canada. Several attributes have contributed to the fast-paced evolution of *Neisseria*. Their genome is a true mosaic of repeated elements, many of which are still cryptic. Moreover, these species are naturally competent, meaning that they can acquire environmental DNA through transformation, although the regulation of this mechanism is poorly understood.

This fundamental study focuses on a small hypothetical protein from the *Neisseria* genus, of which several characteristics suggest important biological and evolutionary roles. We named this protein SsnA for Specific single-strand nuclease. Its gene is surrounded by numerous copies of a repeated sequence with unknown functions, named NTS. With the aim of unlocking the secrets of these two elements, we proceeded to comprehensively characterize them using complementary approaches, including bioinformatics, biochemistry, phenotypic studies, and animal models. Our results demonstrate that SsnA is a single-strand DNA endonuclease specifically targeting the NTS repeated sequences, found in hundreds of copies in the *N. meningitidis* genome. Due to this unique enzymatic activity, SsnA regulates the integration of foreign DNA by transformation, which is one of the main means of horizontal gene transfers in *Neisseria*. Surprisingly, our study also reveals that this nuclease is a major virulence determinant of *N. meningitidis*. Finally, our analyses suggest that SsnA is the first characterized member of a vast nuclease family with diverse specificities, the biological functions of which are yet to be defined.

In conclusion, this study marks several important discoveries for the fundamental research of *Neisseria*, as well as the biotechnology field. Indeed, in addition to shedding light on a novel virulence determinant of *N. meningitidis* and an important horizontal gene transfer mediator, this study reveals a broad family of nucleases that specifically cut single-stranded DNA, a particularly rare enzymatic activity that could be exploited for genetic engineering.

Keywords : Microbiology; Bacteriology; Molecular biology; *Neisseria*; Nuclease; Virulence; Natural competence; Evolution; Bioinformatics; Biotechnology

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ.....	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES FIGURES	XV
LISTE DES TABLEAUX.....	XIX
LISTE DES SIGLES ET ABRÉVIATIONS	XXI
REVUE DE LITTÉRATURE.....	1
1 NUCLEASES.....	1
1.1 Classification et grandes familles de nucléases	1
1.1.1 Exonucléases	2
1.1.2 Endonucléases.....	2
1.2 Fonctions biologiques chez les procaryotes	5
1.2.1 Réparation d'ADN	5
1.2.2 Réplication d'ADN	6
1.2.3 Recombinaison d'ADN	6
1.2.4 Mécanismes de défense	6
1.2.5 Régulation post-transcriptionnelle	7
1.2.6 Mobilité génétique	7
1.2.7 Apport métabolique	8
1.3 Applications biotechnologiques et thérapeutiques	8
1.3.1 DNases et RNases non spécifiques.....	8
1.3.2 Enzymes de restriction	8
1.3.3 Nucléases programmables.....	9
1.4 Superfamille des nucléases GIY-YIG	11
1.4.1 UvrC-like.....	12
1.4.2 Endonucléases de type « homing » (HEases)	13
1.4.3 Penelope-like Elements (PLEs).....	13
1.4.4 COG3680	13
1.4.5 COG1833	14

1.4.6	<i>Famille Slx1</i>	14
1.4.7	<i>Enzymes de restriction (REases)</i>	14
1.4.8	<i>MSH1</i>	15
1.4.9	<i>Nucléases GIY-YIG non classifiées</i>	15
2	BACTÉRIES DU GENRE NEISSERIA	17
2.1	Généralités.....	17
2.1.1	<i>Importance en santé publique</i>	17
2.1.2	<i>Informations génomiques</i>	20
2.1.3	<i>Diversité écologique, morphologique et pathologique</i>	21
2.2	Évolution et adaptation.....	22
2.2.1	<i>Émergence des Neisseria pathogènes</i>	22
2.2.2	<i>Facteurs de virulence</i>	24
2.2.3	<i>Mécanismes de variabilité génétique</i>	26
2.3	Compétence naturelle.....	27
2.3.1	<i>Origines de l'ADN transformant</i>	28
2.3.2	<i>Import d'ADN</i>	30
2.3.3	<i>Traitement de l'ADN et intégration au génome</i>	31
2.3.4	<i>Fonctions biologiques et rôles dans l'évolution</i>	33
2.4	Séquences répétées des <i>Neisseria</i>	34
2.4.1	<i>DNA uptake sequence (DUS)</i>	36
2.4.2	<i>dRS3/NIME</i>	37
2.4.3	<i>Répétitions Correia (CR) et Éléments Correia (CREE/NEMIS)</i>	38
2.4.4	<i>Répétitions AT-riche (ATR)</i>	38
2.4.5	<i>Séquences répétées non uniques aux Neisseria</i>	39
	PROBLÉMATIQUE, HYPOTHÈSES ET OBJECTIFS	41
1	CHAPITRE I – ÉTUDE BIO-INFORMATIQUE	45
	MATÉRIEL ET MÉTHODES.....	45
1.1	Analyses des séquences répétées.....	45
1.1.1	<i>Quantification des séquences répétées</i>	45
1.1.2	<i>Représentations graphiques</i>	46
1.2	Génomés analysés et phylogénie des <i>Neisseriaceae</i>	47

1.3	Analyses des homologues de SsnA	48
	RÉSULTATS ET DISCUSSION	49
1.4	Identification d'une séquence répétée analogue aux dRS3	49
1.5	Distribution des séquences répétées chez les <i>Neisseriaceae</i>	51
1.6	Identification d'un gène associé aux NTS (<i>ssnA</i>)	55
1.7	Association <i>ssnA</i> -NTS.....	56
1.7.1	<i>Association ssnA-NTS chez les Neisseria</i>	56
1.7.2	<i>Identification d'une variante de la NTS chez certaines Neisseria</i>	58
1.7.3	<i>SsnA chez les Neisseria</i>	59
1.7.4	<i>Association ssnA-NTS chez les non-Neisseria</i>	61
1.8	SsnA est une nucléase hypothétique de la superfamille GIY-YIG	63
	CONCLUSIONS	66
2	CHAPITRE II – CARACTÉRISATION BIOCHIMIQUE DE SSNA	67
	MATÉRIEL ET MÉTHODES	67
2.1	Souches de <i>E. coli</i> et conditions de culture.....	67
2.2	Clonage, expression et purification de SsnA.....	68
2.3	Interactions protéines-ADN	71
2.3.1	<i>Oligonucléotides fluorescents</i>	71
2.4	Tests de liaison à l'ADN	73
2.4.1	<i>Retards sur gel (EMSA)</i>	73
2.4.2	<i>Thermophorèse</i>	73
2.5	Tests de nucléase	74
2.5.1	<i>Sur gels d'agarose</i>	74
2.5.2	<i>Sur gels d'acrylamide</i>	74
	RÉSULTATS ET DISCUSSION	77
2.6	Purifications de protéines	77
2.7	Activité enzymatique.....	78
2.7.1	<i>Détermination du substrat de SsnA</i>	78

2.7.2	<i>Conditions optimales de clivage</i>	83
2.7.3	<i>Cinétiques de clivage</i>	85
2.7.4	<i>Spécificité et site de clivage</i>	86
2.7.5	<i>Clivage d'autres copies de la NTS</i>	90
2.7.6	<i>Affinité (K_D)</i>	92
CONCLUSIONS		95
3	CHAPITRE III - CARACTÉRISATION PHÉNOTYPIQUE DE SSNA	96
MATÉRIEL ET MÉTHODES		96
3.1	Souches de <i>Neisseria</i> et conditions de culture.....	96
3.2	Courbes de croissance et transcriptomique	99
3.3	Tests de transformation	100
3.3.1	<i>Transformations avec de l'ADN purifié</i>	100
3.3.2	<i>Transformations par cocultures</i>	102
3.4	Variation antigénique de <i>pilE</i>	103
3.5	Tests de virulence	105
3.5.1	<i>Infections in vitro</i>	105
3.5.2	<i>Infections de souris</i>	106
3.6	Tests statistiques.....	107
RÉSULTATS ET DISCUSSION		108
3.7	Génération et validation des mutants	108
3.8	Étude transcriptomique.....	109
3.9	Rôle dans la transformation par compétence naturelle	111
3.9.1	<i>Transformation d'ADN sans NTS dans les zones de recombinaison</i>	112
3.9.2	<i>Transformation d'ADN avec NTS dans les zones de recombinaison</i>	114
3.9.3	<i>Transformations de plasmides avec des NTS mutées</i>	116
3.9.4	<i>Autres tests de transformation</i>	117
3.9.5	<i>Transformations par co-culture</i>	120
3.10	Impact sur la variation antigénique de <i>PilE</i>	122
3.10.1	<i>Variation antigénique en absence d'ADN transformant</i>	124

3.10.2	<i>Variation antigénique en présence d'ADN transformant</i>	126
3.11	Implications dans la virulence de <i>N. meningitidis</i>	127
3.11.1	<i>Infections in vitro</i>	128
3.11.2	<i>Infection systémique de souris</i>	131
3.12	Autres phénotypes testés	137
	CONCLUSIONS	138
4	CHAPITRE IV – FAMILLE DES SSN	139
	MATÉRIEL ET MÉTHODES	139
4.1	Identification et classification des homologues de SsnA	139
4.2	Recherche d'éléments répétés associés aux homologues de SsnA	140
4.3	Purifications de protéines	140
4.4	Tests de nucléase	141
	RÉSULTATS ET DISCUSSION	143
4.5	Protéines Ssn dans le domaine bactérien	143
4.6	Association des Ssn avec des séquences similaires aux NTS.....	147
4.7	Caractérisation enzymatique d'homologues distants de SsnA.....	150
4.7.1	<i>Expression et purifications de protéines</i>	150
4.7.2	<i>Tests d'activité nucléase</i>	151
	CONCLUSIONS	155
5	DISCUSSION GÉNÉRALE	157
5.1	Importance de l'interaction SsnA-NTS chez les <i>Neisseria</i>	157
5.1.1	<i>NTS : Nouvelle séquence répétée des Neisseria</i>	157
5.1.2	<i>Médiation des transferts horizontaux de gènes par SsnA</i>	158
5.1.3	<i>Nouveau mécanisme d'interférence de la compétence naturelle</i>	162
5.1.4	<i>Antagonisme des séquences répétées chez les Neisseria</i>	163
5.1.5	<i>Paradigme des dRS3/NTS</i>	164
5.2	Importance de SsnA dans la virulence de <i>N. meningitidis</i>	166
5.3	Nouvelle famille de protéines GIY-YIG	167

5.3.1	<i>Activité enzymatique unique de SsnA</i>	169
5.3.2	<i>Spécificités attendues</i>	171
5.3.3	<i>Nucléases, ou plutôt transposases domestiquées ?</i>	172
5.4	<i>Perspectives futures</i>	174
5.4.1	<i>Interactome de SsnA</i>	174
5.4.2	<i>Études structurales de SsnA</i>	175
5.4.3	<i>Importance de SsnA dans la virulence du méningocoque</i>	175
5.4.4	<i>Potentiel biotechnologique des Ssn</i>	177
CONCLUSION		179
BIBLIOGRAPHIE		183
ANNEXE I – MODÉLISATIONS <i>IN SILICO</i>		201
ANNEXE II – BREVET		203

LISTE DES FIGURES

Figure 1 Classifications des nucléases	1
Figure 2 Principales familles de nucléases utilisées en ingénierie génétique	11
Figure 3 Incidence de la gonorrhée au Canada	18
Figure 4 Phylogénie du genre <i>Neisseria</i>	23
Figure 5 Origine et évolution des <i>Neisseria</i> commensales et pathogènes	24
Figure 6 Schéma simplifié de transformation par compétence naturelle.....	28
Figure 7 Don et import d'ADN par les <i>Neisseria</i>	29
Figure 8 Mécanisme d'import d'ADN par compétence naturelle chez les <i>Neisseria</i>	31
Figure 9 Traitement et intégration chromosomique de l'ADN transformant.....	32
Figure 10 Distribution des clusters de dRS3 chez <i>N. meningitidis</i>	41
Figure 11 Association des dRS3/NTS avec un gène non caractérisé.....	42
Figure 12 La dRS3 est plus longue et moins conservée qu'initialement décrite.....	49
Figure 13 Différences entre les dRS3 et les NTS	50
Figure 14 Cartographie des séquences répétées chez les <i>Neisseria</i>	52
Figure 15 Distribution et abondance des éléments répétés chez les <i>Neisseria</i>	53
Figure 16 Analyse des motifs répétés à proximité de <i>ssnA</i> chez les <i>Neisseria</i>	57
Figure 17 Analyse de nouveaux motifs répétés à proximité de <i>ssnA</i>	58
Figure 18 Comparaison entre la NTS et la NTS _{var}	59
Figure 19 Synténie des régions codant pour <i>ssnA</i> chez les <i>Neisseria</i>	60
Figure 20 Alignement multiple des protéines <i>SsnA</i> chez les <i>Neisseria</i>	61
Figure 21 <i>SsnA</i> appartient à la superfamille des nucléases GIY-YIG	64
Figure 22 Hiérarchie de la superfamille GIY-YIG selon la base de données CDD	65
Figure 23 Purification de <i>SsnA</i> sauvage et mutée.....	78

Figure 24 Tests d'activité nucléase résolus sur gel d'agarose	79
Figure 25 Tests d'endonucléase sur de l'ADN double-brin en gel de polyacrylamide	80
Figure 26 Activité catalytique de SsnA sur de l'ADN simple-brin	82
Figure 27 Optimisation des conditions de clivage de SsnA	84
Figure 28 Détermination de la constante catalytique de SsnA.....	86
Figure 29 Détermination du site de clivage et de la taille minimale du substrat	87
Figure 30 Cartographie de la spécificité de séquence de SsnA.....	90
Figure 31 Interaction de SsnA avec d'autres NTS et modèle proposé	91
Figure 32 Détermination de l'affinité de SsnA pour son substrat.....	93
Figure 33 Validation des mutants de SsnA chez N. meningitidis 8013.....	108
Figure 34 Analyse transcriptomique de mutants de N. meningitidis	110
Figure 35 Tests de transformation de Nm avec des substrats sans NTS	113
Figure 36 Transformation de Nm avec des plasmides contenant la NTS.....	115
Figure 37 Transformation de plasmides avec des NTS intactes et mutées	117
Figure 38 Transformations d'autres substrats avec NTS	119
Figure 39 Transformation de Nm par coculture avec une souche donneuse	121
Figure 40 Méthodologie des tests de variation antigénique sans ADN transformant.....	123
Figure 41 Variation antigénique de Nm en absence d'ADN transformant.....	125
Figure 42 Variation antigénique de Nm en présence d'ADN transformant	127
Figure 43 Adhésion et invasion de macrophages THP-1.....	129
Figure 44 Adhésion et invasion de cellules épithéliales Calu-3	130
Figure 45 Infections systémiques de souris avec Nm LNP20553 lux	133
Figure 46 Infections systémiques de souris avec Nm LNP24198 lux_Km	135
Figure 47 Analyse CLANS de la famille cd10448	144
Figure 48 Phylogénie de la famille cd10448	146
Figure 49 Homologues de SsnA associés à des éléments répétés similaires aux NTS ..	148

Figure 50 Homologues de ssnA associés à d'autres répétitions palindromiques	149
Figure 51 Purifications de protéines homologues à SsnA	151
Figure 52 Activité nucléase des homologues SsnA sur leurs REP respectives.....	153
Figure 53 Spécificité de clivage des homologues de SsnA.....	153
Figure 54 Modèle de transformation par compétence naturelle.....	159
Figure 55 Mécanisme proposé de restriction d'ADN transformant par SsnA.....	161
Figure 56 Hypervariabilité des régions contenant des dRS3 chez <i>N. meningitidis</i>.....	164
Figure 57 Nouvelle addition à la superfamille des nucléases GIY-YIG.....	169
Figure 58 Mécanisme de repliement du site de clivage de TnpA	170
Figure 59 Structure des séquences d'insertion et des locus ssnA	172

LISTE DES TABLEAUX

Tableau 1 Familles de nucléases GIY-YIG	12
Tableau 2 Facteurs impliqués dans la virulence des Neisseria	25
Tableau 3 Éléments répétés chez les Neisseria.....	35
Tableau 4 Détection de NTS à proximité des homologues ssnA	62
Tableau 5 Souches d'expression utilisées pour les purifications de protéine	67
Tableau 6 Concentration en antibiotiques des milieux de culture.....	68
Tableau 7 Liste des plasmides d'expression utilisés dans cette étude	68
Tableau 8 Oligonucléotides fluorescents	72
Tableau 9 Constantes enzymatiques de protéines agissant sur l'ADNsb.....	94
Tableau 10 Souches de Neisseria utilisées	97
Tableau 11 Liste des plasmides pour Neisseria utilisés dans cette étude	101
Tableau 12 Tests phénotypiques sans effets observés liés à SsnA.....	137
Tableau 13 Liste des oligonucléotides fluorescents	142

LISTE DES SIGLES ET ABRÉVIATIONS

aa	Acides aminés
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire, obtenu par rétrotranscription d'ARNm
ADNdb	ADN double-brin
ADNg	ADN génomique
ADNsb	ADN simple-brin
Ago	Argonaute. Famille de nucléases.
AP	Site abasite (apurinique/apyrimidique)
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNt	ARN de transfert
ATP	Adénosine triphosphate
ATR	Répétitions AT-riches
BER	Réparation par excision de base
Cas	Protéine associée aux CRISPR
CDC	Centre de prévention et de contrôle des infections
CDD	Base de données des domaines conservés
CFU	Unité formant une colonie
CLANS	Cluster ANalysis of Sequences
Cm	Chloramphénicol
Compl	Complémentation
CPS	Comptes par seconde (luminescence)
CR/CREE	Répétition Correia / Élément répété Correia
CRISPR	Courtes répétitions en palindrome regroupées et régulièrement espacées
Cy5	Cyanine 5 (fluorophore)
DNase	Nucléase ciblant l'ADN
DO600/DO595	Densité optique à 600 ou 595 nm
dRS3	Séquence répétée directe 3
DUS	Séquence d'import d'ADN
EDTA	Acide éthylènediaminetétraacétique
EMS	Methanesulfonate d'éthyle
EMSA	Retard sur gel
Ery	Erythromycine
FCB	Fibrobacteria, Chlorobiota, Bacteroidota. Superphylum bactérien
Genta	Gentamycine
GGI	Ilôt génétique du gonocoque
GST	Glutathion S-transférase
HEase	Endonucléase de homing
His/6x-His	Étiquette Histidine/polyhistidine
HRR	Réparation par recombinaison homologue
IHF	Facteur d'intégration de l'hôte
IS	Séquence d'insertion

ITS	Infection transmise sexuellement
Km	Kanamycine
KO	Knock-out, délétion
LOS	Lipooligosaccharide
Mg	Magnésium
MGE	Élément génétique mobile
MITE	Élément transposable espacé en mosaïque
MMR	Réparation de mésappariement
Mn	Manganèse
MOI	Multiplicité de l'infection
Nal	Acide nalidixique
NER	Réparation par excision de nucléotide
NHEJ	Jonction d'extrémités non homologues
NIME	Élément espacé en mosaïque des <i>Neisseria</i>
Nm	<i>Neisseria meningitidis</i>
ns	Non significatif
nt	Nucléotide
NUMOD	Domaine de liaison à l'ADN modulaire associé à une nucléase
OE	Surexpression
Oligo	Oligonucléotide
OMS	Organisation mondiale de la santé
OMV	Vésicule de la membrane externe
PAGE/SDS-PAGE	Électrophorèse sur gel de polyacrylamide / En présence de dodécylsulfate de sodium
PAM	Motif adjacent à un protoespaceur
pb	Paire de bases
PBS	Tampon phosphate salin
PCA	Analyse par composante principale
PCR/RT-qPCR	Réaction en chaîne par polymérase / couplée à une transcription inverse
PFU	Unité de formation de plaque (virus)
PI	Post-infection
PLE	Élément similaire aux Penelope
PMT	Tubes photomultiplicateurs
PRR	Répétition pentatricopeptide
RAYT	Transposase tyrosine associée à un REP
REP	Palindrome répétitif extragénique
RM/R-M	Restriction-Modification
RNase	Nucléase ciblant l'ARN
rpm	Rotations par minute
RPMI	Milieu de culture cellulaire (milieu de l'institut Roswell Park Memorial)
RT	Rétrotranscription ou transcriptase inverse
SAM	S-adenosyl methionine
Sm	Streptomycine
Sp	Spectinomycine
ssDNase	Nucléase ciblant l'ADN simple-brin

SSM	Mésappariement par glissement de brin
SVF	Sérum de veau foetal
T4SS	Système de sécrétion de type 4
TALE/TALEN	Nucléases effectrices de type activateur de transcription
TAP-MS	Spectrométrie de masse couplée à la purification d'affinité en tandem
TBE	Tris-Borate-EDTA
tBOOH	tertbutyle d'hydroperoxyde
Tet	Tétracycline
TFP	Pili de type 4
USS	Séquence signal d'import
UV	Ultraviolets
VIH	Virus de l'immunodéficience humain
WT	Souche ou protéine sauvage
Y2H	Double hybride de levure
ZFN	Nucléase à doigt de zinc

REVUE DE LITTÉRATURE

1 Nucléases

Les nucléases représentent une famille extrêmement vaste d'enzymes capables de cliver ou de dégrader des acides nucléiques en hydrolysant les liaisons phosphodiester qui relient les nucléotides entre eux. Ces protéines, retrouvées dans tous les domaines du vivant, sont impliquées dans une panoplie de processus biologiques comme la régulation du génome, le métabolisme ou encore la réplication virale (Yang, 2011). L'origine des nucléases est intimement liée à leur interaction avec l'ADN et l'ARN. La coévolution longue et complexe de ces deux éléments a donné naissance à des mécanismes plus ou moins sophistiqués qui reflètent bien la diversité des organismes qui les produisent. Du fait de leur diversité, leurs fonctions biologiques et leur potentiel industriel, ces protéines sont largement étudiées en génétique, médecine ou encore biotechnologie, tant en recherche fondamentale qu'en recherche appliquée et clinique.

1.1 Classification et grandes familles de nucléases

L'énorme diversité des nucléases rend leur classification complexe. En effet, ces enzymes peuvent être divisées selon plusieurs critères non mutuellement exclusifs comme leur origine, leur spécificité de substrat, leur mécanisme de clivage ou encore leurs fonctions biologiques (**Figure 1**). Il existe cependant deux grandes classes de nucléases, les exonucléases et les endonucléases, elles-mêmes divisées en désoxyribonucléases et ribonucléases selon leur spécificité pour de l'ADN ou de l'ARN.

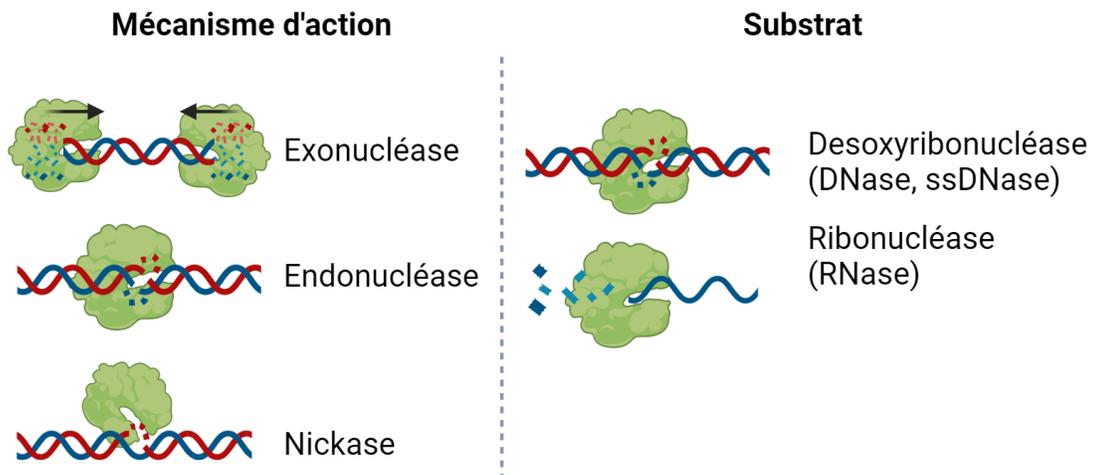


Figure 1 Classifications des nucléases
Figure personnelle créée avec BioRender.com

1.1.1 Exonucléases

Les exonucléases ont comme particularité de cliver des acides nucléiques à partir de leurs extrémités. En effet, ces enzymes peuvent hydrolyser successivement les liaisons phosphodiester entre les sucres formant une chaîne d'ADN ou d'ARN, relâchant au passage des mononucléotides. Les exonucléases se divisent en deux sous-familles selon la polarité de leur activité catalytique : 5' à 3' ou 3' à 5'.

La dégradation de l'ADN est un processus essentiel au maintien de l'intégrité génomique. En ce sens, les exonucléases sont particulièrement importantes dans de nombreux mécanismes de réparation d'ADN, où elles permettent d'éliminer les régions endommagées qui pourront ensuite être réparées par d'autres enzymes (Nishino & Morikawa, 2002). Les exonucléases interagissent avec un élément à la base du matériel génétique, les sucres, et n'ont généralement pas de spécificité de séquence. Elles interviennent régulièrement dans des mécanismes complexes où une autre enzyme crée d'abord une cassure spécifique, générant ainsi l'extrémité nécessaire à leur activité catalytique (Manils *et al.*, 2022). Une enzyme peut également posséder des domaines d'interaction spécifiques avec l'ADN en plus d'un domaine exonucléase.

Bien que les domaines fonctionnels d'exonucléases n'aient pas de spécificité de séquence, ces enzymes peuvent avoir une affinité particulière pour de l'ADN double-brin (ADNdb), simple-brin (ADNsb), ou encore de l'ARN auquel cas il s'agit d'exoribonucléases. Ces dernières contrôlent l'expression de gènes via la dégradation d'ARN messager (ARNm) et la maturation d'ARN ribosomiaux (ARNr) et d'ARN de transfert (ARNt) (Deutscher & Li, 2000).

1.1.2 Endonucléases

Contrairement aux exonucléases qui dégradent les acides nucléiques à partir des extrémités, les endonucléases peuvent cliver à l'intérieur de l'ADN ou de l'ARN. Ces enzymes ont un large éventail de spécificités et peuvent reconnaître des séquences précises, des mésappariements de nucléotides, différents types de dommages dans l'ADN ou même diverses structures secondaires indépendamment de leur séquence. Les grandes classes d'endonucléases sont décrites ci-dessous, mais il est à noter qu'elles ne sont pas mutuellement exclusives. Il existe également plusieurs protéines appelées endo-exonucléases qui peuvent à la fois couper aux extrémités et à l'intérieur de leur substrat.

Enzymes de restriction

Les endonucléases ou enzymes de restriction représentent une très vaste famille de protéines retrouvées chez les procaryotes qui ciblent spécifiquement de courtes séquences d'ADN (4-8 pb). Ces enzymes tirent leur nom de leur fonction de protection de l'hôte via un mécanisme de restriction d'ADN étranger (Bertani & Weigle, 1953; Luria & Human, 1952). Le génome de l'hôte est immunisé contre ses propres enzymes de restriction via l'action de méthyltransférases qui méthylent ses propres sites de restriction, les rendant ainsi inaccessibles aux nucléases. Des milliers d'enzymes de restriction ont été découvertes à ce jour, arborant plusieurs centaines de séquences de reconnaissance différentes (Roberts *et al.*, 2015). Ces enzymes sont subdivisées en quatre types et davantage de sous-types selon leur besoin en cofacteur, leur spécificité et position de clivage, ainsi que leurs domaines fonctionnels. Les enzymes de type I forment de gros complexes protéiques composés de plusieurs sous-unités. Ces dernières clivent l'ADN à distance de leur site de reconnaissance et nécessitent du magnésium pour leur activité catalytique, mais aussi de l'adénosine triphosphate (ATP) et du S-adenosyl méthionine (SAM) (Loenen *et al.*, 2014). Puisque leur site de clivage n'est pas clairement défini, ces enzymes ont peu d'utilité en biotechnologie. À l'inverse, les enzymes de restriction de type II ont énormément d'applications en raison de leur spécificité de clivage bien définie dans ou à proximité de leur site de reconnaissance. Ces enzymes n'ont besoin que de magnésium et agissent majoritairement sous forme d'homodimères (Di Felice *et al.*, 2019). Les enzymes de type III quant à elles nécessitent deux copies inversées de leur séquence de reconnaissance pour cliver à distance de ceux-ci en présence de magnésium et d'ATP (Rao *et al.*, 2014). Enfin, les enzymes de type IV se distinguent par leur reconnaissance de modifications d'ADN comme la méthylation, plutôt qu'une spécificité de séquence (Loenen & Raleigh, 2014).

Endonucléases de type « homing »

Les endonucléases de type « homing » ont la particularité de cibler de façon hautement spécifique une séquence assez longue d'ADN (>12 pb), retrouvée en une seule ou très peu de copies dans leur génome d'origine. Habituellement, les gènes codant pour ces nucléases interrompent leur propre séquence de reconnaissance, ce qui leur permet de s'insérer uniquement dans des sites où il n'y a pas encore le gène par un mécanisme de copier-coller, d'où l'expression « homing » (guidage/ciblage). Contrairement aux enzymes de restriction qui servent de mécanismes de défense, les endonucléases de homing sont souvent retrouvées dans des éléments génétiques mobiles (MGEs) et parasitent leur hôte sans autre fonction que leur propre mobilité. Cependant, certaines de ces nucléases ont évolué avec leur hôte de façon à accomplir

des fonctions biologiques importantes, un processus appelé « domestication » (Hafez & Hausner, 2012). Ces nucléases se divisent en six familles selon leurs motifs structuraux ; les LAGLIDADG, H-N-H, His-Cys box, PD-(D/E)xK, Vsr-like/EDxHD, et les GIY-YIG. Cette dernière famille, qui ne se limite pas à des endonucléases de homing, sera détaillée plus loin.

Nickases

Les nickases sont très similaires aux enzymes de restriction en termes de spécificité de séquence et de mécanisme de clivage. Cependant, elles se distinguent par leur capacité à ne couper qu'un brin d'ADN dans leur site de reconnaissance, qui lui est double-brin. Techniquement, plusieurs enzymes de restriction et endonucléases de homing sont en fait des nickases qui clivent séquentiellement un brin d'ADN à la fois, ou encore qui s'homodimérisent de façon à ce que chaque monomère clive un brin (Chan *et al.*, 2011). En raison de leurs nombreuses utilités biotechnologiques et de leur faible nombre découvert à ce jour, des efforts importants sont déployés pour créer des nickases artificielles à partir d'autres endonucléases. En effet, ces dernières sont particulièrement utiles à la mutagenèse dirigée, le marquage d'ADN, et les technologies d'amplification isothermique (Xu, 2015).

Endonucléases guidées

Cette famille regroupe les endonucléases nécessitant une séquence précise d'ADN ou d'ARN pour guider leur spécificité de clivage. On y retrouve notamment les nucléases associées aux CRISPR (Cas), rendues célèbres pour leurs importantes applications biotechnologiques, mais aussi les protéines Argonautes (Ago). Toutes deux servent de mécanismes de défense contre l'invasion d'acides nucléiques exogènes. Les Cas utilisent un guide d'ARN pour médier une immunité dite adaptative chez les procaryotes, tandis que les Ago utilisent un guide d'ADN pour médier une immunité innée. À la différence des précédentes catégories d'enzymes, les nucléases guidées n'ont pas besoin d'être modifiées systématiquement pour changer leur spécificité. Seule la séquence guide doit être adaptée, d'où leur énorme versatilité (Koonin, 2017).

Nucléases à spécificité de structure

Ces enzymes, comme leur nom l'indique, reconnaissent et clivent des structures secondaires d'ADN ou d'ARN plutôt que des séquences précises. Ces nucléases sont particulièrement importantes dans les mécanismes de réparation, de réplication, et de recombinaison d'ADN, où ces structures se forment temporairement à la suite de l'action de diverses protéines (Dehe & Gaillard, 2017). Des défauts dans ces enzymes sont à l'origine de plusieurs maladies génétiques

graves, soulignant leur importance dans des processus biologiques essentiels (Minocherhomji & Hickson, 2014).

DNases et RNases non-spécifiques

Contrairement aux précédentes familles d'endonucléases, les désoxyribonucléases (DNases) et ribonucléases (RNases) non-spécifiques dégradent l'ADN et l'ARN de façon non ciblée en hydrolysant les liaisons phosphodiester reliant les acides nucléiques. Cette dégradation est importante pour la gestion des déchets cellulaires ainsi que l'apoptose et la régulation des transcrits d'ARN (Evans & Aguilera, 2003; Kishi *et al.*, 2001; Saramago *et al.*, 2014).

1.2 Fonctions biologiques chez les procaryotes

Étant donné la diversité énorme des nucléases telle que décrite dans la section précédente, il n'est pas étonnant qu'une panoplie de fonctions biologiques leur soient associées. Ces principales fonctions sont décrites ci-dessous.

1.2.1 Réparation d'ADN

Des nucléases sont impliquées dans les cinq principaux mécanismes de réparation d'ADN, soit la réparation par excision de base (BER), la réparation par excision de nucléotide (NER), la réparation des mésappariements (MMR), la réparation par jonction d'extrémités non homologues (NHEJ) ainsi que la réparation par recombinaison homologue (HRR). Lorsqu'un nucléotide précis est endommagé, la machinerie cellulaire l'enlève par la voie BER, générant ce qu'on appelle un site abasique (AP site). Les endonucléases AP ciblent spécifiquement ces lésions pour ouvrir la voie à d'autres enzymes de réparation (Mol *et al.*, 2000). Lorsqu'une lésion cause une déformation de l'hélice d'ADN, la voie NER peut être activée, auquel cas deux nucléases, UvrB et UvrC, clivent de part et d'autre de la lésion pour la retirer (Nishino & Morikawa, 2002). Lorsqu'une base est mal appariée, la voie MMR est activée, permettant à la nickase MutH de cliver uniquement le brin nouvellement synthétisé, puis à des exonucléases d'exciser l'erreur (Marinus, 2012). Les cassures double-brin peuvent être réparées par les voies NHEJ ou HRR. La voie NHEJ utilise la nickase Artemis qui reconnaît des structures secondaires intermédiaires lors de la réparation (Lieber, 2010). Lors de la voie HRR, d'autres types de structure secondaires comme le D-loop et la jonction Holliday sont spécifiquement clivées par des nucléases pour permettre la recombinaison homologue entre un fragment d'ADN intact et celui endommagé (Nishino & Morikawa, 2002).

1.2.2 Réplication d'ADN

Les polymérase d'ADN affichent un taux d'erreur assez important qui se caractérise par le mésappariement ponctuel de nucléotides. Afin de corriger ces erreurs, des exonucléases accompagnent les polymérase afin d'immédiatement retirer toute base mal appariée (Shevelev & Hübscher, 2002). D'autres nucléases sont essentielles à la dégradation des fragments d'Okazaki lors de la réplication du brin tardif (lagging strand) (Nishino & Morikawa, 2002). Enfin, pour permettre la séparation des deux brins d'ADN lors de la réplication ou la transcription, des nickases spécialisées appelées topoisomérase clivent temporairement un brin, relâchant ainsi la tension créée par le surenroulement.

1.2.3 Recombinaison d'ADN

La recombinaison homologue est un processus important à plusieurs mécanismes de réparation d'ADN et de transferts horizontaux de gènes via l'appariement de séquences similaires, en plus d'être impliquée dans des réarrangements chromosomiques tels que des délétions ou duplications. Elle est notamment essentielle à l'intégration d'ADN exogène par compétence naturelle ainsi qu'à la variation antigénique chez les *Neisseria*, deux mécanismes qui seront abordés plus en détail dans cet ouvrage. Composée de trois étapes, soit la présynapse, la synapse et la post-synapse, la recombinaison homologue fait intervenir diverses nucléases spécifiques. Tout d'abord, le complexe RecBCD ou RecFOR génère des extrémités simple-brin sur l'ADN donneur grâce à leur activité exonucléase et hélicase. Ces complexes recrutent ensuite la recombinase RecA sur ce brin, qui va l'associer à de l'ADN double-brin homologue par un procédé appelé invasion de brin, générant par la même occasion une structure d'ADN en hétéroduplex appelée jonction de Holliday (Mullany, 2010). Des endonucléases structure-spécifiques appelées résolvas reconnaissent enfin ces jonctions et les clivent pour permettre l'échange de brin (Wyatt & West, 2014).

1.2.4 Mécanismes de défense

Pour se protéger des bactériophages, les procaryotes ont développé plusieurs systèmes de défense faisant appel à diverses nucléases. Bien que ces organismes n'aient pas de systèmes immunitaires à proprement parler, plusieurs similarités frappantes apparentent ces derniers à une immunité innée et même adaptative. Les enzymes de restriction par exemple, aidées par des méthylases qui immunisent leur propre génome, reconnaissent et clivent tout ADN étranger qui ne possède pas les mêmes modifications, à la manière d'une immunité innée (Koonin *et al.*,

2017). De leur côté, les nucléases Argonautes (Ago) et les protéines associées aux CRISPR (Cas) se rapprochent plus d'une immunité adaptative (Koonin *et al.*, 2017). Les Argonautes utilisent de courtes séquences guides d'ADN ou d'ARN pour cliver des séquences complémentaires retrouvées dans l'ADN viral ou dans des éléments génétiques mobiles (Willkomm *et al.*, 2018). De leur côté, les systèmes CRISPR-Cas fragmentent et incorporent l'ADN intrus dans leur génome, permettant ensuite aux nucléases Cas d'utiliser les transcrits correspondant pour cliver ces mêmes séquences lors d'une prochaine infection (Barrangou, 2015).

1.2.5 Régulation post-transcriptionnelle

Plusieurs nucléases dont les argonautes sont essentielles au mécanisme d'interférence d'ARN, qui mène à la dégradation spécifique de certains ARN messagers et donc à une régulation post-transcriptionnelle (Wilson & Doudna, 2013). De nombreuses RNases sont également impliquées dans le processus de dégradation de l'ARN mais aussi dans la maturation des ARN de transfert et des ARN ribosomiaux. À titre d'exemple, les précurseurs de l'ARN 16S, 23S et 5S sont transcrits en un bloc et nécessitent l'action de la RNase III pour les dissocier, puis de plusieurs autres ribonucléases pour modifier leurs extrémités et les amener à maturité (Deutscher, 2006).

1.2.6 Mobilité génétique

Bien que camouflées sous divers noms comme recombinases, intégrases ou encore transposases, les éléments génétiques mobiles (MGE) encodent habituellement des nucléases qui leur permettent de s'insérer ailleurs par un mécanisme de copier-coller ou de couper-coller. Les MGE regroupent une variété de plasmides, d'éléments transposables et de prophages. Ces éléments peuvent être bénéfiques pour la cellule hôte d'un point de vue évolutif, comme certains plasmides ou des îlots génétiques contenant des gènes de virulence ou de résistance aux antibiotiques. Ils contribuent également à la diversité et à la variabilité génétique. Cependant, les MGE peuvent aussi être délétères, en s'insérant dans des gènes importants ou en introduisant des gènes causant des défauts ou des maladies (Singh *et al.*, 2014). Les MGE ont une importance particulière chez les *Neisseria*. À titre d'exemples, le génome de *N. gonorrhoeae* abrite un îlot génétique appelé *Gonococcal Genetic Island* encodant un système de sécrétion de type IV, tandis que certaines souches de *N. meningitidis* possèdent l'îlot *Meningococcal Disease Associated*, associé à un meilleur potentiel d'invasion (Bille *et al.*, 2005).

1.2.7 Apport métabolique

Certaines bactéries et archées sont capables d'importer de l'ADN extracellulaire dans leur cytoplasme par un mécanisme appelé compétence naturelle. Bien que cet ADN puisse servir à la réparation du génome et à l'acquisition de gènes, plusieurs études démontrent que l'ADN exogène peut également servir comme source de nutriments, tout particulièrement d'azote, de carbone et de phosphore (Ibanez de Aldecoa *et al.*, 2017; Vorkapic *et al.*, 2016). Diverses DNases seraient impliquées dans ce mécanisme de recyclage d'acides nucléiques en nutriments et en énergie (Palchevskiy & Finkel, 2009). De plus, les exonucléases, qui digèrent les acides nucléiques en mononucléotides, permettraient la réutilisation de ces nucléotides afin de réduire la nécessité d'en synthétiser *de novo* (Wakamatsu *et al.*, 2011).

1.3 Applications biotechnologiques et thérapeutiques

La grande diversité des nucléases a permis le développement de nombreuses applications, tant dans le domaine de la recherche fondamentale que dans des procédés industriels et des perspectives thérapeutiques. Plusieurs usages concrets sont détaillés ci-dessous.

1.3.1 DNases et RNases non spécifiques

Les nucléases non spécifiques sont couramment utilisées pour la purification d'acides nucléiques et de protéines, et servent à fragmenter l'ADN lors de différentes méthodes de séquençage et d'analyses d'interactions ADN-protéines. D'un point de vue clinique, plusieurs traitements à base de DNase I sont actuellement sur le marché ou en essais cliniques pour traiter les symptômes d'infections respiratoires (Lauková *et al.*, 2020). À titre d'exemple, cette enzyme est couramment utilisée pour réduire la viscosité du mucus dans les voies respiratoires de patients atteints de fibrose kystique, ce qui facilite son élimination (Pressler, 2008).

1.3.2 Enzymes de restriction

Aux débuts de la biologie moléculaire dans les années 70 et 80, les enzymes de restrictions ont été énormément utilisées pour cartographier l'ADN. Grâce à leurs nombreuses spécificités de séquence, ces enzymes permettent de comparer différentes séquences selon leur profil de restriction, chose qui a été particulièrement utilisée en médecine légale (Di Felice *et al.*, 2019). À la suite de la synthèse, de l'optimisation, et de la commercialisation de plusieurs centaines d'enzymes de restriction, ces dernières servent encore aujourd'hui à diverses techniques de clonage à des fins de manipulation génétique. En combinant l'usage de ces ciseaux moléculaires

avec d'autres enzymes appelées ligases, n'importe quel fragment d'ADN peut être recombinaison à un autre, permettant la génération de mutants et de protéines recombinantes (Cohen, 2013).

1.3.3 Nucléases programmables

À l'inverse des enzymes de restriction dont la spécificité est intrinsèque à la séquence protéique et difficilement modifiable, plusieurs familles de nucléases ont été adaptées pour être programmables et ainsi couper n'importe quelle séquence d'intérêt (**Figure 2**). Cette modularité a permis l'expansion assez récente d'un domaine de recherche extrêmement prometteur appelé ingénierie génétique.

Méganucléases

Les méganucléases sont les premières enzymes ayant été étudiées pour leur potentiel en ingénierie génétique. En effet, ces dernières sont associées à de grands sites de reconnaissance (20-30 bp) leur conférant une forte spécificité et limitant énormément les risques de clivage hors cible, qui peuvent être catastrophiques dans un contexte clinique. Bien que plusieurs études supportent les applications *in vivo* de ces enzymes, celles-ci sont particulièrement difficiles à reprogrammer, ce qui a mené au développement d'autres stratégies plus prometteuses (Epinat *et al.*, 2013).

Nucléases à doigts de zinc (ZFN)

Les ZFN sont composées de plusieurs domaines de liaison à l'ADN appelés doigts de zinc, fusionnés au domaine de clivage de l'enzyme de restriction FokI. FokI a comme particularité de cliver à une distance précise de son site de reconnaissance, tandis que les doigts de zinc peuvent lier de façon séquence-spécifique des triplets de nucléotides. La combinaison de plusieurs doigts de zinc permet de programmer le site de liaison de l'enzyme chimérique, qui clivera ensuite l'ADN quelques nucléotides plus loin grâce au domaine de FokI. Puisque FokI ne fonctionne que sous forme dimérique, une ZFN est constituée de deux sous-unités ciblant chacune un brin d'ADN pour permettre une cassure double-brin. Les cassures spécifiques créées par les ZFN sont ensuite réparées par la machinerie de l'hôte, ce qui permet de muter, déléter ou même d'insérer une séquence d'intérêt en fournissant de l'ADN donneur contenant la modification voulue (Urnov *et al.*, 2010). En raison de la simplicité de ces systèmes, les ZFN sont les premières nucléases programmables à avoir été utilisées en thérapie génique, notamment contre le VIH et les glioblastomes (Chandrasegaran, 2017). Le principal obstacle à leur utilisation clinique repose sur leur spécificité imparfaite, qui occasionne parfois des coupures non spécifiques pouvant mener à une forte toxicité (Urnov *et al.*, 2010).

Nucléases effectrices de type activateur de transcription (TALEN)

Dans une optique d'optimiser la liaison à l'ADN des ZFN, les domaines à doigts de zinc ont récemment été substitués par d'autres domaines provenant de *transcription activator-like effectors* (TALE), tout en gardant le domaine nucléase de FokI. Alors que les modules ZFN reconnaissent des triplets de nucléotides, les modules TALE sont spécifiques à chaque nucléotide. De plus, cette spécificité est liée à seulement deux acides aminés, ce qui rends les TALE beaucoup plus modulables et faciles à programmer que les ZFN (Tröder & Zevnik, 2022). Bien que beaucoup d'applications thérapeutiques des TALEN cèdent progressivement la place aux systèmes CRISPR-Cas, ces enzymes demeurent particulièrement intéressantes pour cibler l'ADN mitochondrial, apporter des modifications épigénétiques ou agir comme modulateurs transcriptionnels (Bhardwaj & Nain, 2021).

CRISPR-Cas et Argonautes (endonucléases guidées)

Les systèmes CRISPR-Cas ont initialement été identifiés comme des mécanismes de défense chez les procaryotes s'apparentant à une immunité adaptative. Récemment, un de ces systèmes a été détourné pour servir en ingénierie génétique, révolutionnant ce domaine et octroyant un prix Nobel de chimie à ses pionnières, Jennifer Doudna et Emmanuelle Charpentier. La spécificité de clivage des nucléases Cas pouvant être modulée simplement à l'aide d'ARN guide, cette nouvelle technologie est beaucoup plus accessible et versatile que les précédentes. Les systèmes CRISPR-Cas ne se limitent pas aux thérapies géniques, ils ouvrent également la voie à une multitude d'autres applications comme de nouvelles techniques d'imagerie cellulaire, de criblage génomiques à haut débit et même comme outils diagnostiques (Knott & Doudna, 2018). Cependant, comme toute technologie récente, ces systèmes ont plusieurs limitations dont le besoin en motifs particuliers appelés PAM. Les protéines Argonautes, dont certaines sont guidées par de l'ADN plutôt que de l'ARN, ne nécessitent pas de motifs particuliers ce qui leur permet de cibler davantage de séquences que les CRISPR-Cas. De plus, les guides d'ADN sont plus faciles et moins chers à produire que les guides d'ARN, et les nucléases Argonautes sont plus petites que les Cas, facilitant leur insertion dans les cellules cibles. Cependant, ces enzymes ne sont actives qu'à haute température. Leur développement à des fins d'ingénierie génétique ne date que de quelques années et est encore préliminaire, mais ces protéines se présentent comme une alternative prometteuse aux systèmes CRISPR-Cas (Yang *et al.*, 2020).

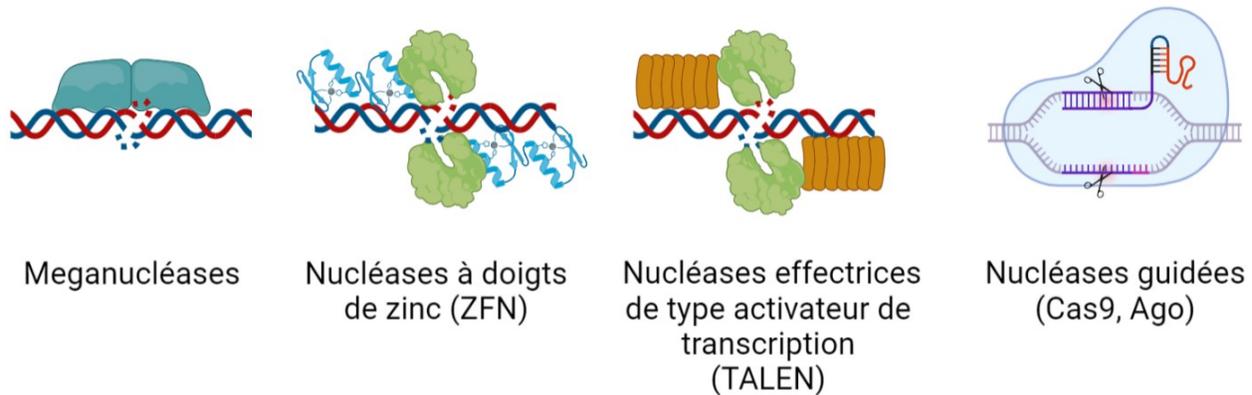


Figure 2 Principales familles de nucléases utilisées en ingénierie génétique

Les méganucléases permettent des coupures précises dans l'ADN double-brin à l'aide de leurs grandes séquences de reconnaissance. Les ZFN couplent un domaine nucléase à des modules de liaison d'ADN spécifiques appelés doigts de zinc. Les TALEN sont également des nucléases chimériques dont les domaines de liaison à l'ADN proviennent des protéines TAL. Les nucléases guidées n'ont pas besoin d'être modifiées pour changer leur spécificité, il suffit de leur fournir un ADN ou un ARN guide qui va moduler leur site de clivage. Figure personnelle créée avec BioRender.com

1.4 Superfamille des nucléases GIY-YIG

Les nucléases GIY-YIG représentent une grande famille d'enzymes clivant l'ADN ou l'ARN provenant de tous les domaines du vivant. Ces protéines tirent leur nom d'un domaine fonctionnel d'une centaine d'acides aminés contenant les motifs conservés Gly-Ile-Tyr et Tyr-Ile-Gly (GIY-YIG), séparés par une dizaine d'acides aminés variables. Les domaines GIY-YIG se démarquent également par un repliement tridimensionnel composé de trois feuillets bêta antiparallèles entourés de plusieurs hélices alpha. Les protéines de cette superfamille possèdent pratiquement toutes des domaines fonctionnels supplémentaires qui guident leur spécificité et leurs fonctions biologiques (**Tableau 1**). Ces nucléases sont notamment impliquées dans la maintenance du génome, l'expression de gènes ainsi que des mécanismes de défense (Dunin-Horkawicz *et al.*, 2006). Les différentes familles de GIY-YIG ainsi que leurs fonctions biologiques sont détaillées ci-dessous.

Tableau 1 Familles de nucléases GIY-YIG

Familles	Domaines fonctionnels*	Cibles	Fonctions
UvrC-like	UvrBb, EndoV, HhH, Cho, ExoIII	Lésions d'ADNdb	Réparation de l'ADN
Endonucléase de type « hopping » (HEase)	NUMOD, HTH	ADNdb (12-40 bp), clive à distance du site de reconnaissance	Mobilité d'éléments génétiques, principalement de phages
Penelope-Like Elements (PLE)	RVT	ADNdb	Mobilité de rétroéléments par l'intermédiaire d'ARN
COG3680	ANK, LEM	Chromatides en ségrégation	Réparation d'ADN et segregation de chromosome
COG1833	EndoIII, His/Cys-rich	ADNdb	Réparation d'ADN
Six1	RING finger	ADNsb, ADNdb, structures d'ADN embranchées	Réplication d'ADN ribosomal et réparation d'ADN
Enzymes de restriction (REase)	Extensions**	ADNdb (4-6 bp)	Protection du génome contre l'ADN exogène
MSH1	MutS-like	Structures d'ADN embranchées	Réparation d'ADN mitochondrial et chloroplastique

ANK Ankyrin, LEM *Nuclear membrane-associated* RVT transcriptase inverse, HTH *helix-turn-helix*, HhH *helix-hairpin-helix*, NUMOD *nuclease-associated modular DNA-binding domain*.

*En plus du domaine GIY-YIG. Liste non-exhaustive.

**Extensions importantes à l'activité enzymatique, mais non-identifiées comme des domaines distincts.

1.4.1 UvrC-like

Cette famille est composée des protéines UvrC et Cho (UvrC-homologue) majoritairement retrouvées chez les bactéries. UvrC est une composante du complexe UvrABC, crucial dans la réparation de dommages à l'ADN chez les procaryotes par la voie d'excision de nucléotides (NER). UvrC lie spécifiquement l'ADN embranché créé lors du NER grâce à son domaine C-terminal HhH, tandis que ses domaines EndoV et GIY-YIG catalysent respectivement le clivage en 5' et en 3' de la lésion (Singh *et al.*, 2002). Une étude très récente a également montré que UvrC de *H. pylori* est essentielle à l'intégration de courts fragments d'ADN exogènes au chromosome par compétence naturelle, démontrant un rôle important dans le dynamisme génomique de cette espèce (Ailloud *et al.*, 2022). Cho est un homologue de UvrC également impliqué dans la réparation d'ADN par NER, mais par un mécanisme légèrement différent. Pour Cho, le domaine GIY-YIG catalyse la coupure en 3' de la lésion mais à un endroit différent de UvrC, tandis que le domaine ExoIII est responsable de la coupure en 5' (Van Houten *et al.*, 2002).

1.4.2 Endonucléases de type « homing » (HEases)

Les nucléases GIY-YIG sont reconnues comme l'une des six familles structurales d'endonucléases de type « homing ». Les HEases GIY-YIG proviennent majoritairement de phages ou sont retrouvées dans des éléments génétiques mobiles de l'ADN mitochondrial et chloroplastique. Celles-ci sont typiquement composées du domaine GIY-YIG en N-terminal, responsable de l'activité nucléase, et de plusieurs modules de liaison à l'ADN appelés NUMOD pour *nuclease-associated modular DNA-binding domains* (Dunin-Horkawicz *et al.*, 2006). Comme leur nom l'indique, cette famille de nucléases sert à la mobilité de l'élément génétique qui les encode vers des allèles qui ne le possèdent pas encore. I-TevI, provenant du bactériophage T4, est l'enzyme la mieux caractérisée appartenant à cette famille. Cette dernière est séparée en deux unités fonctionnelles complètement indépendantes reliées par un pont flexible; le domaine GIY-YIG en N-terminal qui catalyse le clivage, ainsi que la région C-terminale qui permet la liaison spécifique à une large séquence de reconnaissance de 37 pb (Liu *et al.*, 2006). Cette modularité est à la source de nombreuses applications biotechnologiques combinant le domaine nucléase GIY-YIG de I-TevI avec d'autres domaines fonctionnels (Kleinstiver *et al.*, 2012; Pereira *et al.*, 2018; Wolfs *et al.*, 2016). Contrairement aux autres familles de nucléases GIY-YIG, les HEases agissent sous forme de monomères qui clivent séquentiellement chaque brin d'ADN (Kleinstiver *et al.*, 2013). D'autres membres de cette famille, SegA-H, ont une fonction similaire à la différence que leur site de clivage peut se situer à plusieurs centaines de paires de bases de leur site de reconnaissance (Edgell *et al.*, 2010).

1.4.3 Penelope-like Elements (PLEs)

Cette famille de nucléases GIY-YIG est exclusivement retrouvée chez les eucaryotes et constitue une classe de rétroéléments peu étudiés. Les PLEs comportent en plus de leur domaine GIY-YIG un domaine de transcriptase inverse en N-terminal, ce qui permet leur mobilité à l'aide d'un intermédiaire d'ARN. Outre leur fonction d'élément génétique mobile, les PLEs sont associées à des mutations délétères chez la drosophile (Evgen'ev & Arkhipova, 2005).

1.4.4 COG3680

Peu étudiée, cette famille de protéines GIY-YIG est retrouvée chez les eucaryotes ainsi que chez quelques bactéries, dont les *Neisseria*. La majorité de ces nucléases demeurent hypothétiques puisqu'elles n'ont pas été caractérisées, à l'exception de la protéine LEM-3 chez *C. elegans*. Cette dernière contient en N-terminal des motifs ankyrin (ANK) et un domaine LEM, et est encore

une fois impliquée dans la réparation de l'ADN par la voie NER, mais aussi dans la ségrégation de chromosome (Dittrich *et al.*, 2012).

1.4.5 COG1833

Présentes principalement chez les archées et quelques bactéries, aucune étude fonctionnelle n'a été réalisée sur cette famille. Cependant, certaines protéines COG1833 possèdent un domaine EndoIII, laissant supposer un rôle dans la réparation d'ADN par la voie BER (Dunin-Horkawicz *et al.*, 2006).

1.4.6 Famille Slx1

Cette famille de nucléases est impliquée dans la réplication de l'ADN ribosomal chez les eucaryotes, ainsi que dans la réparation de l'ADN chez les levures. Ces nucléases sont structure-spécifiques (Fricke & Brill, 2003). En plus de leur domaine GIY-YIG, plusieurs Slx1 arborent en C-terminal un domaine RING finger qui leur permet d'interagir avec leur cible d'ADN ou d'autres partenaires protéiques (Dunin-Horkawicz *et al.*, 2006)(1). Seule la protéine YhbQ de *E. coli* ne possède que le domaine GIY-YIG dans la famille Slx1. En absence de domaines supplémentaires pour restreindre sa spécificité, YhbQ clive de nombreux substrats dont l'ADN simple-brin, double-brin, et certaines structures secondaires. Cette protéine est impliquée dans des mécanismes de réparation d'ADN (Kumar *et al.*, 2016). Très récemment, une enzyme de la famille Slx1 a été caractérisée chez les plantes ; At-HIGLE. Bien qu'elle ne possède pas d'autre domaine fonctionnel annoté, celle-ci affiche une extension de près de 200 acides aminés qui autorégule l'activité catalytique du domaine GIY-YIG en N-terminal. Tout comme YhbQ, At-HIGLE est capable de cliver différentes structures embranchées d'ADN via sa dimérisation (Verma *et al.*, 2022).

1.4.7 Enzymes de restriction (REases)

Seules quelques enzymes de restriction caractérisées appartiennent à la superfamille des GIY-YIG. Contrairement aux autres sous-familles, les membres de celle-ci ne possèdent pas de domaines fonctionnels additionnels, mais plutôt des extensions à leur domaine GIY-YIG qui pourraient expliquer leurs spécificités de séquence (Dunin-Horkawicz *et al.*, 2006). Les nucléases R.Eco29kl, R.NgoMIII, R.Mral et R.Cfr42I proviennent de différentes bactéries et affichent différentes séquences protéiques, mais sont des isoschizomères, c'est-à-dire qu'ils ont la même spécificité de liaison et de clivage (CCGC/GG). Ces quatre nucléases sont encodées directement

à côté de méthyltransférases possédant la même spécificité, confirmant leur rôle dans la protection du génome contre l'ADN exogène par mécanisme de Restriction-Modification (système RM). À l'exception de R.Cfr41I qui forme un tétramère pour lier deux copies séparées de son site de reconnaissance, ces nucléases sont actives sous forme de dimères (Ibryashkina *et al.*, 2009). R.Hpy188I est une autre enzyme de restriction de cette famille accomplissant la même fonction. Cependant, cette dernière possède une lysine à la place de la tyrosine conservée du motif « YIG » et affiche une spécificité de séquence complètement différente des précédentes enzymes (Kaminska *et al.*, 2008).

1.4.8 MSH1

Initialement attribuées à la famille Slx1, les MutS homologs 1 (MSH1) ont récemment été regroupés dans une famille distincte de nucléases GIY-YIG. Ces dernières sont les homologues chez les plantes des protéines MutS prokaryotes importantes à la voie MMR de réparation d'ADN. Exprimées dans les mitochondries et les chloroplastes, les MSH1 sont importantes à la maintenance du génome de ces organelles. Ces protéines abritent un domaine GIY-YIG en C-terminal, et un ou plusieurs domaines MutS-like en N-terminal. À l'inverse de toutes les autres familles présentées jusque-là, c'est le domaine GIY-YIG des MSH1 qui leur permet de se lier à différentes structures embranchées d'ADN, tandis que les domaines MutS-like sont essentiels à leur activité nucléase (Fukui *et al.*, 2018). Cette récente découverte laisse envisager une diversité encore inexplorée d'activités catalytiques associée aux domaines fonctionnels GIY-YIG.

1.4.9 Nucléases GIY-YIG non classifiées

Il existe des milliers de petites protéines réparties dans l'ensemble du vivant qui arborent comme seul domaine fonctionnel le domaine GIY-YIG. Ces dernières sont annotées comme des nucléases hypothétiques ou des pseudogènes en raison de leur petite taille et de l'absence d'homologues caractérisés. Vu la grande diversité de la superfamille des GIY-YIG, tant au niveau de sa distribution phylogénétique que de ses fonctions biologiques, il ne serait pas surprenant d'y voir plusieurs ajouts dans le futur comme la très récente famille MSH1.

2 BACTÉRIES DU GENRE *NEISSERIA*

2.1 Généralités

Appartenant à la classe des betaprotéobactéries, à l'ordre des *Neisseriales* et à la famille des *Neisseriaceae*, les *Neisseria* sont des bactéries à Gram négatif retrouvées fréquemment dans les muqueuses des mammifères. Ces microorganismes tirent leur nom du chercheur qui les a identifiés à la fin du 19^e siècle, Albert Neisser. Composé majoritairement d'espèces commensales, le genre *Neisseria* comporte seulement deux espèces pathogènes chez l'humain : *Neisseria meningitidis* et *Neisseria gonorrhoeae*. *N. meningitidis* est l'un des pathogènes responsables de la méningite bactérienne et de septicémies, tandis que *N. gonorrhoeae* est le seul agent causatif de la gonorrhée, une infection transmise sexuellement.

2.1.1 Importance en santé publique

Neisseria gonorrhoeae

La gonorrhée est l'une de quatre grandes infections transmises sexuellement (ITS) curables aux côtés de la chlamydie, la syphilis et la trichomonase. Selon l'Organisation mondiale de la santé (OMS), plus de 82 millions de nouveaux cas de gonorrhée ont été reportés en 2020 (WHO, 2021b). Non traitée, cette infection peut mener à plusieurs conséquences graves dont l'infertilité, le syndrome inflammatoire pelvien ou encore une septicémie. La transmission mère-enfant de cette pathologie est fréquente et peut mener à une fausse-couche, un accouchement prématuré et même la cécité chez le nouveau-né (Dolange *et al.*, 2018). Comme pour les autres ITS, la gonorrhée est également un facteur de risque important de contracter le VIH (Jones *et al.*, 2019).

Contrairement à *N. meningitidis*, il n'existe aucun vaccin efficace et approuvé pour prévenir la gonorrhée. Nous avons donc recours aux traitements antibiotiques pour traiter cette infection. Cependant, cette voie thérapeutique est particulièrement mise en danger par des taux de plus en plus élevés de résistances aux antibiotiques chez *N. gonorrhoeae*. En effet, des résistances sont observées pour les six précédents traitements antibiotiques recommandés, ne laissant comme dernier recours que les céphalosporines de dernière génération et l'azithromycine. Nous atteignons déjà la limite des traitements actuels, puisque plusieurs cas de résistance à ces derniers ont déjà été observés dans plusieurs pays, et la susceptibilité de *N. gonorrhoeae* à ces antibiotiques est en constante baisse (Unemo & Shafer, 2014). Concrètement, ce pathogène a développé des résistances à tous les antibiotiques utilisés en première ligne dans les 80 dernières

années, ce qui risque de nous placer très bientôt dans des situations d'impasse thérapeutique qui pourraient avoir de graves conséquences en santé publique. En raison de son incidence croissante et de ses multiples résistances aux antibiotiques, *N. gonorrhoeae* a récemment été identifiée comme une « priorité élevée » de recherche de nouveaux antibiotiques par l'OMS, ainsi que comme « menace urgente » par le *Center for Disease Control* (CDC).

Bien que l'incidence de la gonorrhée soit beaucoup plus importante dans certains pays et chez des groupes démographiques précis, cette infection est également un problème de santé publique majeur au Canada (**Figure 3**). En effet, son incidence a triplé au cours des dix dernières années, la classant au 2^e rang des ITS au pays avec plus de 35 000 cas déclarés en 2019 (ASPC, 2019). Parmi les souches isolées en 2017 et 2018, près du deux tiers étaient résistantes à au moins un antibiotique tandis qu'une souche sur cinq était multirésistante. L'Agence de la santé publique du Canada a lancé dans les dernières années le Programme de surveillance antimicrobienne du gonocoque (PSAG), preuve de l'actualité de cette menace.

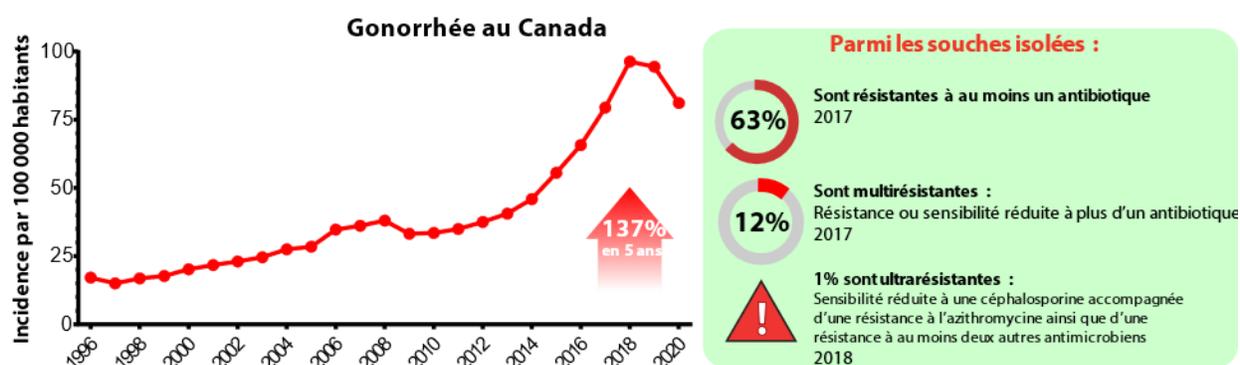


Figure 3 Incidence de la gonorrhée au Canada

Figure adaptée de mon article dans le média The Conversation (Chenal, 2021). Données de l'Agence de la santé publique du Canada (ASPC). La baisse marquée en 2020 est fort probablement liée à la pandémie de COVID-19.

Neisseria meningitidis

N. meningitidis est responsable de plusieurs maladies graves dont bien sûr la méningite bactérienne mais aussi des septicémies appelées méningococcies et parfois des pneumonies. Cette bactérie transmise par aérosols est fréquemment trouvée de façon asymptomatique dans le nasopharynx humain, mais arrive parfois à envahir la circulation sanguine où elle se réplique rapidement et mène à des infections aiguës. Les maladies à méningocoques sont associées à un très haut taux de mortalité, soit de 85%, qui demeure élevé même lorsque traitées (15%).

Même guéris, les survivants sont à haut risque de garder des séquelles permanentes comme des dommages au cerveau, une perte de l'ouïe ou encore une amputation (Rouphael & Stephens, 2012).

N. meningitidis possède une capsule polysaccharidique qui permet sa classification en 12 sérogroupes, dont six sont associés à la grande majorité des cas recensés. Contrairement à *N. gonorrhoeae*, plusieurs vaccins sont disponibles pour protéger contre les méningocoques, chacun ciblant un ou plusieurs sérogroupes. Cependant, l'accès limité à ces vaccins dans certaines régions du monde ainsi que la couverture limitée de chaque vaccin en termes de sérotypes freinent l'éradication de la maladie et ne parviennent pas à prévenir des épidémies ponctuelles (Tsang, 2021). À titre d'exemple, les États-Unis traversent depuis 2021 une importante épidémie de méningococcie dans la communauté LGBT (CDC, 2022). En 2021, l'OMS s'est donné comme objectif de vaincre la méningite bactérienne d'ici 2030, c'est-à-dire de prévenir toute nouvelle épidémie tout en diminuant drastiquement le nombre de décès et la morbidité qui sont associés à ces infections (WHO, 2021a).

Bien que moins inquiétante que pour *N. gonorrhoeae*, la résistance aux antimicrobiens chez *N. meningitidis* demeure préoccupante. En effet, vu la rapidité et la gravité des infections à méningocoques, un échec thérapeutique dû à une résistance est bien souvent létal (Rostamian *et al.*, 2022). Dans les trois dernières années, plusieurs isolats de *N. meningitidis* résistants à la pénicilline et à la ciprofloxacine ont été détectés aux États-Unis (McNamara *et al.*, 2020).

***Neisseria* commensales**

Bien que considérées comme non pathogènes, certaines espèces commensales de *Neisseria* peuvent agir comme pathogènes opportunistes et causer des infections chez des individus immunosupprimés ou à la suite d'une morsure d'animal (Liu *et al.*, 2015). Des infections invasives ont notamment été associées à *N. sicca*, *N. cinerea*, *N. mucosa* et *N. flavescens* suite à la prise d'un immunosuppresseur, l'eculizumab (Crew *et al.*, 2019).

À l'inverse, la capacité de certaines *Neisseria* commensales à induire une réaction immunitaire potentiellement protectrice contre les espèces pathogènes est à l'origine de nombreuses études visant à développer des vaccins (Dorey *et al.*, 2019). *N. lactamica* en particulier, est reconnue comme inhibant la colonisation par *N. meningitidis*. Cette action protectrice a été liée très récemment à la génération d'immunité croisée via la production d'anticorps anti-méningocoque (Dale *et al.*, 2022).

En plus de leur potentiel infectieux, les *Neisseria* commensales ont une importance en santé publique en raison de leur capacité à transmettre des résistances aux antibiotiques à leurs homologues pathogènes. En effet, les *Neisseria* sont reconnues comme étant naturellement compétente de façon constitutive, ce qui leur permet de forts taux de transferts horizontaux de gènes (HGT) intra et interespèces. Porteuses de plusieurs mutations et gènes de résistance, les espèces commensales agissent ainsi comme réservoirs génétiques. Plusieurs études récentes suggèrent que de nombreuses résistances observées chez le méningocoque et le gonocoque proviennent d'espèces commensales (Fiore *et al.*, 2020; Vanbaelen *et al.*, 2022).

2.1.2 Informations génomiques

Les *Neisseria* possèdent un génome d'environ 2.2 millions de paires de bases comprenant 51.5% de GC et abritant près de 2400 cadres de lectures ouverts. Leur génome, autant chez les espèces commensales que pathogènes, est considéré comme ouvert ce qui reflète leur grande diversité (Lu *et al.*, 2019). Pourtant, en raison de leur fréquence élevée de transferts horizontaux de gènes (HGT) interespèces, il est établi que les espèces pathogènes et commensales partagent le même répertoire de gènes (Linz *et al.*, 2000). Les *Neisseria* abritent un nombre variable de plasmides. Tandis que la majorité des souches *N. gonorrhoeae* en possèdent au moins un, la présence de ces éléments génétiques est très limitée chez *N. meningitidis* (van Passel *et al.*, 2006). Par contre, toutes deux possèdent un génome extrêmement riche en éléments répétés, qui seront discutés plus loin.

Polyplœdie

À l'inverse des espèces commensales qui ne possèdent qu'une copie de leur chromosome, les *Neisseria* pathogènes sont polyplœides homozygotes. Posséder deux à cinq copies identiques de leur chromosome contribuerait potentiellement au mécanisme de variation antigénique, important dans la génération de diversité génétique et dans la virulence des *Neisseria* (Tobiason & Seifert, 2010).

Systèmes de restriction-modification (RM)

Les systèmes de restriction-modification sont des mécanismes de défense bactériens contre l'ADN exogène provenant principalement de phages. Ces systèmes sont composés d'une enzyme de restriction associée à une méthyltransférase, qui reconnaissent toutes deux la même séquence. La méthyltransférase méthyle les sites de reconnaissance du génome de l'hôte, le protégeant ainsi de l'action de sa propre nucléase. En moyenne, les bactéries possèdent deux

systèmes RM. Chez les *Neisseria* pathogènes, jusqu'à 14 systèmes RM différents peuvent être trouvés dans une seule souche (Stein *et al.*, 1995). Cette énorme diversité est également retrouvée chez *Haemophilus influenzae* et *Helicobacter pylori*, deux autres espèces naturellement compétentes qui affichent également de très hauts taux de transferts horizontaux de gènes (HGT), confirmant le rôle de ces systèmes dans la protection du génome contre l'ADN invasif.

2.1.3 Diversité écologique, morphologique et pathologique

Comme mentionné précédemment, toutes les *Neisseria* pathogènes comme commensales partagent globalement le même réservoir de gènes, voulant dire qu'il existe très peu de gènes spécifiques à une seule espèce (Linz *et al.*, 2000). Pourtant, les différentes espèces de *Neisseria* ont des tropismes très variés, autant en termes d'hôtes que de tissus colonisés. À titre d'exemples, *N. musculi* a été isolée de la bouche d'une souris, *N. dentiae* de la plaque dentaire d'un bovin, *N. macacae* de la gorge d'un macaque et *N. iguanae* d'abcès d'un iguane (Liu *et al.*, 2015; Weyand *et al.*, 2016). D'autres encore ont été retrouvées dans l'environnement, suggérant qu'elles ne sont pas toutes des symbiotes obligatoires (Liu *et al.*, 2015).

Chose rare au sein d'un même genre bactérien, les *Neisseria* se démarquent également par la présence de plusieurs morphologies. En effet, on y retrouve des diplocoques, des coccobacilles et des bacilles. Si on s'étend à la famille des *Neisseriaceae*, on y retrouve même plusieurs assemblages de bacilles multicellulaires avec différentes polarités et modes de division (Nyongesa *et al.*, 2022b).

D'un point de vue pathologique, les infections à méningocoque et à gonocoque sont également très différentes malgré leur forte ressemblance génétique. *N. meningitidis* est parfois considérée comme commensale puisqu'elle fait partie de la flore normale des voies aériennes supérieures dans une certaine proportion de la population, mais peut évoluer en infection systémique soudaine associée à un taux très élevé de mortalité. En contraste, *N. gonorrhoeae* est considérée comme pathogène obligatoire du tractus urogénital même si elle est majoritairement asymptomatique ou associée à des symptômes légers. Cependant, de nombreux cas d'infections du tractus urogénital ont été associés à *N. meningitidis*, auquel cas la pathologie est presque indissociable de la gonorrhée (Burns & Rhoads, 2022).

2.2 Évolution et adaptation

2.2.1 Émergence des *Neisseria* pathogènes

Contrairement à d'autres genres bactériens, l'ARN ribosomal 16S ne parvient pas à bien dissocier les espèces de *Neisseria* en raison de leur forte similarité et des nombreux transferts horizontaux de gènes. Pour avoir une bonne résolution, il est nécessaire d'employer des approches basées sur de nombreux locus génétiques comme le génome noyau (*core genome*, 400 gènes), c'est-à-dire les gènes qui sont retrouvés dans la majorité des espèces (Maiden & Harrison, 2016). Les phylogénies inférées de cette façon suggèrent que toutes les *Neisseria* associées à l'humain proviennent d'un ancêtre commun (**Figure 4**). Ce dernier aurait colonisé la cavité buccale d'un humanoïde et ensuite perdu la capacité de survivre en dehors de son hôte (Seifert, 2019). Notre équipe a précédemment montré que cet ancêtre était un bacille qui a subi une transition morphologique vers un diplocoque pour lui permettre de coloniser le nasopharynx. Ce changement morphologique aurait permis une meilleure adhésion aux tissus ciblés tout en réduisant son exposition au système immunitaire (Veyrier *et al.*, 2015). Les populations ancestrales de la bouche et du nasopharynx auraient ensuite divergé et mené à l'émergence des différentes espèces commensales que l'on connaît aujourd'hui (**Figure 5**). Étant donné leur similarité génétique et phénotypique, les deux espèces pathogènes de *Neisseria* ainsi que les espèces proches comme *N. lactamica*, *N. polysaccharea* et *N. bergeri* ont vraisemblablement divergé assez récemment à partir du même ancêtre plutôt qu'émergé indépendamment. Ce clade ne possédant pas de nom dans la littérature, est défini comme clade « prépathogène » dans cet ouvrage (**Figure 4**).

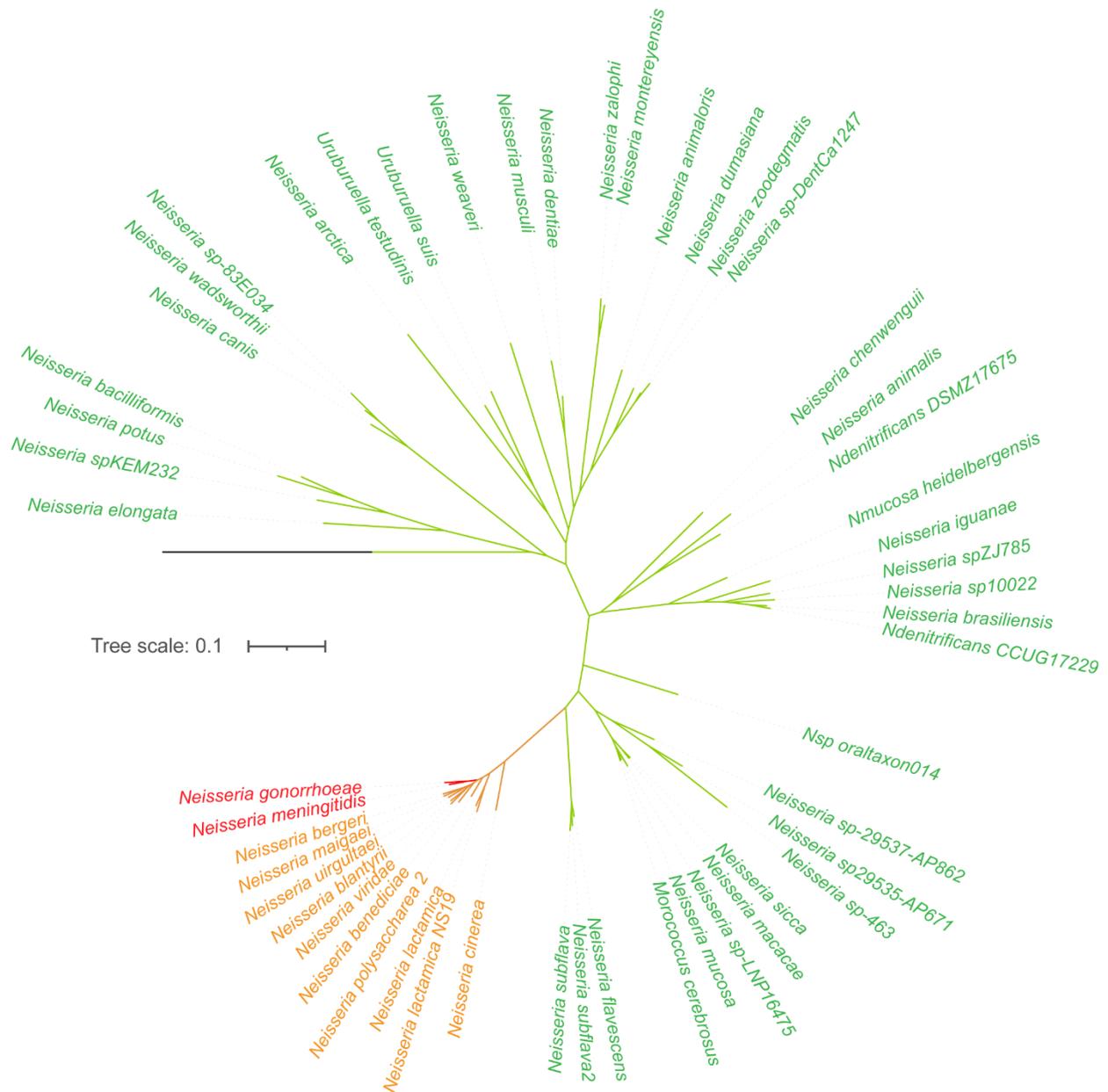


Figure 4 Phylogénie du genre *Neisseria*

Le génome noyau de 52 espèces de *Neisseria* a été utilisé pour générer cette phylogénie par l'algorithme *maximum likelihood*. Les couleurs représentent la progression vers la pathogénicité. Un ancêtre commensal (vert) aurait mené à l'émergence du clade en orange désigné ici « prépathogène », puis aux deux seules espèces considérées pathogènes chez l'humain (rouge) : *N. meningitidis* et *N. gonorrhoeae*. Image personnelle non publiée.

L'ordre précis des étapes ayant mené à l'émergence des deux pathogènes n'est pas encore bien défini. La première étape serait la capacité de leur ancêtre à induire une inflammation dans le nasopharynx, un élément clé qui différencie les espèces pathogènes des commensales (**Figure 5**)(Criss & Seifert, 2012). *N. meningitidis* aurait émergé à partir du nasopharynx en acquérant la

capacité à traverser la barrière hématoencéphalique et en survivant dans la circulation sanguine, tandis que d'autres espèces commensales auraient divergé sans développer davantage de facteurs de virulence. Un clone du même ancêtre se serait également relocalisé et adapté au tractus urogénital, menant à l'émergence de *N. gonorrhoeae* (Priniski & Seifert, 2018).

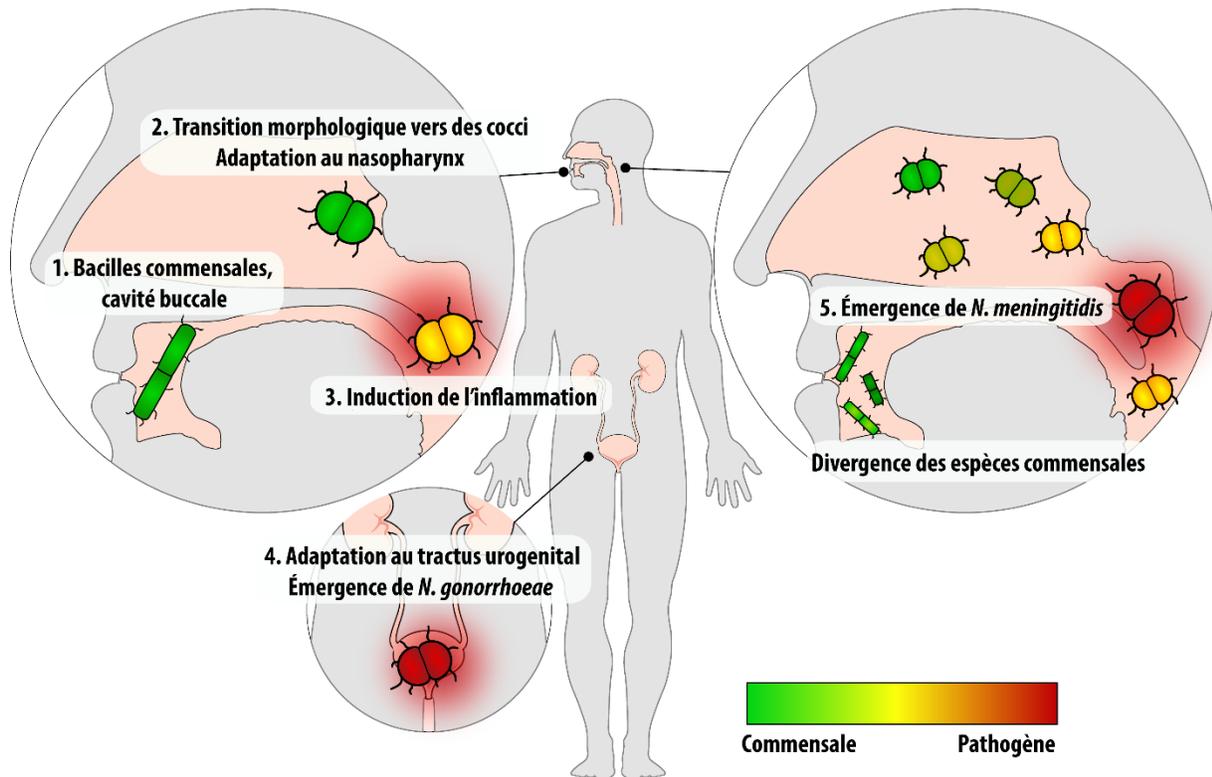


Figure 5 Origine et évolution des *Neisseria* commensales et pathogènes
Image personnelle non publiée.

2.2.2 Facteurs de virulence

Malgré leurs différences pathologiques, les *Neisseria* partagent globalement le même réservoir génétique, signifiant que très peu de gènes sont spécifiques à une seule espèce. Cette observation est également vraie pour les gènes dits de virulence, qui comme les autres sont fréquemment échangés entre les espèces dû à leur compétence naturelle et à leur colocalisation dans le nasopharynx (Marri *et al.*, 2010). Les principaux déterminants impliqués dans la virulence des *Neisseria* pathogènes ainsi que leurs fonctions et distribution au sein de ce genre sont détaillés dans le **Tableau 2**.

Tableau 2 Facteurs impliqués dans la virulence des *Neisseria*

	Facteurs	Fonctions	Distribution
Pathogènes + commensaux	Systèmes d'acquisition de fer (<i>tbp, lbp, hmb, hpu, fbp, fet, tdfF</i>)	Récepteurs et transporteurs de fer. Permettent la survie dans l'hôte, ainsi que la survie intracellulaire (<i>tdfF</i>).	<i>hmb, fet, fbp</i> : Toutes les <i>Neiss</i> <i>tbp, lbp, hpu</i> : Clade prépathogène. <i>tdfF</i> : <i>Nm, Ngo</i> <i>tdfG</i> : <i>Ngo</i>
	Variation de phase (<i>modA, modB, mutS, mutL</i>)	Expression ON/OFF d'une centaine de gènes impliqués dans le métabolisme et la virulence.	Clade prépathogène
	LOS (<i>lgt, rfa, waa, lpt</i>)	Endotoxine importante au recrutement des neutrophiles.	Toutes les <i>Neiss</i>
	Pili (<i>pil, comP</i>)	Adhésion, motilité et compétence naturelle.	Toutes les <i>Neiss</i>
	Opa (<i>opa</i>)	Adhésion, invasion, biofilms.	Clade prépathogène
	Adhésines (<i>opa, opc, nadA</i>)	Adhésion et invasion.	<i>opa</i> : Clade prépathogène <i>opc, nadA</i> : <i>Nm</i>
	Porines (<i>porA, porB</i>)	Adhésion et résistance au complément.	<i>porA</i> : <i>Nm</i> <i>porB</i> : Toutes les <i>Neiss</i>
	Systèmes toxine/antitoxine (<i>mafB/mafl, tpsA-C/tpsl</i>)	Toxines/antitoxines servant à la compétition interspèce.	<i>maf</i> : Clade prépathogène <i>tps</i> : <i>Nm</i>
Pathogènes seulement	Protéases (<i>iga, nalP</i>)	Protéases sécrétées servant à la résistance au complément.	<i>iga</i> : <i>Nm, Ngo, Nla</i> <i>nalP</i> : <i>Nm</i>
	Prophage MDAΦ	Colonisation, agrégation et invasion.	<i>Nm</i>
	Relâche de peptidoglycane (<i>lgtA, lgtD</i>)	Inflammation, recrutement de neutrophiles.	<i>Ngo</i>
	Gonococcal Genetic Island (GGI)	Encode pour un système de sécrétion de type IV	<i>Ngo</i> et quelques souches de <i>Nm</i> .
	Capsule (<i>cps</i>)	Invasion, protection contre le système immunitaire. Permet la classification en sérogroupes.	<i>Nm</i> seulement
	Variation antigénique (<i>pilE, pilS, recA, recO</i>)	Module la séquence ou l'expression de gènes codant pour des protéines de surface (LOS, Opa, PilE).	<i>Nm</i> et <i>Ngo</i>

Les gènes principaux impliqués dans chaque système sont indiqués entre parenthèses. Le clade prépathogène désigne le groupe monophylétique incluant toutes les espèces de *N. meningitidis* à *N. cinerea* dans la **Figure 4**. *Ngo* *N. gonorrhoeae*, *Nm* *N. meningitidis*, *Nla* *N. lactamica*. Tableau inspiré de (Criss & Seifert, 2012; Marri *et al.*, 2010; Priniski & Seifert, 2018).

2.2.3 Mécanismes de variabilité génétique

Peut-être en raison de sa vaste distribution écologique, *E. coli* possède de nombreux systèmes de régulation génétique comme des facteurs sigma et des systèmes à deux composantes pour s'adapter aux conditions changeantes de son environnement. À l'inverse, les *Neisseria* pathogènes sont des symbiotes obligatoires de l'humain dont le génome beaucoup plus petit contient très peu de ce genre de systèmes régulateurs (Davidsen & Tønjum, 2006). Plutôt que de répondre activement aux stimuli environnementaux, *N. meningitidis* et *N. gonorrhoeae* ont développés des systèmes pour générer constitutivement et aléatoirement une énorme diversité génétique qui contribue à la plasticité de leur génome et à leur potentiel pathogène.

Éléments répétés

Les *Neisseria*, tout particulièrement les espèces pathogènes, abritent au sein de leur génome un riche et abondant répertoire de séquences répétées de toutes les tailles. Outre leurs fonctions spécifiques, qui seront discutées plus loin, ces éléments contribuent à l'instabilité génétique des *Neisseria* par différents mécanismes dont la recombinaison et le *slipped strand mispairing* qui mènent à des délétions ou duplications.

Variation de phase

La variation de phase permet de rapidement et réversiblement changer le statut d'expression d'un gène (on/off) ou de passer d'un variant à un autre. Chez les *Neisseria*, plus de 100 gènes sont variables par ce système dont beaucoup sont impliqués dans la virulence. Ces derniers codent entre autres pour des adhésines, la capsule, les lipooligosaccharides (LOS) ou encore les pili (Davidsen & Tønjum, 2006). La variation de phase est médiée par un mécanisme appelé *slipped strand mispairing* (SSM) qui intervient lorsque de courtes répétitions d'ADN sont présentes dans une séquence codante ou son promoteur. On retrouve plus de 4000 courtes répétitions de ce genre (1-6 bp) dans le génome de *N. meningitidis*, avec une prédominance dans les gènes codant pour des protéines de la membrane externe (Siena *et al.*, 2016). Lors de la réplication ou de la réparation d'ADN, il n'est pas rare que la machinerie cellulaire apparie les mauvaises séquences entre elles en alignant incorrectement des régions répétées, occasionnant ainsi un glissement de brin qui peut changer le cadre de lecture ou altérer la région promotrice.

Concrètement, la variation de phase permet aux *Neisseria* de générer des populations hétérogènes, autant en termes de métabolisme qu'au niveau de leurs antigènes de surface. La présence de nombreux variants dans une population favorise l'adaptation à des changements environnementaux soudains et contribue également à l'évasion immunitaire (Davidsen & Tønjum, 2006).

Variation antigénique des pili

Contrairement à la génération de la diversité d'autres antigènes comme les adhésines Opa et les LOS qui intervient par variation de phase, les pili de type IV (Tfp) utilisent un autre mécanisme pour se diversifier, qui implique la recombinaison homologue par RecA. Tandis que la variation de phase affecte uniquement l'expression des gènes, la variation antigénique modifie directement la séquence des gènes codant pour les pili de type IV (Tfp) (Rotman & Seifert, 2014). De plus, seules les deux *Neisseria* pathogènes sont capables de variation antigénique. Les Tfp sont composés majoritairement de la piline encodée par *pilE*. En plus de *pilE*, les *Neisseria* pathogènes encodent un nombre variable allant jusqu'à 19 pseudogènes de piline tronqués et sans promoteurs appelés *pilS*, qui servent de réservoir génétique. Ces copies silencieuses *pilS* possèdent toutes une certaine homologie avec *pilE* qui permet leur recombinaison à divers endroits de la piline. Étant donné le nombre de différents *pilS* dans une souche donnée et le caractère aléatoire de la recombinaison homologue, qui peut échanger de très courtes comme de très longues séquences, ce mécanisme peut générer une incroyable diversité chez les Tfp, ce qui favorise l'évasion immunitaire (Criss *et al.*, 2005).

2.3 Compétence naturelle

Plusieurs espèces d'archées et de bactéries, dont les *Neisseria*, ont la capacité d'acquérir de l'ADN de leur environnement par un mécanisme de transformation (**Figure 6**). Cette particularité, appelée compétence naturelle, est retrouvée de façon très éparse dans la phylogénie des procaryotes et est modulée par des systèmes de régulation complètement différents selon les espèces, suggérant plusieurs apparitions et disparitions au cours de l'évolution (Johnston *et al.*, 2014). La transformation par compétence naturelle est un mécanisme majeur de transfert horizontal de gènes (HGT) reconnu comme ayant un fort impact sur l'évolution du genre *Neisseria*. En effet, cet attribut est important non seulement pour la réparation de dommages à l'ADN mais aussi pour l'acquisition de gènes et la génération de variabilité génétique (Mao & Lu, 2016).

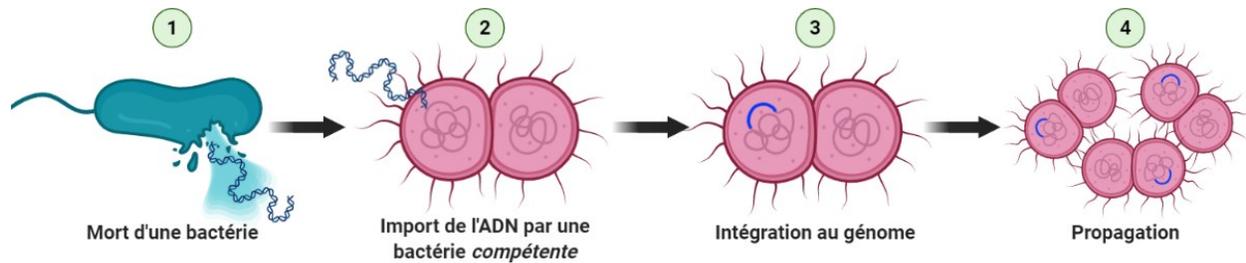


Figure 6 Schéma simplifié de transformation par compétence naturelle

Image créée avec BioRender.com, tirée de mon article dans La Conversation (Chenal, 2021).

Contrairement à la compétence « artificielle » utilisée en laboratoire pour transformer différentes bactéries, qui nécessite de physiquement ou chimiquement perméabiliser les membranes, la compétence naturelle est associée à une machinerie protéique complexe intrinsèque aux bactéries. Des protéines spécifiques sont impliquées dans toutes les étapes de la transformation, soit l'import d'ADN, son traitement et sa recombinaison au génome de l'hôte.

2.3.1 Origines de l'ADN transformant

L'ADN est retrouvé de façon ubiquitaire dans l'environnement, y compris dans la niche écologique des *Neisseria*, soit les voies aériennes supérieures et le tractus urogénital. En effet, ce matériel génétique est relâché en continu par les cellules de l'hôte et d'autres constituants du microbiote à leur mort. Afin de favoriser l'import de séquences possédant une bonne homologie au génome, ce qui facilite leur intégration par recombinaison homologue, les *Neisseria* ont développé plusieurs mécanismes de don d'ADN (**Figure 7**).

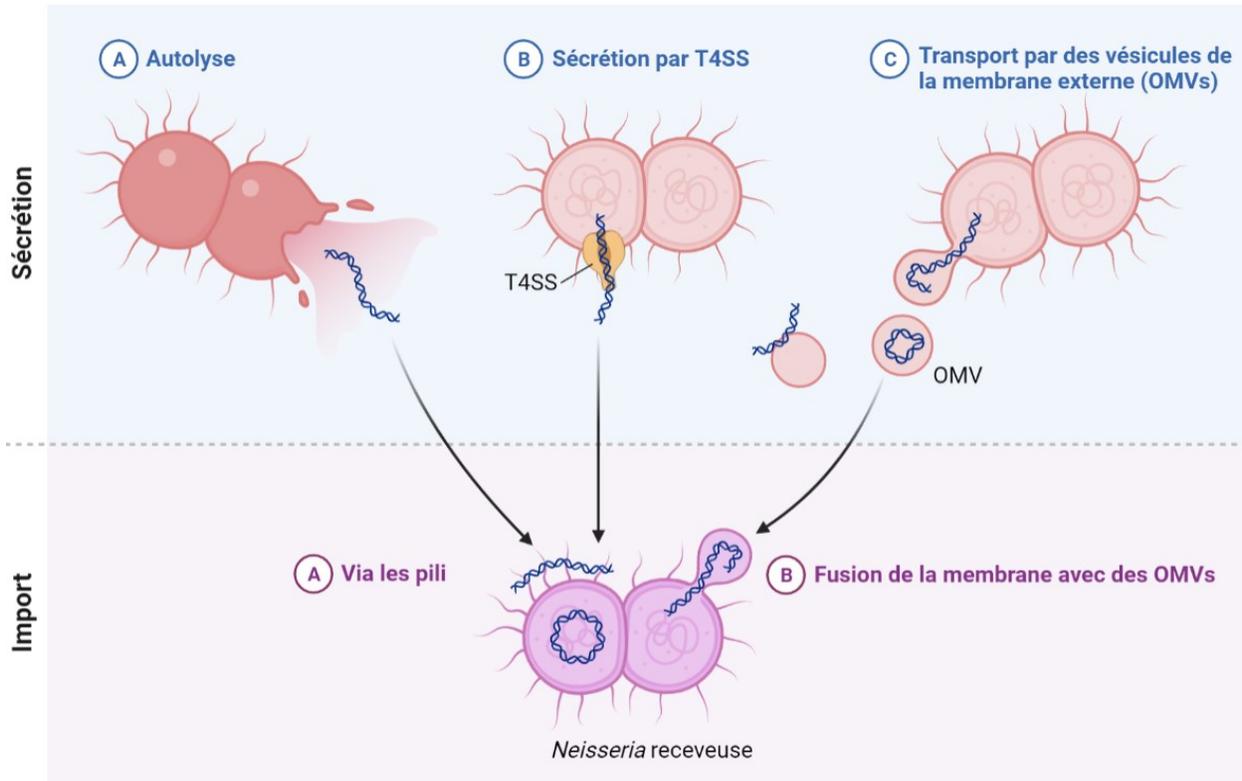


Figure 7 Don et import d'ADN par les *Neisseria*

Adaptation d'une image personnelle créée sur BioRender.com et republiée après édition (Chenal & Huang, 2023).

Autolyse

Dès l'atteinte de la phase stationnaire ou en réponse à un stress physiologique, les *Neisseria* pathogènes ont tendance à perdre très rapidement leur viabilité, ce qui se caractérise par un fort taux d'autolyse (Hebeler & Young, 1976). Cette autolyse est médiée par une expression modulable d'enzymes appelées amidases et transglycosylases qui dégradent la paroi cellulaire (Chan *et al.*, 2012). Ce phénomène serait un moyen de fournir des nutriments aux cellules plus « jeunes » lors de carences nutritionnelles et servirait aussi à fournir de l'ADN pour la transformation par compétence naturelle (Hamilton & Dillard, 2006).

Sécrétion active

Bien que l'autolyse soit une façon efficace de propager de l'ADN chromosomique dans son environnement, ce mécanisme implique forcément la mort de la cellule donneuse. Pour éviter ce sort, quelques souches de *N. meningitidis* et la majorité des souches de *N. gonorrhoeae* abritent au sein de leur génome un îlot génétique appelé *gonococcal genetic island* (GGI) qui encode pour un système de sécrétion de type IV (T4SS) (Dillard & Seifert, 2001). Ce système est reconnu

comme pouvant sécréter activement de l'ADN chromosomique dans le milieu extracellulaire, qui peut alors servir à la compétence naturelle. En plus d'être associé aux souches invasives à fort potentiel pathogène, ce système facilite la propagation de gènes de résistance aux antibiotiques (Callaghan *et al.*, 2017).

Transfert par des vésicules

Les vésicules de la membrane externe (OMVs) sont des liposomes sphériques de diverses tailles contenant toutes les composantes de la double membrane des bactéries à Gram négatif, mais aussi des protéines cytoplasmiques. Ces structures ont été associées à de nombreuses fonctions dont le trafic intracellulaire, la formation de biofilm, la résistance aux antibiotiques, la virulence, et la défense contre les phages (Reyes-Robles *et al.*, 2018; Schwechheimer & Kuehn, 2015). En raison de leur forte immunogénicité, les OMVs des *Neisseria* sont une cible de choix pour le développement de vaccins, y compris pour *N. gonorrhoeae* pour laquelle il n'existe pas encore de vaccin autorisé sur le marché (Christodoulides & Heckels, 2017; Petousis-Harris & Radcliff, 2019). En plus de protéines cytoplasmiques et membranaires, du matériel génétique peut également se retrouver dans des OMVs. D'ailleurs, c'est chez *N. gonorrhoeae* qu'a été démontré pour la première fois du transfert d'ADN génomique et plasmidique via des OMVs (Dorward *et al.*, 1989). Contrairement aux deux précédents systèmes, le matériel génétique retrouvé dans les OMVs est protégé des nucléases extracellulaires par la membrane lipidique.

2.3.2 Import d'ADN

Chez les *Neisseria*, l'ADN double-brin extracellulaire est capté par des structures spécifiques appelées pili de type IV (Tf4, *type IV pili*). Ces dernières sont composées de la piline majeure PilE et de plusieurs pilines mineures dont ComP et PilV. Les *Neisseria* ont la particularité de favoriser l'import d'ADN d'autres *Neisseria* grâce à la reconnaissance par ComP de séquences répétées retrouvées exclusivement et en forte abondance dans leur génome. Ces séquences, appelées *DNA uptake sequences* (DUS), diffèrent légèrement entre les espèces pathogènes et commensales de *Neisseria* ce qui leur permet de distinguer et de maximiser l'import d'ADN fortement homologue (Frye *et al.*, 2013). Elles sont détaillées plus loin dans cet ouvrage.

Une fois l'ADN lié aux Tfp, ces pili peuvent se rétracter grâce à l'action d'ATPases, tirant au passage l'ADN transformant au travers de pores de la membrane externe formées par la porine PilQ (**Figure 8**). Un récepteur d'ADN, ComE, lie l'ADN transformant immédiatement après son entrée dans le périplasma, l'empêchant de ressortir par un mécanisme de cliquet moléculaire (Hepp & Maier, 2016). Pour des raisons et par un mécanisme encore inconnus chez les *Neisseria*,

un brin est directement dégradé dans le périplasma, tandis que l'autre est transloqué au travers de la membrane interne via ComA (Hamilton & Dillard, 2006).

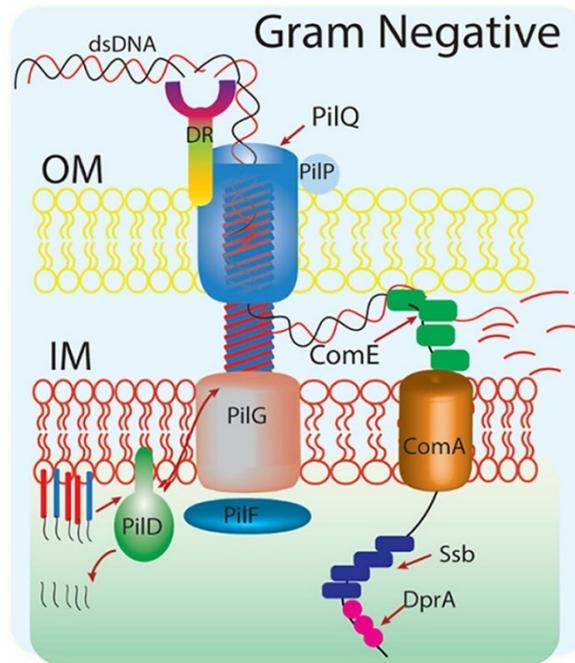


Figure 8 Mécanisme d'import d'ADN par compétence naturelle chez les *Neisseria*

L'ADN extracellulaire est capté par les pili de type IV via un récepteur spécifique (DR). La rétraction du pilus entraîne ensuite l'import de cet ADN double-brin dans le périplasma. Seul un brin d'ADN est transloqué au cytoplasme tandis que l'autre est dégradé dans le périplasma. OM membrane externe, IM membrane interne. Adaptée de (Sharma *et al.*, 2023).

2.3.3 Traitement de l'ADN et intégration au génome

Dès que de l'ADN simple-brin (ADNsb) se retrouve dans le cytoplasme, des protéines spécifiques vont s'y lier dont SsbA, SsbB et DprA. Ces dernières le protègent des nucléases cytoplasmiques, en plus de permettre la fixation de la protéine RecA, essentielle à l'intégration chromosomique par recombinaison homologue (**Figure 9**). Grâce à ses différents sites de liaison à l'ADN, RecA lie à la fois l'ADNsb transformant et l'ADN double-brin (ADNdb) de l'hôte. Cette recombinase scanne passivement les deux molécules à la recherche d'homologie en exploitant le déplacement aléatoire de celles-ci par diffusion (Bell & Kowalczykowski, 2016). Lorsque qu'une homologie suffisante est rencontrée, RecA catalyse l'échange de brin avec l'aide d'autres enzymes appelées résolvases. En absence d'homologie suffisante pour être recombiné au génome, l'ADN transformant est dégradé dans le cytoplasme (Hamilton & Dillard, 2006). De larges séquences non homologues peuvent être insérées si elles sont flanquées de séquences homologues au génome. Si jamais l'ADN transformant affiche un profil de méthylation différent, son intégration

au génome sera suivie d'une compétition entre les méthyltransférases et les nucléases des différents systèmes restriction-modification (RM) pour le protéger ou le dégrader. Ce phénomène a été suggéré très récemment lors de l'étude de transferts horizontaux de gènes par compétence naturelle entre des *Neisseria* commensales et pathogènes. En effet, les espèces commensales encodent moins de méthyltransférases que les pathogènes, rendant leur ADN potentiellement toxique lorsqu'intégré au chromosome puisqu'il pourrait être dégradé par les nucléases de l'hôte (Kim *et al.*, 2019).

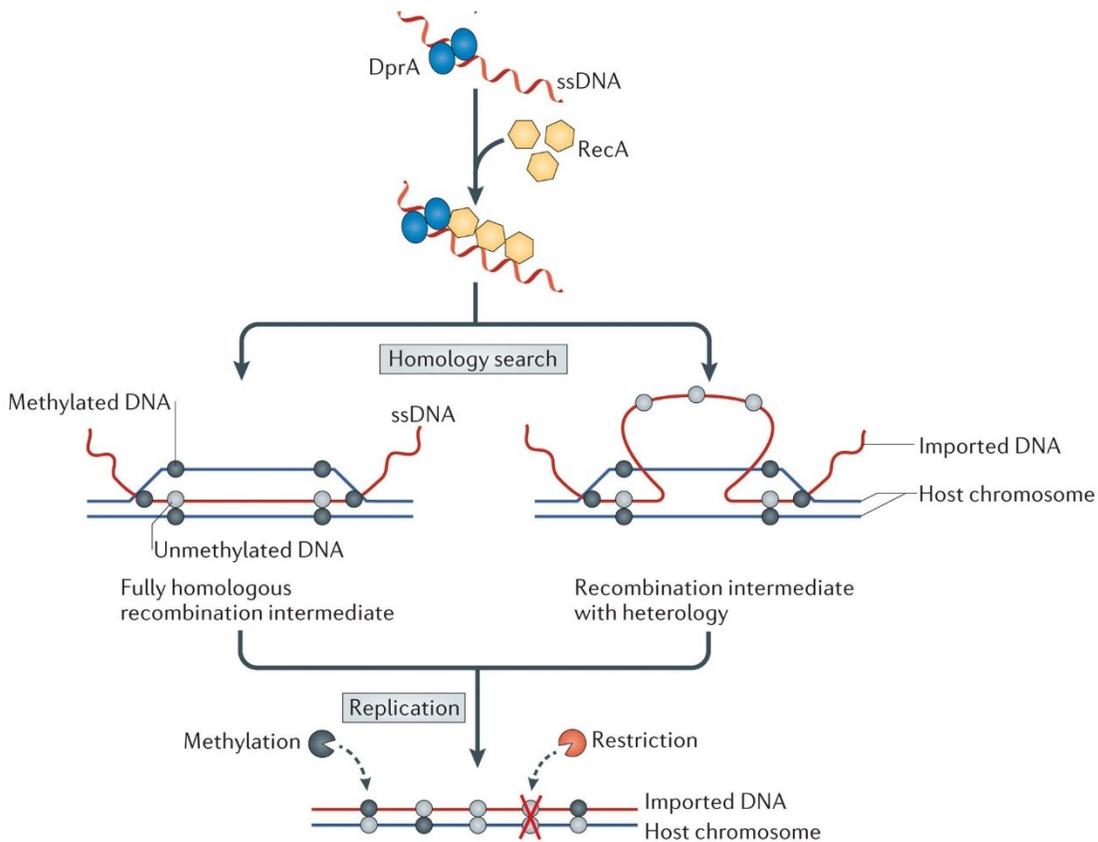


Figure 9 Traitement et intégration chromosomique de l'ADN transformant

Recrutée par des protéines liant spécifiquement l'ADN simple-brin, RecA permet l'intégration de l'ADN transformant au génome de l'hôte par recombinaison homologue. Si des régions non homologues se retrouvent entre des régions homologues, celles-ci peuvent être insérées dans le génome. Les systèmes de restriction-modification de l'hôte vont alors compétitionner pour méthyler ou cliver ces nouvelles séquences. Adaptée de (Johnston *et al.*, 2014).

2.3.4 Fonctions biologiques et rôles dans l'évolution

Dans la plupart des bactéries compétentes, les gènes codant pour la machinerie de compétence naturelle ne sont pas exprimés constitutivement. Ces derniers nécessitent des conditions particulières pour être induits, suggérant que la compétence n'est pas toujours bénéfique. De plus, cette capacité est très éparpillée dans la phylogénie des procaryotes, ce qui soulève des questions sur l'avantage évolutif qu'elle confère (Johnston *et al.*, 2014). En effet, vu le nombre de protéines impliquées, la machinerie de compétence entraîne nécessairement un coût énergétique non négligeable. Il existe trois grandes hypothèses quant à l'utilité de la compétence naturelle et de l'ADN importé : l'ADN comme source de nutriments, l'ADN comme substrat de réparation du génome, et la compétence comme mécanisme de diversification génomique. Ces hypothèses ne sont pas mutuellement exclusives et peuvent être applicables à certaines espèces plutôt que d'autres puisque la régulation de la compétence diffère beaucoup entre les bactéries compétentes.

ADN comme source de nutriments (*DNA for food*)

Cette première hypothèse se base sur le postulat que l'ADN est retrouvé de façon ubiquitaire et contient plusieurs nutriments essentiels comme des sucres, des phosphates et de l'azote. Plusieurs études démontrent que certaines espèces bactériennes peuvent utiliser l'ADN comme seule source de certains nutriments (Finkel & Kolter, 2001; Vorkapic *et al.*, 2016). Importer des acides nucléiques de l'environnement permet de minimiser leur synthèse *de novo*. L'ADN est particulièrement abondant dans les niches écologiques de plusieurs espèces naturellement compétentes, dont les *Neisseria*, *H. influenzae* et *S. pneumoniae* qui résident dans les voies aériennes de l'humain. L'ADN dans ces tissus provient en grande majorité de l'hôte, ce qui en fait un mauvais substrat pour la réparation de gènes bactériens et limite grandement ses chances d'être intégré au génome faute d'homologie suffisante. De plus, l'ADN environnemental est particulièrement exposé à une panoplie de nucléases extracellulaires qui le dégradent en mononucléotides (Redfield, 1993). Pour supporter cette hypothèse, la compétence chez *H. influenzae* est induite lors de carence nutritionnelle, en plus d'être maximale en début de phase stationnaire (Sinha *et al.*, 2013).

Compétence pour la réparation du génome

Chez plusieurs espèces compétentes dont *S. pneumoniae*, *H. pylori* et *L. pneumophila*, la compétence naturelle est induite par la présence d'agents endommageant l'ADN, suggérant

fortement un rôle dans la réparation du génome (Johnston *et al.*, 2014). Chez les *Neisseria*, la délétion de certains gènes importants dans la compétence naturelle s'accompagne de défauts de réparation d'ADN (Kline & Seifert, 2005; Stohl & Seifert, 2001). Le biais vers l'import d'ADN de la même espèce de certaines bactéries supporte également cette hypothèse. Cependant, vu que l'ADN homologue provient de bactéries mortes, on pourrait s'attendre à un certain biais vers la présence d'allèles mutés qui seraient à l'inverse délétères aux bactéries compétentes encore vivantes (Mell & Redfield, 2014).

Compétence pour la diversification du génome et l'acquisition de nouveaux allèles

Cette troisième hypothèse sur l'utilité de la compétence naturelle découle de nombreuses observations de HGT ayant mené à la transmission de résistances aux antibiotiques, à l'acquisition de gènes de virulence ou plus globalement à un avantage sélectif (Baltrus *et al.*, 2008; Blokesch, 2016). Sur les onze bactéries pathogènes identifiées comme « priorités de recherche » par l'OMS en raison de leurs multiples résistances aux antibiotiques, au moins huit sont naturellement compétentes (Lerminiaux & Cameron, 2019). Chez *S. pneumoniae*, une méthylase spécifique d'ADN simple-brin (DpnA) a été identifiée comme étant importante à la protection de l'ADN étranger lors de son entrée au cytoplasme par compétence naturelle. Exprimée uniquement lors du stade compétent du streptocoque, DpnA permet l'acquisition de larges fragments d'ADN hétérologues comme des îlots de pathogénicité. En absence de DpnA, ces fragments seraient dégradés par les enzymes de restriction de l'hôte une fois intégré au génome, causant sa mort (Johnston *et al.*, 2013). Il s'agit du premier mécanisme identifié favorisant spécifiquement l'intégration d'ADN étranger, ce qui contribue à la diversité génomique.

2.4 Séquences répétées des *Neisseria*

Le génome des *Neisseria* abrite un nombre et une diversité énorme de séquences répétées (**Tableau 3**). Ces éléments représentent environ 20% de leur génome total, une des proportions les plus élevées du domaine bactérien (Achaz *et al.*, 2002). En fournissant des zones de recombinaison, en modulant l'import d'ADN ou encore en participant à la dégradation d'ADN exogène, certaines séquences répétées contribuent à la stabilité et l'intégrité génomique tandis que d'autres fournissent au contraire des mécanismes de diversification génétique (Treangen *et al.*, 2009).

Tableau 3 Éléments répétés chez les *Neisseria*

Souche	DUS (12 bp)	dRS3 (20 bp)	CR (25-27 bp)	CREE* (100-150 bp)	ATR (183 bp)	Références
<i>N. meningitidis</i> MC58	1935	756	516	262	13	(Bentley <i>et al.</i> , 2007; Liu <i>et al.</i> , 2002)
<i>N. meningitidis</i> Z2491	1892	772	533	286	19	(Bentley <i>et al.</i> , 2007; Liu <i>et al.</i> , 2002)
<i>N. meningitidis</i> FAM18	1888	718	-	274	13	(Bentley <i>et al.</i> , 2007)
<i>N. meningitidis</i> alpha14	1851	646	-	269	12	(Budroni <i>et al.</i> , 2011; Schoen <i>et al.</i> , 2009)
<i>N. meningitidis</i> 053442	1858	725	-	234	18	(Budroni <i>et al.</i> , 2011; Roberts <i>et al.</i> , 2016; Schoen <i>et al.</i> , 2009)
<i>N. gonorrhoeae</i> FA1090	1965	208	254	122	-	(Bentley <i>et al.</i> , 2007)
<i>N. gonorrhoeae</i> NCCP11945	1966	-	-	131	-	(Bentley <i>et al.</i> , 2007; Roberts <i>et al.</i> , 2016)
<i>N. lactamica</i> ATCC 23970	2178	197	121	49	-	(Bentley <i>et al.</i> , 2007)
<i>N. lactamica</i> Y92-1009	2025	454	-	86	-	(Pandey <i>et al.</i> , 2017)
<i>N. lactamica</i> 020- 06	2244	297	-	92	-	(Pandey <i>et al.</i> , 2017)
<i>N. polysaccharea</i> 43768	2147	153	159	75	-	(Bentley <i>et al.</i> , 2007; Roberts <i>et al.</i> , 2016)
<i>N. cinerea</i> ATCC 14685	1915	5	28	13	-	(Bentley <i>et al.</i> , 2007)
<i>N. subflava</i> NJ9703	2875	2	1	0	-	(Bentley <i>et al.</i> , 2007)
<i>N. flavescens</i> NRL30031	2833	9	53	26	-	(Bentley <i>et al.</i> , 2007)
<i>N. sicca</i> 29256	3729**	35	570	257	-	(Bentley <i>et al.</i> , 2007)
<i>N. mucosa</i> 25996	3427**	63	210	92	-	(Bentley <i>et al.</i> , 2007)
<i>N. elongata</i> 29315	2141	24	116	56	-	(Bentley <i>et al.</i> , 2007)

Le nombre total de chaque élément répété par génome est répertorié.

DUS : DNA Uptake Sequence, dRS3 : duplicated Repeat Sequence 3, CR : Correia Repeat, CREE : Correia Repeat-Enclosed Element, ATR : AT-Rich Repeat

*Les paramètres d'analyses varient selon les études, donnant des chiffres légèrement différents pour une souche donnée. **Variante espèce-spécifique de DUS

2.4.1 DNA uptake sequence (DUS)

Les DUS permettent à la machinerie de compétence naturelle des *Neisseria* d'importer sélectivement de l'ADN provenant d'autres *Neisseria*. Ces séquences non palindromiques de 10 ou 12 pb représentent l'élément répété le plus abondant de toute la famille des *Neisseriaceae*, correspondant à environ 1% de leur génome (Frye *et al.*, 2013). Ces séquences ne sont pas ou très peu retrouvées dans d'autres bactéries, bien que les *Pasteurellaceae* possèdent leur propre équivalent appelé *uptake signal sequence* (USS).

Répartis dans l'ensemble du génome mais jamais dans des séquences codantes, les DUS sont distribués de façon légèrement biaisée à proximité de gènes conservés impliqués dans la maintenance du génome, leur suggérant un rôle dans l'intégrité génomique (Davidsen *et al.*, 2004). Deux tiers des DUS s'assemblent en paires inversées en 3' de gènes, permettant la formation de structures d'ARN en tête d'épingle qui agiraient comme terminateurs transcriptionnels *rho*-indépendant (Spencer-Smith *et al.*, 2016).

Les DUS sont reconnus et liés spécifiquement par la piline mineure ComP présente à la surface des pili de type IV (Cehovin *et al.*, 2013). Ces séquences ne sont pas absolument nécessaires à l'import d'ADN par la machinerie de compétence, mais leur présence augmente les taux de transformation des *Neisseria* de plusieurs ordres de magnitude. Pour des raisons et par un mécanisme encore inconnu, différents taux de transformation sont observés selon l'orientation du DUS sur un ADN simple-brin, suggérant que cette séquence possède d'autres rôles une fois importée dans le périplasme ou le cytoplasme des *Neisseria* (Duffin & Seifert, 2012).

En plus de favoriser l'import d'ADN d'autres *Neisseria*, les DUS permettent également de discriminer les espèces entre elles. En effet, les deux espèces pathogènes de *Neisseria* possèdent un DUS légèrement différent des espèces commensales. Au total, huit différents variants de DUS ont été identifiés, certains étant uniques à des espèces précises de *Neisseriaceae*. Le nombre total de ces répétitions est similaire entre les espèces pathogènes et commensales. La présence dans l'ADN transformant du variant correspondant à l'espèce receveuse augmente les taux de transformation (Frye *et al.*, 2013). Ainsi, les DUS servent également de barrière contre les HGT interspèces.

2.4.2 dRS3/NIME

Les *duplicated repeat sequence 3* (dRS3) sont des séquences semi-palindromiques de 20 pb représentant le deuxième élément répété le plus abondant chez les *Neisseria*. Elles sont beaucoup plus abondantes chez les espèces pathogènes que chez les espèces commensales, et ce tout particulièrement chez *N. meningitidis*. Tout comme les DUS, les dRS3 sont majoritairement retrouvées dans des régions non codantes sous forme de paires inversées, mais autant en 5' qu'en 3' de gènes. Cependant, à l'inverse des DUS, les dRS3 ne sont pas distribuées dans l'ensemble du génome. Elles s'accumulent plutôt dans des clusters à proximité de gènes hautement variables codant en grande majorité pour des protéines de surface, souvent impliquées dans la virulence (Parkhill *et al.*, 2000). Ces clusters, appelés NIMEs pour *Neisseria interspaced mosaic elements*, ne sont cependant pas retrouvés dans le génome des espèces commensales (Marri *et al.*, 2010). Bien qu'aucune explication n'ait été proposée, une étude montre une corrélation inverse significative entre la présence de dRS3 et de DUS, signifiant que ces deux éléments ont tendance à ne pas se retrouver à proximité l'un de l'autre.

En raison de leur distribution particulière, les dRS3 sont probablement impliquées dans la virulence et dans la variabilité génétique des *Neisseria* pathogènes en promouvant les transferts horizontaux de gènes et la variation antigénique (Bentley *et al.*, 2007). Une étude bio-informatique a en effet associé des clusters de dRS3 à des zones hypervariables appelées *hotspots* de recombinaison, et identifié plusieurs réarrangements ayant eu lieu précisément entre deux copies de ces éléments répétés (Budroni *et al.*, 2011; Schoen *et al.*, 2008). Cependant, aucune étude expérimentale n'a encore confirmé ou infirmé ces hypothèses. Toutefois, deux équipes ont simultanément prouvé chez *N. meningitidis* que les dRS3 sont la cible d'un type de phage filamenteux associé à une plus forte virulence. Ce dernier possède une transposase qui coupe spécifiquement l'ADN double-brin au milieu de certaines dRS3 (Bille *et al.*, 2005; Kawai *et al.*, 2005). Ces découvertes n'expliquent cependant pas pourquoi et comment les dRS3 s'accumulent autour de gènes hypervariables, signifiant que d'autres effecteurs protéiques interagissent probablement avec ces séquences.

Bien que les dRS3 et les NIMEs soient presque exclusivement retrouvés chez les *Neisseria*, ces éléments partagent beaucoup de similarités avec les familles d'éléments répétés REP (*repetitive extragenic palindromes*) et BIME (*bacterial interspaced mosaic elements*). Ces derniers, qui possèdent également des séquences semi-palindromiques et qui s'amassent aussi en large clusters autour de gènes, sont impliqués dans une myriade de processus biologiques, dont la terminaison transcriptionnelle, la liaison à des protéines de l'hôte, et des réarrangements

chromosomiques (Treangen *et al.*, 2009). Chez certaines bactéries, la présence et l'abondance de REP sont associées à des transposases clivant spécifiquement leurs séquences répétées respectives. Ces nucléases sont suggérées comme étant responsable de l'accumulation de ces REP (Nunvar *et al.*, 2013).

2.4.3 Répétitions Correia (CR) et Éléments Correia (CREE/NEMIS)

Les éléments Correia sont composés d'un noyau central de 50-100 pb flanqué de répétitions inversées de 25-27 pb appelées répétitions Correia. De structure similaire aux séquences d'insertion (IS), les CREE n'encodent cependant pas de transposase pour médier leur mobilité. Considérés comme éléments génétiques mobiles non autonomes, ils se classent dans la famille des répétitions inversées miniatures transposables (MITE). Les CREE sont distribués dans l'ensemble du génome, souvent à proximité de gènes avec lesquels ils sont co-transcrits ou près d'ARN non codants (Mazzone *et al.*, 2001; Roberts *et al.*, 2016). Tout comme les DUS et les dRS3, les CREE sont plus abondants chez les *Neisseria* pathogènes que les commensales, à l'exception de *N. sicca* qui en contient également beaucoup. Cette disparité n'est toujours pas expliquée, mais suggère un rôle des CREE dans l'évolution des *Neisseria* (Liu *et al.*, 2002).

Sous forme d'ARN, les CR forment des structures en épingle à cheveux qui interagissent avec la RNase III et jouent un rôle dans la régulation post-transcriptionnelle (De Gregorio *et al.*, 2002). Les CREE contiennent également un site de liaison de la protéine IHF (*integration host factor*), un régulateur transcriptionnel (Rouquette-Loughlin *et al.*, 2004). Enfin, une étude leur suggère un rôle dans la variation de phase de certains gènes chez *N. meningitidis* dû à la présence de promoteurs forts dans certaines de ces séquences (Siddique *et al.*, 2011).

2.4.4 Répétitions AT-riche (ATR)

Une vingtaine de séquences répétées de 183 pb avec un bas taux de GC sont présentes dans le génome de *N. meningitidis*. Ces éléments possèdent des répétitions inversées aux extrémités très similaires à celles de plusieurs séquences d'insertion (IS), mais n'encodent pas de transposase (Parkhill *et al.*, 2000). Dans certaines souches de méningocoque, une ATR est insérée entre le gène *fHbp* et son promoteur, modulant sa transcription (Sanders *et al.*, 2012). Les études sur ces éléments étant très limitées, on ne connaît pas leur distribution chez les *Neisseria* commensales, ni leurs autres implications biologiques.

2.4.5 Séquences répétées non uniques aux *Neisseria*

CRISPR

Les *clustered regularly interspaced palindromic repeats* (CRISPR) sont composés de courtes répétitions directes (24-47 pb) séparées par des espaceurs (*spacers*) variables (26-72 pb). Ces éléments agissent comme système immunitaire adaptatif des bactéries en les protégeant contre l'ADN étranger, majoritairement viral. Les *spacers* sont des fragments d'ADN de phages provenant d'infections antérieures, qui lorsque transcrits en ARN servent de guides aux nucléases Cas (CRISPR-associated) pour dégrader l'ADN étranger contenant les mêmes séquences.

Plusieurs espèces de *Neisseria* possèdent un système CRISPR-Cas de type IIC, dont *N. meningitidis* et quelques espèces commensales, mais pas *N. gonorrhoeae*. Cependant, ces systèmes sont limités à quelques souches, suggérant une transmission horizontale dans le nasopharynx (Zhang, 2017). Le système CRISPR-Cas de *N. meningitidis* est particulièrement intéressant d'un point de vue biotechnologique car il est plus simple que les systèmes précédemment découverts et nécessite moins d'étapes (Zhang *et al.*, 2013). De plus, la Cas9 du méningocoque possède une activité ssDNase guidée par l'ARN en plus de l'activité classique de clivage d'ADN double-brin, ce qui ouvre la porte à d'autres applications (Zhang *et al.*, 2015). Le système CRISPR-Cas9 a le potentiel de complètement bloquer la transformation d'ADN contenant des séquences homologues aux *spacers* retrouvées dans son génome. Ce mécanisme d'interférence de transformation par compétence naturelle limite l'intégration d'ADN viral au génome, tout en permettant l'import d'ADN d'autres *Neisseria* (Zhang *et al.*, 2013).

Séquences d'insertion (IS)

Les séquences d'insertion sont les plus petits éléments transposables autonomes car composées uniquement d'une transposase flanquée de répétitions inverses. Fortement associés au mode de vie des bactéries, ces éléments joueraient un rôle majeur dans leur évolution. En effet, on observe une accumulation massive d'IS associée à une restructuration du génome dans des organismes ayant récemment adopté un mode de vie symbiotique. À l'inverse, les symbiotes plus anciens possèdent très peu d'IS et un génome considérablement réduit (Siguiet *et al.*, 2014). *N. meningitidis* possède une des plus fortes concentrations en IS du domaine bactérien, appartenant principalement aux familles IS5, IS30 et ISNCY (Mahillon & Chandler, 1998). Ces IS sont associées à des réarrangements chromosomiques, des inactivations de gènes et même à la variation de phase de la capsule du méningocoque (Budroni *et al.*, 2011; Hilse *et al.*, 2000).

PROBLÉMATIQUE, HYPOTHÈSES ET OBJECTIFS

PROBLÉMATIQUE

On retrouve au sein des génomes des *Neisseria* plusieurs centaines de copies d'une séquence répétée appelée dRS3, presque exclusive à ce genre bactérien. La dRS3 est beaucoup plus abondante chez les espèces pathogènes que les commensales et affiche une distribution intrigante autour de gènes codant pour des protéines de surface, dont plusieurs facteurs de virulence (**Figure 10**). Aucune étude n'a démontré expérimentalement le rôle des dRS3, ni a su expliquer leur distribution particulière et leur relation avec la virulence des *Neisseria*. À l'exception d'une recombinase de phage, on ne connaît pas les protéines qui interagissent avec cet élément répété.

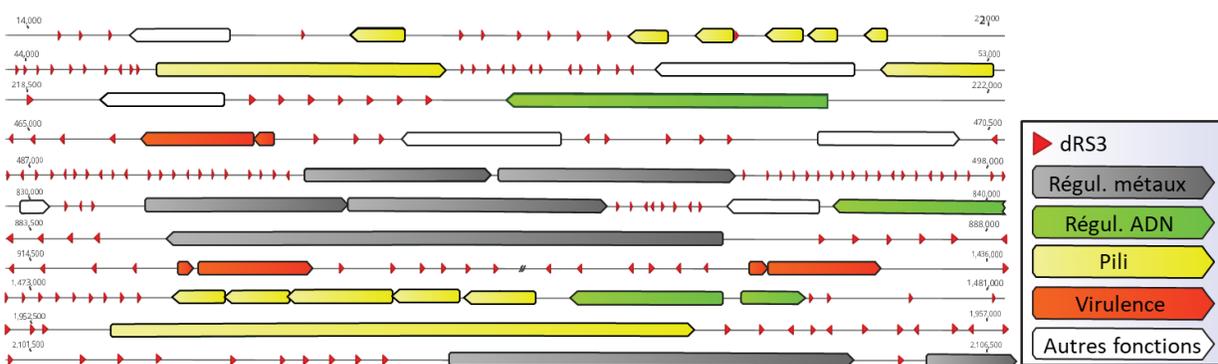
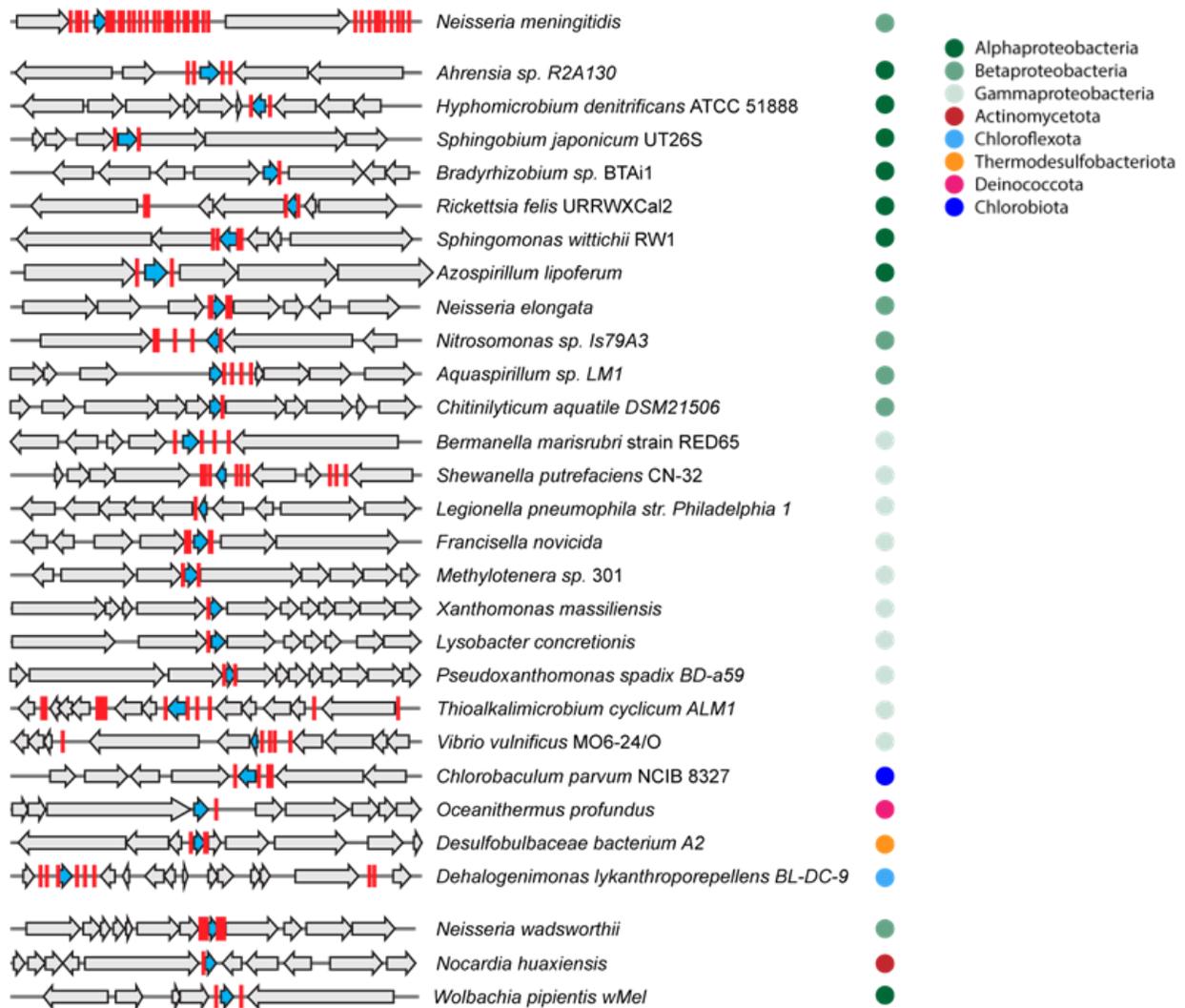


Figure 10 Distribution des clusters de dRS3 chez *N. meningitidis*

Les régions du génome de *N. meningitidis* 8013 abritant une forte concentration en dRS3 sont illustrées. Les dRS3 sont représentées par de courtes flèches rouges, tandis que les gènes à proximité sont illustrés par de plus longues flèches dont la couleur indique la fonction.

Au tout début de ce projet, notre équipe a réalisé des analyses bio-informatiques préliminaires sur des séquences répétées très similaires aux dRS3 que nous avons appelées NTS. Concrètement, ces analyses visaient à identifier des gènes présents uniquement dans des génomes contenant des NTS, et absents dans les génomes n'en contenant aucune. Ces analyses, conduites sur l'ensemble du domaine bactérien, ont révélé un gène précis qui est presque systématiquement flanqué de plusieurs dRS3/NTS (**Figure 11**). Ce gène, surnommé *ssnA*, encode pour une minuscule nucléase hypothétique.



OBJECTIFS

L'objectif global de cette étude est de caractériser de façon exhaustive la protéine encodée par le gène *NMV_0044* (SsnA), et d'explorer son interaction prédite avec les séquences répétées dRS3/NTS retrouvées chez les *Neisseria*. Afin d'avoir une image globale de l'impact de ces deux éléments dans l'évolution et l'adaptation des *Neisseria* pathogènes, des approches bio-informatiques, biochimiques et phénotypiques sont utilisées.

Les objectifs spécifiques suivants sont séparés en chapitres dans cet ouvrage :

- 1- Caractérisation bio-informatique – Quantifier et cartographier les dRS3 et les NTS chez les *Neisseria*. Confirmer l'association entre le gène *NMV_0044* (*ssnA*) et ces séquences répétées.
- 2- Caractérisation biochimique – Purifier SsnA et définir son activité enzymatique. Vérifier expérimentalement son interaction avec les séquences répétées.
- 3- Caractérisation phénotypique – Identifier les fonctions biologiques de SsnA. Élucider son rôle dans la compétence naturelle et la virulence de *N. meningitidis*.
- 4- Identifier des homologues de SsnA chez d'autres bactéries. Vérifier si ces derniers sont également associés à des séquences répétées et explorer leur potentielle interaction avec ces éléments.

1 CHAPITRE I – ÉTUDE BIO-INFORMATIQUE

Ce premier chapitre est dédié aux analyses bio-informatiques et génomiques visant à confirmer le lien entre les séquences répétées dRS3/NTS et un gène préalablement identifié dans notre équipe, que l'on a appelé *ssnA*.

MATÉRIEL ET MÉTHODES

1.1 Analyses des séquences répétées

1.1.1 Quantification des séquences répétées

Chez les *Neisseriaceae*

Afin de chercher de façon ciblée des séquences répétées précises comme les DUS, dRS3 ou NTS dans la famille entière des *Neisseriaceae*, un pipeline d'analyse maison basé sur Python a été généré. Ce dernier combine le programme fuzznuc de la suite EMBOSS, qui cherche des motifs répétés dans des séquences nucléotidiques, et l'outil de visualisation WebLogo, qui permet d'illustrer la séquence consensus des motifs trouvés (Crooks *et al.*, 2004; Rice *et al.*, 2000). Les dRS3 ont été cherchées uniquement dans le brin matrice des génomes en raison de leur séquence parfaitement palindromique, tandis que les DUS, NTS, et NTS_{var} ont également été cherchées dans le brin complémentaire. Aucune substitution n'a été permise pour les DUS et dRS3, tandis qu'une substitution à la séquence requête a été autorisée pour les NTS. Voici les séquences requêtes de fuzznuc pour chaque élément répété :

```
DUS :      GCCGTCTGAA
dRS3 :    ATTCCNNNNNNNNGGGAAT
NTS :     CGTCATTCCCGCGMAVGC GGGGAATCYRG
NTSvar :  CGTCATACTCGGGNNNNNCCCGAGTATC
```

Pour les DUS, en plus de la séquence générique présentée ci-dessus, six autres variantes ont été cherchées et compilées sans substitution possible à l'aide de fuzznuc, afin d'obtenir la totalité des répétitions séquence-spécifiques dans toutes les *Neisseriaceae* analysées (Frye *et al.*, 2013).

wadDUS : CCTGTCTGAA
mucDUS : GTCGTCTGAA
simDUS : GCTGCCTGAA
kingDUS : GCAGCCTGAA
king3DUS : GCAGCCTGCA
eikDUS : GCTACCTGAA

Chez les autres bactéries

Pour dénombrer les NTS associées au gène *ssnA* dans l'ensemble du domaine bactérien, des homologues de ce gène ont d'abord été identifiés par BLAST sur NCBI en utilisant un seuil de 50% d'identité. Parmi les 256 génomes identifiés, 17 espèces provenant de familles et de genres différents ont été sélectionnées aléatoirement. L'outil fuzznuc a ensuite servi à la détection de NTS dans les génomes entiers en permettant quatre substitutions à la séquence consensus détaillée plus haut. Pour dénombrer les NTS exclusivement associées aux gènes *ssnA*, les régions intergéniques immédiatement en amont et en aval de ces gènes, représentant 500-1000 pb, ont été extraites de NCBI, puis soumises à la même analyse. Comme contrôle, une analyse identique a été réalisée sur les régions flanquant le gène essentiel *gyrA*. Afin d'estimer le facteur d'enrichissement en NTS dans la région encodant *ssnA*, le nombre total de NTS a été divisé par la taille de chaque génome, donnant une fréquence théorique de NTS par kilobase si ces répétitions étaient distribuées aléatoirement. La fréquence réelle de NTS associée à *ssnA* a ensuite été divisée par cette fréquence théorique.

1.1.2 Représentations graphiques

Séquences consensus (logos)

Les logos ont été générés avec WebLogo 3.7.8 sans barres d'erreur et avec la palette de couleur classique, à partir des alignements générés par l'outil fuzznuc (Crooks *et al.*, 2004).

Structures secondaires

Le serveur RNAfold a été utilisé pour inférer la structure secondaire des différentes séquences d'intérêts, à l'aide des paramètres par défaut et de l'algorithme pour l'ADN (Gruber *et al.*, 2008). L'outil RNAalifold a été également utilisé pour visualiser la structure consensus plutôt que la structure de la séquence consensus (Bernhart *et al.*, 2008). Cette analyse n'est cependant pas présentée car les résultats étaient presque identiques à l'analyse précédente.

Représentations circulaires (Circos)

Afin d'illustrer l'abondance et la distribution des DUS, dRS3 et NTS, les génomes de *N. meningitidis* 8013, *N. gonorrhoeae* FA1090 ainsi que de *N. elongata* subsp. *glycolytica* ATCC 29315 ont été extraits de NCBI. Chaque élément répété a été annoté séparément sur ces génomes à l'aide du serveur Benchling (<https://benchling.com/>). Enfin, les représentations graphiques circulaires ont été générées avec l'outil ShinyCircos-V2.0 en ajoutant chaque jeu de données sous forme de rectangles continus de couleurs différentes, rassemblés sur un unique chromosome (Yu *et al.*, 2018).

Recherche de motifs (MEME)

Pour la recherche non ciblée de motifs répétés associés au gène *ssnA*, c'est-à-dire sans fournir de séquence requête, le serveur MEME v5.5.1 a été utilisé (Bailey & Elkan, 1994; Bailey *et al.*, 2015). Parmi 75 espèces de *Neisseriaceae* analysées, 19 encodent le gène *ssnA*. Les régions correspondant à 5000 pb en amont et en aval de ce gène ont été extraites puis soumises au serveur MEME avec l'algorithme *any number of repetitions (anr)*. Deux motifs ont été recherchés avec une contrainte de taille de 10 à 40 pb, et ce dans les deux brins d'ADN. Pour identifier la NTS_{var}, les cinq génomes ne contenant aucune NTS classique autour de *ssnA* ont été soumis à une analyse indépendante similaire, cette fois uniquement sur les 1000 pb en 5' et 3' du gène.

1.2 Génomes analysés et phylogénie des *Neisseriaceae*

Un total de 75 génomes de *Neisseriaceae* a été utilisé pour générer une phylogénie et chercher les séquences répétées. Cette liste, composée d'un génome représentatif par espèce, combine la base de données NCBI à plusieurs génomes séquencés par notre équipe (Nyongesa *et al.*, 2022b). La phylogénie basée sur le génome noyau des *Neisseriaceae* a été générée de façon similaire à une publication récente de notre groupe (Nyongesa *et al.*, 2022b). Brièvement, l'outil Roary avec l'algorithme d'alignement multiple MAFFT a servi à l'identification de gènes orthologues présents dans 80% des espèces et possédant au moins 50% d'identité, totalisant 401 gènes. Une phylogénie par maximum de vraisemblance a ensuite été inférée à partir de ce génome noyau à l'aide de l'outil IQ-TREE avec 10 000 itérations (Nguyen *et al.*, 2015). L'arbre phylogénétique a été tracé à l'aide de FigTree v1.4.4, avec *Crenobacter* comme racine (<http://tree.bio.ed.ac.uk/software/figtree/>).

1.3 Analyses des homologues de SsnA

Le gène *ssnA* est annoté comme NMV_0044, NMB0046 et Ngo_1911 dans *N. meningitidis* 8013, *N. meningitidis* MC58 et *N. gonorrhoeae* FA1090 respectivement. Les homologues de SsnA ont été identifiés par BLASTP de la séquence protéique de *N. meningitidis* sur la base de données du NCBI selon les critères suivants : (1) taille entre 80 et 150 acides aminés, (2) au moins 50% d'identité avec la protéine de référence.

Synténie de *ssnA* chez les *Neisseria*

La représentation graphique de la synténie des régions abritant *ssnA* chez les *Neisseria* a été générée à l'aide du serveur SyntTax (Oberto, 2013). La protéine de *N. meningitidis* 8013 a servi de référence pour chercher tous les homologues avec seuil d'identité minimale de 40%. Les régions centrées sur *ssnA* correspondant à 15 000 pb sont illustrées.

Alignements multiples protéiques

Les séquences protéiques correspondant à 17 homologues de SsnA provenant des *Neisseriaceae* ont été alignées par Clustal Omega, ordonnées selon leurs liens phylogénétiques par *Neighbour-joining*, et illustrées à l'aide du programme Jalview (Sievers *et al.*, 2011; Waterhouse *et al.*, 2009).

Pour comparer le domaine GIY-YIG de SsnA à ceux des autres membres de cette superfamille de nucléases, les séquences correspondant uniquement à ce domaine provenant d'un membre représentatif de chaque famille ont été extraites des bases de données Uniprot et CDDsearch (Consortium, 2019; Lu *et al.*, 2020). L'algorithme T-Coffee a été utilisé pour l'alignement multiple de ces séquences, car il identifie plus efficacement les acides aminés fortement conservés parmi des séquences très différentes (Notredame *et al.*, 2000).

Prédiction structurelle de SsnA

La structure tridimensionnelle de SsnA a d'abord été prédite avec l'outil Phyre2, puis validée avec AlphaFold lorsque cet outil a été rendu disponible (Jumper *et al.*, 2021; Kelley *et al.*, 2015). 89/94 résidus ont été modélisés avec plus de 90% de confiance selon Phyre2, donnant une structure prédite fiable. Le programme PyMOL a servi à orienter et colorer les différentes structures protéiques (DeLano, 2002).

RÉSULTATS ET DISCUSSION

1.4 Identification d'une séquence répétée analogue aux dRS3

Identifiée il y a plus de vingt ans, la dRS3 se définit comme la séquence suivante : ATTCCC[N₈]GGGAAT. Cet élément est composé d'un noyau variable flanqué de séquences palindromiques formant une structure secondaire en épingle à cheveux (tige-boucle). En 2005, une étude a proposé d'y ajouter le « sous-type Nf1 » dont le consensus est ATTCCRCCTRCGCGGRAAK. Ce sous-type, qui représente près de la moitié des copies de la dRS3, est le site d'insertion d'un phage filamentueux impliqué dans la virulence de certaines souches de *N. meningitidis* (Bille *et al.*, 2005; Kawai *et al.*, 2005).

En réanalysant cette séquence répétée avec des outils plus récents, nous avons remarqué que celle-ci est plus complexe et moins conservée qu'initialement décrite dans la littérature (**Figure 12**). En effet, lorsqu'on regarde les 759 copies retrouvées dans le génome de *N. meningitidis* 8013, on constate que les régions palindromiques qui flanquent le noyau variable sont légèrement plus longues que prévu, ce qui semble correspondre au sous-type Nf1. En élargissant notre analyse aux régions en 5' et en 3' de ces séquences répétées, nous y avons également identifié plusieurs nucléotides hautement conservés, dont une cytosine et une guanine exactement quatre nucléotides avant et après la dRS3.

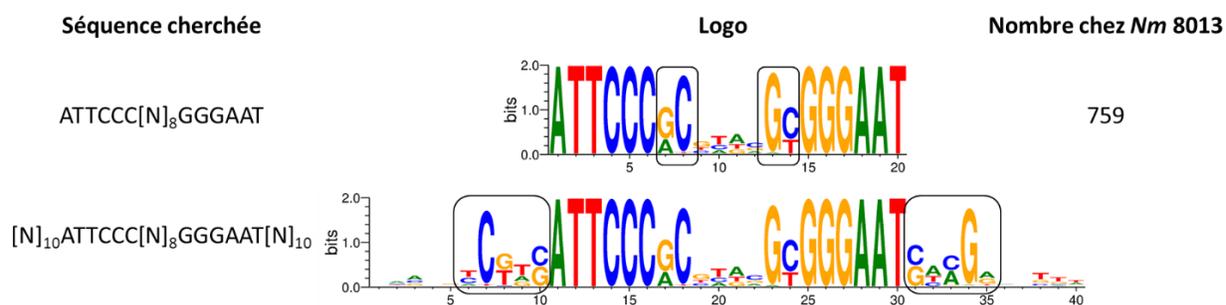


Figure 12 La dRS3 est plus longue et moins conservée qu'initialement décrite

Toutes les dRS3 (ATTCCC[N₈]GGGAAT) du génome de *N. meningitidis* 8013 ont été extraites avec l'outil fuzznuc en ne tolérant aucune substitution. L'outil Weblogo a ensuite été utilisé pour tracer les logos. La hauteur d'un nucléotide indique sa proportion relative à l'ensemble des séquences analysées. Les encadrés indiquent les extensions à la dRS3 initialement décrite.

La séquence étendue la plus représentée dans le génome de *N. meningitidis* forme une tige plus longue et nettement plus stable que la dRS3 classique en raison d'un noyau palindromique plus large qu'initialement décrit (**Figure 13A**). Cependant, les extensions conservées en 5' et en 3' ne sont pas palindromiques, ce qui implique que ces éléments répétés ont une orientation. L'existence d'une telle polarité des dRS3 n'a jamais été soulevée auparavant. En raison de ces différences majeures et d'autres découvertes décrites plus loin dans cet ouvrage, nous avons nommé cette séquence répétée étendue NTS pour *Neisseria transformation sequence*. La séquence consensus des NTS est CGTCATTCCCGCGMAVGCGGGAATCYRG, avec la région palindromique soulignée. Ainsi, toutes les NTS contiennent une dRS3 mais forment tout de même un élément répété distinct à nos yeux (**Figure 13B**). Il est à noter ici que la structure secondaire de la séquence consensus (NTS) obtenue avec l'outil RNAfold est identique à la structure consensus obtenue avec l'outil RNAalifold à partir de toutes les répétitions analysées.

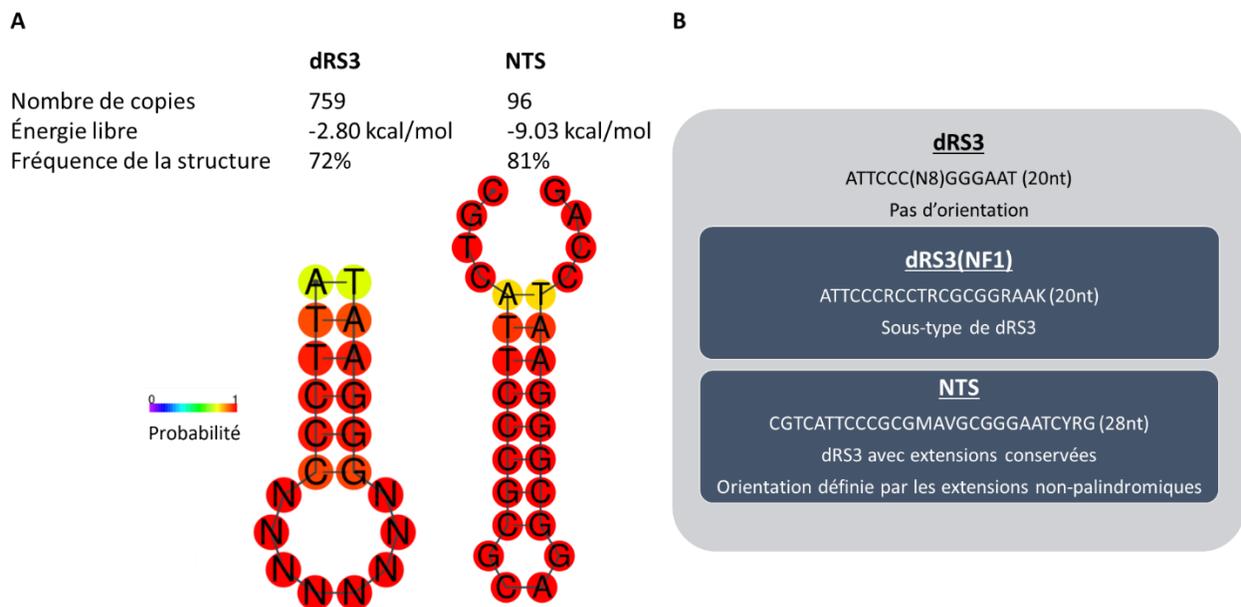


Figure 13 Différences entre les dRS3 et les NTS

L'outil RNAfold a été utilisé pour prédire les structures et leurs paramètres de stabilité. Les couleurs indiquent la probabilité de la position de chaque nucléotide. L'énergie libre de chaque structure est indiquée, de même que sa fréquence parmi toutes les structures possibles. Le nombre de copies correspond au génome de *N. meningitidis* 8013 2C4.3 sans aucune substitution autorisée.

1.5 Distribution des séquences répétées chez les *Neisseriaceae*

Nous nous sommes ensuite intéressés à la distribution de ces éléments répétés chez les *Neisseriaceae*. Les dRS3, les NTS, ainsi que les séquences répétées DUS ont été cartographiées dans l'ensemble du génome des deux espèces pathogènes ainsi que sur celui de *N. elongata*, une espèce commensale (**Figure 14**). Ces représentations graphiques ont permis d'observer plusieurs choses. D'abord, les DUS sont réparties dans l'ensemble du génome, sans différences marquantes entre les espèces pathogènes et commensales, tel que décrit dans la littérature (Frye *et al.*, 2013). À l'inverse, les dRS3 ainsi que les NTS s'accumulent à des endroits précis du génome et sont nettement plus abondantes chez les espèces pathogènes, suggérant un rôle dans la virulence. Ces clusters sont d'ailleurs systématiquement retrouvés dans des régions intergéniques. On n'observe pas de différences flagrantes entre les dRS3 et les NTS chez les espèces pathogènes, ce qui était attendu puisque ces dernières sont des séquences légèrement plus longues dérivées des dRS3. Cependant, on constate que le ratio NTS/dRS3 de *N. elongata* est plus faible que celui de *Nm* et *Ngo*, ce qui supporte davantage un potentiel rôle des NTS dans la virulence. Autre observation plus subtile : la présence de NTS dans une région donnée semble inversement proportionnelle à la présence de DUS chez les espèces pathogènes, ce qui avait déjà été démontré pour les dRS3 (Budroni *et al.*, 2011). Cette légère corrélation inverse pourrait suggérer des fonctions antagonistes à ces éléments répétés. Pour rappel, le rôle des DUS dans la promotion de l'import d'ADN par transformation naturelle est bien connu, mais on ne sait encore rien des fonctions biologiques des dRS3 et des NTS.

Aux vues de ces différences importantes d'abondance et de distribution des séquences répétées entre les *Neisseria* commensales et pathogènes, nous avons étendu nos analyses à l'ensemble des espèces de la famille *Neisseriaceae*. Pour ce faire, nous avons rassemblé un maximum de génomes d'espèces différentes, incluant celles identifiées très récemment, afin de générer la phylogénie la plus exhaustive de cette famille à ce jour, basée sur le génome noyau. Nous avons ensuite procédé à la détection et à la quantification des éléments répétés dans tous ces génomes grâce à un pipeline bio-informatique maison basé sur l'outil fuzznuc. Une synthèse de cette analyse (seulement le genre *Neisseria*) est présentée en **Figure 15**.

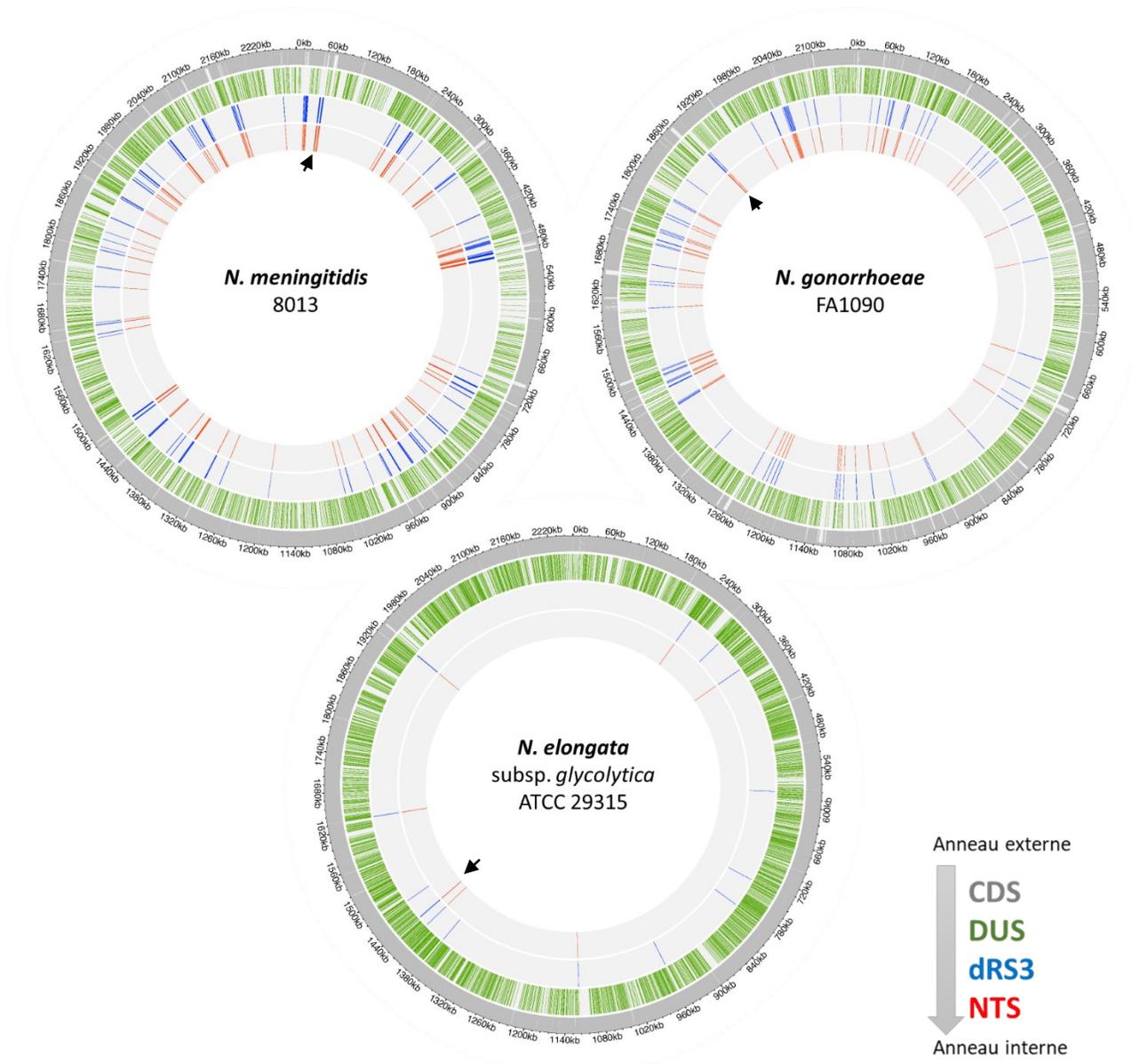


Figure 14 Cartographie des séquences répétées chez les *Neisseria*

Représentation circulaire du génome entier de trois espèces de *Neisseria*, avec la distribution de leurs gènes (CDS) et de plusieurs séquences répétées. La flèche noire pointe vers l'emplacement du gène *ssnA*. Générée avec l'outil shinyCircos V2.0 (Yu *et al.*, 2018).

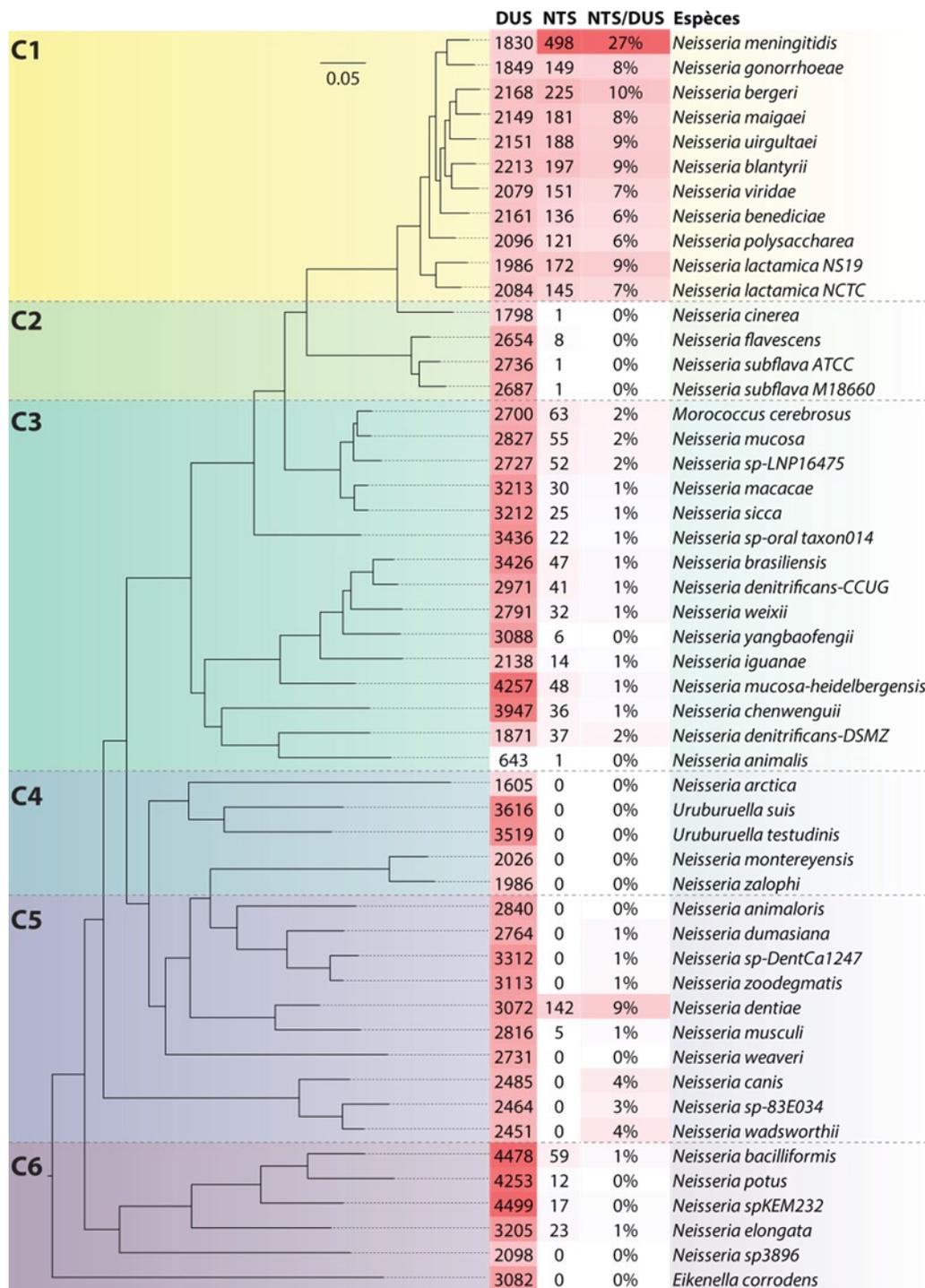


Figure 15 Distribution et abondance des éléments répétés chez les *Neisseria*

Phylogénie basée sur le génome noyau de 75 espèces de *Neisseriaceae*, associée au nombre d'éléments répétés de chacune. Un gradient rouge indique la fréquence relative de chaque répétition par rapport à l'ensemble des espèces. Des cadrans de couleurs soulignent les clades d'espèces affichant des profils différents d'éléments répétés (C1-C6). Aucune substitution n'a été permise pour la détection des DUS, tandis qu'une substitution a été autorisée pour les NTS. Puisque plusieurs variantes de DUS sont associées aux différentes espèces, la somme de toutes les variantes décrites dans la littérature est présentée.

La première observation découlant de ces analyses est que la distribution phylogénétique des *Neisseriaceae* corrèle fortement avec l'abondance et la distribution des différentes séquences répétées, indiquant que ces éléments pourraient être utilisés comme indicateurs de l'évolution et de l'adaptation de cette famille. Le groupe C1, comprenant les deux espèces pathogènes ainsi que le clade prépathogène, détient la plus forte abondance de NTS de toutes les *Neisseriaceae*, avec un ratio NTS/DUS dépassant les 5%. Parmi ce groupe, *N. meningitidis* possède près du triple de NTS comparée aux espèces les plus proches, soit plus de dix fois plus que la majorité des espèces commensales plus éloignées. Malgré sa forte similarité génétique avec le méningocoque, *N. gonorrhoeae* ne se distingue pas des autres membres du groupe C1. Il semble donc y avoir eu une expansion majeure des NTS spécifiquement chez le méningocoque, qui reflète possiblement un avantage évolutif.

De façon générale, la fréquence de NTS dans un génome diminue par paliers plus on s'éloigne des espèces pathogènes. En effet, le groupe C3 affiche quelques dizaines de NTS par génome, tandis que les groupes C4 et C5 n'en possèdent que quelques copies ou pas du tout. Les espèces du groupe C2, soit *N. cinerea*, *N. subflava* et *N. flavescens*, possèdent un nombre de NTS nettement inférieur comparé à leurs voisins phylogénétiques, indiquant possiblement une réduction au travers de l'évolution de ces espèces. À l'inverse, une expansion évolutive des NTS semble être observée chez *N. dentiae*. Cette dernière possède en effet près de 150 copies de la NTS, soit autant que les espèces du groupe C1, alors que ses voisins n'en possèdent aucune. *N. animalis*, quant à elle, abrite très peu de NTS, mais aussi très peu de DUS pour des raisons inconnues. Cette dissonance n'est pas expliquée par la taille du génome ni sa composition en gènes, et suggère donc des différences marquées au niveau des régions intergéniques de cette espèce.

Si on regarde plutôt du côté des DUS, on constate que leur nombre de copies est généralement plus élevé chez les groupes C3-6 que chez le groupe C1, soit l'inverse des NTS. Le groupe C6, dont *N. bacilliformis*, *N. potus* et *N. spKEM232*, exhibe un nombre particulièrement élevé de DUS.

Pour résumer toutes ces observations, la NTS est vraisemblablement apparue dans un ancêtre commun des *Neisseria*, après la divergence des autres genres de la famille des *Neisseriaceae*. Probablement associée à des conséquences évolutives, une expansion de cet élément répété est observée dans le clade prépathogène des *Neisseria*, et davantage encore chez *N. meningitidis*. Quelques espèces, comme *N. cinerea*, semblent au contraire avoir réduit leur nombre de NTS au travers de leur évolution. L'analyse des différents éléments répétés supporte leur utilité dans la classification phylogénétique des *Neisseriaceae*. En effet, certaines espèces

comme *Morococcus cerebrosus*, *Kingella potus* (ici *Neisseria potus*) ainsi que les *Uruburuella* mériteraient d'être reclassifiées comme *Neisseria* d'après leur profil d'éléments répétés. Au moment d'écrire cette thèse, un article en préimpression a été rendu public soulevant exactement ces mêmes points uniquement à partir de l'analyse des DUS (Boutroux *et al.*, 2023). Nos analyses révèlent d'ailleurs plusieurs informations pertinentes jamais répertoriées concernant les différents variants de DUS au sein des *Neisseriaceae*, mais ces résultats sortent du contexte de ce projet de thèse.

1.6 Identification d'un gène associé aux NTS (*ssnA*)

L'abondance et la distribution particulière des NTS chez les *Neisseria* suggèrent que ces éléments répétés sont étroitement liés à l'évolution de ce genre bactérien, et donc qu'ils pourraient être impliqués dans certains processus biologiques comme la virulence de *N. meningitidis*. Cependant, les NTS sont exclusivement retrouvées dans les régions intergéniques et à distance des promoteurs. Pour accomplir des fonctions particulières, ces séquences non codantes doivent donc impérativement interagir avec d'autres effecteurs, comme les protéines de l'hôte.

Dans le but d'identifier des candidats protéiques pouvant potentiellement interagir avec les NTS, notre équipe a réalisé des analyses bio-informatiques préliminaires. Brièvement, la présence de NTS a été vérifiée par BLAST dans une banque de génomes bactériens. Ensuite, un pipeline d'analyse maison dérivé de l'outil mycoHIT a été utilisé pour identifier des protéines exclusivement présentes chez les génomes contenant des NTS, et donc absentes des génomes sans NTS (Veyrier *et al.*, 2009). Ces analyses ont révélé quelques candidats, dont plusieurs protéines hypothétiques (données non montrées, archives du laboratoire). Parmi ces candidats, un seul gène s'est fortement démarqué puisqu'il est presque systématiquement flanqué d'au moins deux NTS (voir **Figure 11** dans la section Hypothèse). N'ayant aucun homologue caractérisé, nous avons appelé ce gène et sa protéine *ssnA*/SsnA.

1.7 Association *ssnA*-NTS

Dans l'optique de confirmer et de mieux caractériser cette association entre *ssnA* et la NTS, nous avons étendu nos recherches à une multitude de génomes, autant chez les *Neisseria* que chez des espèces très éloignées.

1.7.1 Association *ssnA*-NTS chez les *Neisseria*

Chez les *Neisseria* d'abord, nous avons utilisé une approche systématique consistant à identifier les espèces encodant le gène *ssnA*, à extraire les séquences génomiques en amont et en aval de ce gène (5000 bp de chaque côté), puis à détecter et cartographier de façon automatique les motifs nucléotidiques répétés. 19 espèces sur 75 *Neisseriaceae* analysées encodent pour *ssnA*, provenant des groupes C1, C3 et C4 (**Figure 15**). Dans les 19 génomes analysés, les deux motifs retrouvés les plus fréquemment correspondent aux DUS (454 copies) ainsi qu'aux NTS (127 copies) (**Figure 16**). En s'attardant à la localisation de ces motifs dans chaque séquence, on observe que les DUS sont distribuées sur toute leur longueur, tandis que les NTS s'amassent de façon marquée à proximité immédiate de *ssnA* (**Figure 16A**). Par contre, ces régions riches en NTS semblent extrêmement variables puisque le nombre de NTS et leur disposition sont différents entre chaque espèce. Autre observation frappante, le motif identifié automatiquement par l'outil MEME correspond bien à la NTS et non simplement à la dRS3 initialement décrite, composée uniquement du palindrome ATTCCC[N₈]GGGAAT (**Figure 12, Figure 16B**). En effet, la quasi-totalité de ces séquences affiche l'extension CGTC en 5' et C en 3' comparée à la dRS3, en plus d'un noyau palindromique plus étendu et mieux conservé. Ensemble, ces résultats confirment davantage l'association entre *ssnA* et les NTS chez les *Neisseria*, en plus de supporter l'hypothèse que les extensions qui distinguent les NTS des dRS3 sont probablement essentielles aux rôles encore inconnus de ces éléments répétés. La région entourant *ssnA* est hypervariable, suggérant que de nombreux réarrangements chromosomiques y ont lieu, possiblement par l'intermédiaire de ces éléments répétés.

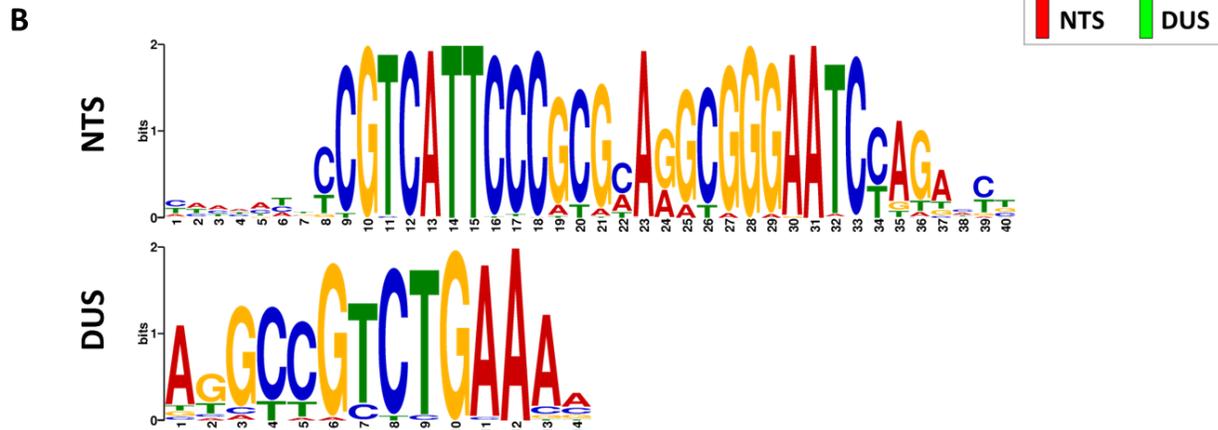
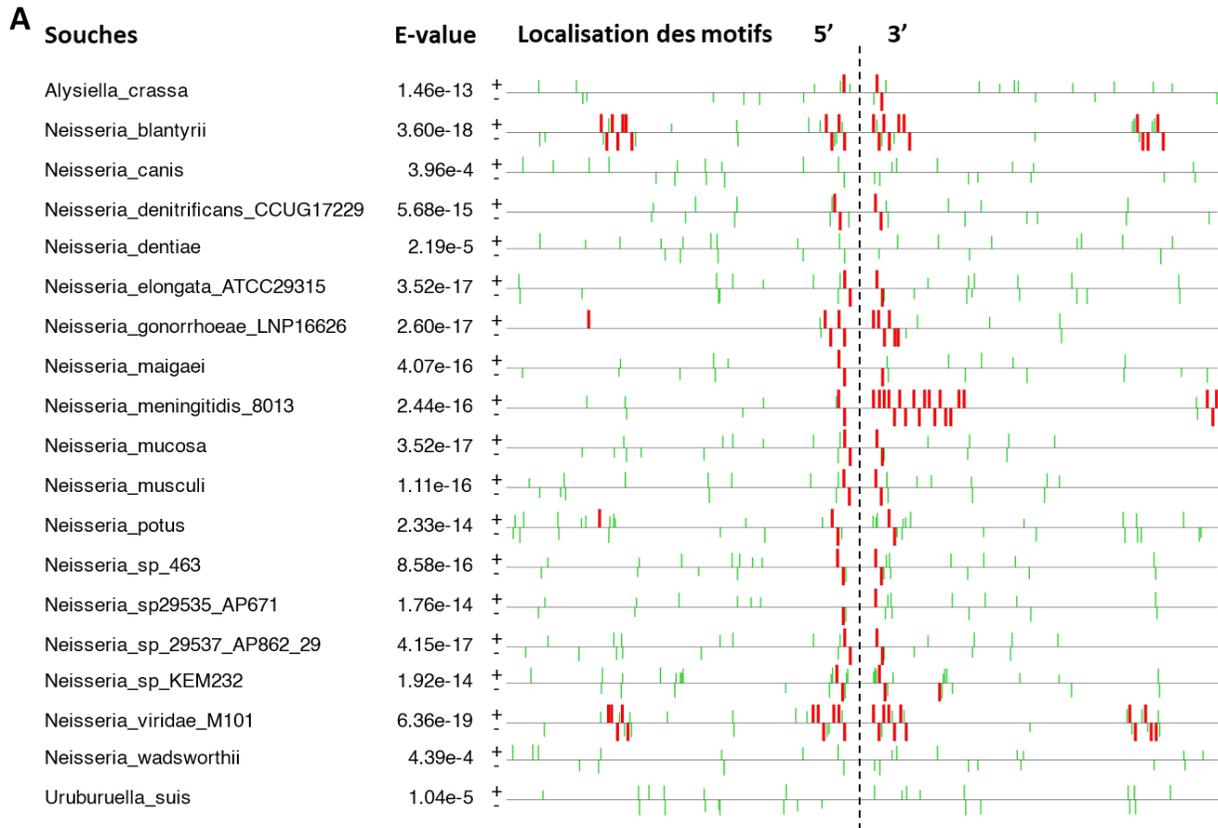


Figure 16 Analyse des motifs répétés à proximité de *ssnA* chez les *Neisseria*

Les régions correspondant à 5000 pb en amont et en aval de *ssnA* chez 19 souches de *Neisseria* ont été soumises à une recherche de motifs nucléotidiques avec l'outil MEME. **A.** Représentation graphique de la localisation des deux motifs identifiés (NTS en rouge, DUS en vert). La ligne pointillée représente l'emplacement du gène *ssnA*. Les E-values représentent la probabilité de détecter le même nombre de motifs similaires dans un ensemble de séquences aléatoires. **B.** Logos résultant de l'alignement de l'ensemble des motifs analysés en A, soit 454 DUS et 127 NTS. La hauteur de chaque nucléotide est proportionnelle à sa fréquence parmi l'ensemble des répétitions analysées.

1.7.2 Identification d'une variante de la NTS chez certaines *Neisseria*

Selon la précédente analyse, seules 4 espèces sur 23 ne possèdent aucune NTS autour de *ssnA*, soit *N. canis*, *N. dentiae*, *N. wadsworthii* et *Uruburuella suis* (voir **Figure 16A**). Pour trois d'entre elles, ceci était attendu puisque leur génome entier ne contient aucune copie de ces éléments répétés (voir **Figure 15**). Nous nous sommes alors demandé si leur gène *ssnA* pouvait être associé à une autre séquence répétée jusque-là inconnue. Ces espèces ont donc été soumises séparément à une analyse de motifs similaire à la précédente (**Figure 17**). En plus des DUS, un tout nouveau motif que l'on a appelé NTS_{var} a été identifié chez toutes ces espèces à l'exception de *U. suis*. Tout comme les NTS, ce motif est systématiquement présent immédiatement devant et derrière *ssnA* (**Figure 17A**). Bien que cet élément répété présente plusieurs similarités avec la NTS, dont le tétranucléotide « CGTC » en 5' et une longue région palindromique, sa séquence diffère considérablement de cette dernière (**Figure 17B, Figure 18A**).

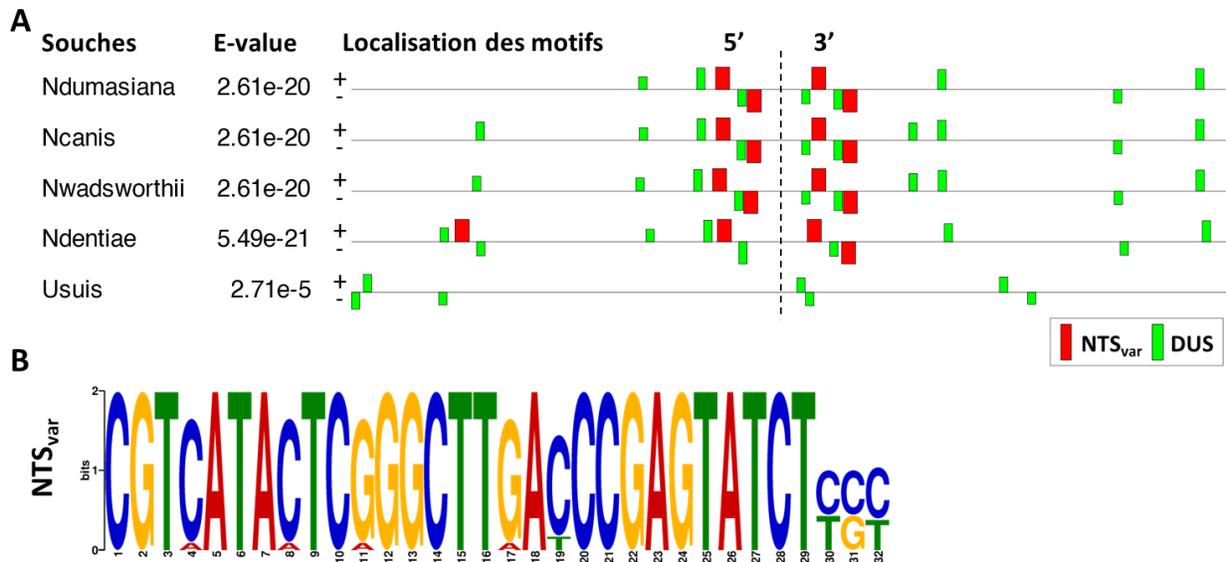


Figure 17 Analyse de nouveaux motifs répétés à proximité de *ssnA*

Les régions correspondant à 1000 bp en amont et en aval de *ssnA* de cinq espèces ont été soumises à une recherche de motifs par l'outil MEME. **A.** Localisation des deux motifs identifiés (NTS_{var} en rouge, DUS en vert). La ligne pointillée représente l'emplacement du gène *ssnA*. Les E-values représentent la probabilité de détecter le même nombre de motifs similaires dans un ensemble de séquences aléatoires. **B.** Logos résultant de l'alignement de l'ensemble des motifs analysés en A, soit 44 DUS et 16 NTS.

Intrigués par cette découverte, nous avons ensuite comparé cette variante de la NTS à la séquence d'origine et évalué son abondance dans plusieurs espèces d'intérêt (**Figure 18**). Malgré sa taille plus longue et sa séquence palindromique différente, la NTS_{var} partage une structure similaire à celle de la NTS, renforçant l'hypothèse qu'il s'agit bien d'un variant et non d'un élément répété complètement distinct. De plus, ce variant est présent en plusieurs dizaines

de copies dans les génomes de *N. dumasiana*, *N. dentiae*, *N. wadsworthii*, et *N. canis*, alors que trois de ces espèces n'abritent aucune NTS classique. À l'inverse, *N. meningitidis* ne contient pas une seule copie de la NTS_{var}. Ces données indiquent que la NTS et la NTS_{var} ont divergé au cours de l'évolution des *Neisseria*, tout en restant associées au gène *ssnA*. *N. cinerea*, qui n'encode pas ce gène, ne possède qu'une seule NTS et aucune NTS_{var}, supportant cette hypothèse. *U. suis* quant à elle, encode *ssnA* mais ne contient aucune NTS ni NTS_{var}, ce qui suggère l'existence d'autres variants encore non-identifiés ou bien marque une acquisition récente du gène *ssnA*. Ainsi, les NTS se déclinent en plusieurs variants distribués selon la phylogénie des *Neisseria*.

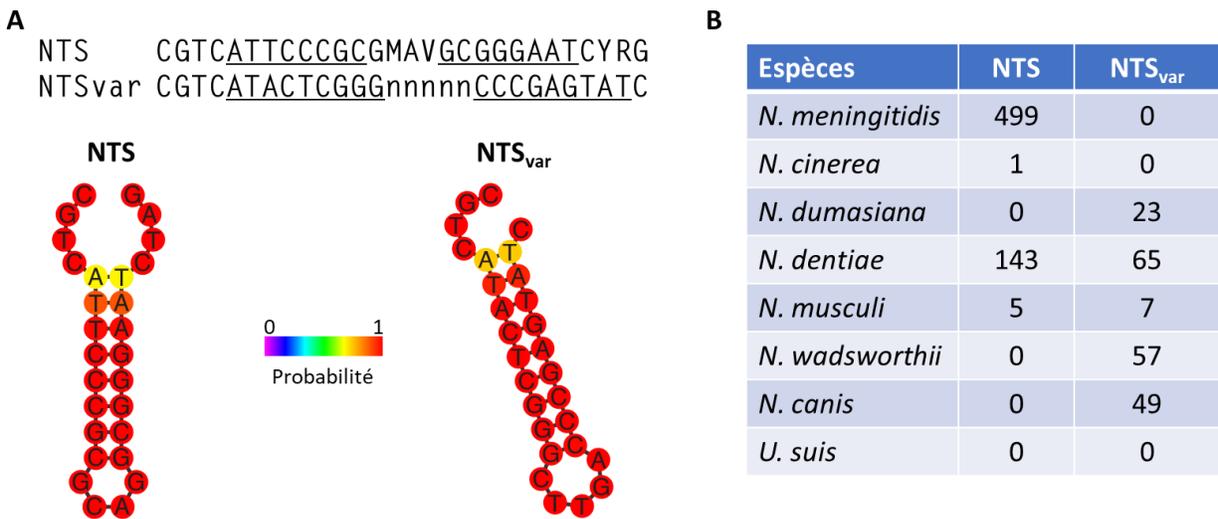


Figure 18 Comparaison entre la NTS et la NTS_{var}

A. Séquence consensus des NTS et NTS_{var} associées à leur structure secondaire prédite. Les nucléotides soulignés indiquent le palindrome formant la tige de la structure en épingle à cheveux. La couleur des nucléotides indique la probabilité de leur position dans la structure, prédite avec l'outil RNAalifold. **B.** Quantifications de séquences présentées en A dans le génome entier de plusieurs *Neisseria*, sans aucune substitution permise.

1.7.3 SsnA chez les *Neisseria*

Toujours chez les *Neisseria*, nous nous sommes ensuite intéressés au contexte génomique de *ssnA*. Nos analyses précédentes montrent que ces gènes se trouvent dans des régions hypervariables riches en NTS ou NTS_{var}, dont le nombre et l'abondance fluctuent beaucoup. Nous avons donc réalisé une analyse de synténie dans le but de déterminer si ces gènes sont encodés au même endroit du génome dans les différentes *Neisseria* (**Figure 19**). Cette analyse montre que la localisation de *ssnA* est conservée au sein d'une même espèce, mais diverge entre les espèces. Les différents groupes observés suivent d'ailleurs la phylogénie des *Neisseria*, confirmant l'acquisition de *ssnA* par un ancêtre commun. Entre les souches de *N. meningitidis* et

diverge des autres espèces abritant la NTS classique. En parallèle, les espèces du clade prépathogène partagent une protéine quasi identique, mais différente des espèces commensales. Ces observations sont en parfaite adéquation avec les analyses précédentes, et confirment une coévolution entre le gène *ssnA* et les séquences répétées propres à chaque espèce, supportant davantage l'interaction prédite entre ces éléments.

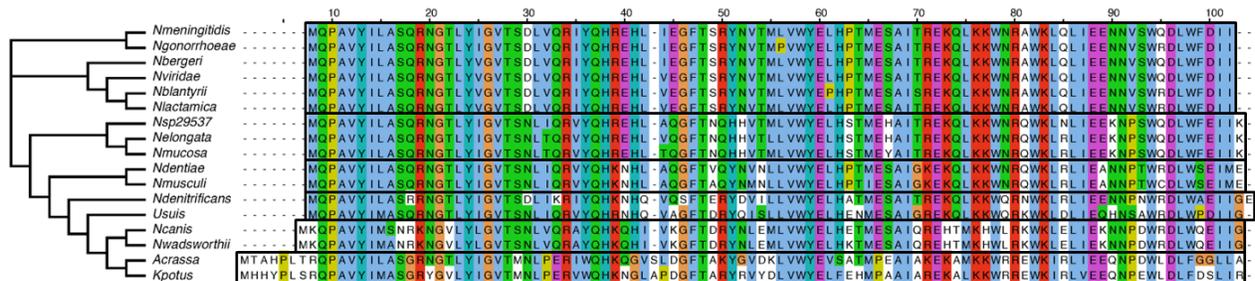


Figure 20 Alignement multiple des protéines *SsnA* chez les *Neisseria*
 Les séquences protéiques correspondant à *ssnA* ont été alignées par Clustal Omega et colorées selon la conservation de chaque acide aminé. Un cladogramme généré par *neighbour joining* illustre les liens entre ces protéines, dont les groupes distincts sont encadrés.

1.7.4 Association *ssnA*-NTS chez les non-*Neisseria*

Qu'en est-il de l'association *ssnA*-NTS en dehors de la famille des *Neisseriaceae* ? Pour répondre à cette question, nous avons identifié des homologues de *ssnA* dans plusieurs espèces très diverses, et nous avons vérifié la présence et la fréquence de NTS à proximité immédiate de ces gènes (**Tableau 4**). Chez la majorité des espèces analysées, on retrouve une ou plusieurs copies de la NTS immédiatement en amont ou en aval de *ssnA*. Dans plusieurs cas, ces NTS associées à *ssnA* sont les seules copies retrouvées dans l'ensemble du génome. Cette association est davantage illustrée par le facteur d'enrichissement en NTS autour de *ssnA*, qui montre que la région intergénique flanquant ce gène contient souvent plusieurs milliers de fois plus de NTS qu'attendu si ces répétitions étaient distribuées aléatoirement dans le génome. Comme contrôle, une analyse identique a été réalisée pour le gène essentiel et universel *gyrA*, révélant que des NTS se trouvent très rarement à proximité de celui-ci, contrairement au gène *ssnA*. Ces chiffres démontrent que l'association *ssnA*-NTS n'est pas limitée aux *Neisseria*, suggérant une interaction et des fonctions biologiques qui s'étendent à l'ensemble du domaine bactérien. Vu l'existence de variantes de la NTS au sein même des *Neisseria*, il est possible que les homologues de *ssnA* ne possédant pas de NTS à proximité soient associés à d'autres variantes non détectées dans cette analyse. Cette hypothèse est explorée dans le quatrième chapitre de cette thèse.

Tableau 4 Détection de NTS à proximité des homologues *ssnA*

Espèces	Phylum/ Classe	NTS dans le génom entier	NTS associées à <i>gyrA</i> *	NTS associées à <i>ssnA</i> *	Facteur d'enrichissement **
<i>Acinetobacter populi</i>	γ	3	0	2	2458
<i>Ahniella affigens</i>	γ	1	0	0	0
<i>Candidatus Nitrospira inopinata</i>	N	388	0	1	8
<i>Dechloromonas denitrificans</i>	β	4	0	4	4238
<i>Dyella japonica A8</i>	γ	41	0	4	465
<i>Hyphomicrobium denitrificans</i> <i>1NES1</i>	α	23	0	0	0
<i>Luteibacter anthropi</i>	γ	3	0	3	4962
<i>Lysobacter capsici</i>	γ	309	2	2	40
<i>Lysobacter soli</i>	γ	8	0	2	989
<i>Nitrosomonas mobilis</i>	β	2	0	1	1542
<i>Nocardia terpenica</i>	A	0	0	0	/
<i>Calycomorphotria hydatis V22</i>	P	3	0	3	5163
<i>Pseudomonas mendocina</i>	γ	3	0	3	5228
<i>Pseudoxanthomonas spadix</i>	γ	419	0	2	16
<i>Rhodanobacter denitrificans</i>	γ	1	0	0	0
<i>Shewanella benthica</i>	γ	164	0	0	0
<i>Thiohalobacter thiocyanaticus</i>	γ	246	2	3	39

γ Gammaprotéobacteria, β Betaproteobacteria, α Alphaproteobacteria, N Nitrospirota, A Actinomycetota, P Planctomycetota

*Nombre de NTS situées dans les régions intergéniques immédiatement en amont et en aval du gène *gyrA* ou *ssnA*.

**Ratio entre le nombre réel de NTS associées à *ssnA* et la fréquence théorique de NTS dans une région génomique de même taille si ces répétitions étaient distribuées aléatoirement dans le génome. Représente l'enrichissement en NTS dans la région contenant *ssnA*.

1.8 SsnA est une nucléase hypothétique de la superfamille GIY-YIG

Le gène *ssnA* encode pour une nucléase hypothétique en raison de son unique domaine fonctionnel appartenant à la superfamille d'endonucléases GIY-YIG. Pour confirmer cette classification, nous avons extrait puis aligné les séquences correspondant uniquement aux domaines GIY-YIG de membres représentatifs de cette superfamille (**Figure 21A**). Comme la majorité des membres de cette famille, SsnA possède la tyrosine du motif « GIY », la seconde partie du motif « YIG » ainsi que les sites actifs conservés représentés par les acides aminés Arginine (R), Histidine (H) et Asparagine (N), en plus d'un acide glutamique (E) prédit comme étant un site de liaison au magnésium. D'après la phylogénie inférée à partir de l'alignement multiple, le domaine GIY-YIG de SsnA se rapproche de celui de MSH1, une sous-famille d'enzymes importantes à la maintenance du génome mitochondrial et chloroplastique interagissant avec des structures d'ADN embranchées (Fukui *et al.*, 2018). Cependant, on constate une très faible similarité de séquence entre SsnA et les autres GIY-YIG, suggérant qu'elle n'appartient pas aux familles décrites.

Puisque la structure tertiaire des protéines GIY-YIG est mieux conservée que leur séquence en acides aminés, nous avons également comparé SsnA aux membres de cette superfamille selon ce critère (**Figure 21B**). En absence de structure expérimentale, nous nous sommes contentés d'une prédiction *in silico* pour SsnA, qui s'est révélée être particulièrement fiable en raison de sa petite taille et de sa simplicité. Tout comme les autres protéines GIY-YIG, SsnA se replie en un noyau formé de trois feuillets bêta antiparallèles ceinturé de plusieurs hélices alpha. SsnA appartient donc sans aucun doute à la superfamille des GIY-YIG, mais sa séquence considérablement différente des familles déjà caractérisées indique qu'elle appartient potentiellement à une nouvelle famille distincte.

Une recherche de la protéine SsnA dans la base de données des domaines conservés du NCBI (CDD) supporte ces conclusions (Lu *et al.*, 2020). En effet, celle-ci est classée dans la famille cd10448 GIY-YIG_unchar_3, composée de 23 000 protéines dont la grande majorité sont hypothétiques et d'origine bactérienne. Conformément à nos analyses, cette famille est distincte des autres familles GIY-YIG et n'a pas encore été caractérisée (**Figure 22**).

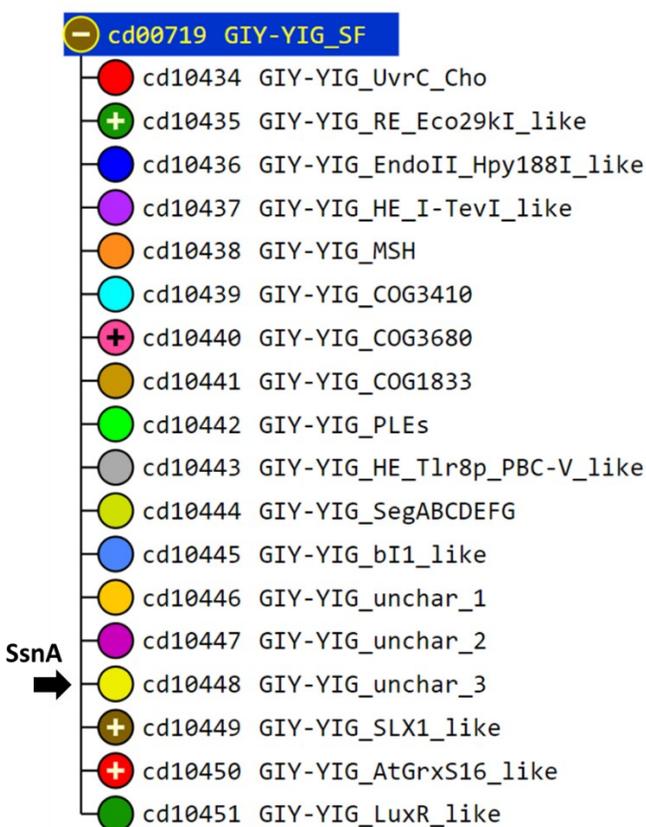


Figure 22 Hiérarchie de la superfamille GIY-YIG selon la base de données CDD
Sous-familles définies par NCBI selon leur structure 3D, leurs motifs conservés et leur organisation phylogénétique.

CONCLUSIONS

Pour résumer brièvement ce premier chapitre, nous avons identifié chez les *Neisseria* un élément répété, appelé NTS, qui ressemble beaucoup aux séquences dRS3 décrites dans la littérature mais dont les fonctions et les interacteurs protéiques sont encore méconnus. Les NTS se distinguent des dRS3 par une séquence palindromique plus longue, précédée et suivie de quelques nucléotides non palindromiques hautement conservés, qui confèrent à ces éléments répétés une orientation. L'association par proximité génomique entre les NTS et le gène *ssnA* a été confirmée *in silico* chez les *Neisseria* mais aussi chez des espèces distantes encodant pour des protéines hypothétiques homologues. De plus, nous avons identifié une variante de la NTS exclusive à certaines espèces de *Neisseria*, qui semble découler d'une coévolution avec leur protéine SsnA. Enfin, nous avons validé l'appartenance de SsnA à la superfamille des endonucléases GIY-YIG, mais sa séquence diverge considérablement des autres familles caractérisées, la classant dans une famille distincte encore mal définie.

2 CHAPITRE II – CARACTÉRISATION BIOCHIMIQUE DE SsnA

SsnA possède comme seul domaine fonctionnel le domaine GIY-YIG, la catégorisant comme une endonucléase hypothétique. Ce second chapitre détaille l'étude fonctionnelle de SsnA d'un point de vue enzymatique. L'objectif de ce dernier est de confirmer qu'il s'agit bien là d'une nucléase, d'en déterminer la spécificité et l'affinité pour son substrat, mais aussi de vérifier expérimentalement son interaction prédite avec les NTS.

MATÉRIEL ET MÉTHODES

2.1 Souches de *E. coli* et conditions de culture

Escherichia coli DH5 α a été utilisée pour les clonages, tandis que la souche BL21(DE3) a servi de vecteur d'expression pour les purifications de protéine. Les souches d'expression sont détaillées dans le **Tableau 5**. Ces souches ont été rendues compétentes par traitement chimique avec un tampon au calcium, et transformées selon un protocole standard (Green & Sambrook, 2018). *E. coli* a été cultivée sur milieu Lysogeny Broth Miller (Fisher) à 37°C. Lors des cultures liquides, les flasques et tubes à essai ont été incubés avec agitation à 250 rpm. Lorsque nécessaire, des antibiotiques ont été ajoutés aux milieux aux concentrations indiquées dans le **Tableau 6**.

Tableau 5 Souches d'expression utilisées pour les purifications de protéine

Souches d'expression	Description
<i>E. coli</i> BL21(DE3)	Vecteur d'expression avec polymérase T7 inducible à l'IPTG
<i>E. coli</i> BL21(DE3) <i>ssnA</i> _Nm <i>ssnA</i> _WT <i>ssnA</i> _Y6A <i>ssnA</i> _Y17A <i>ssnA</i> _E64A	Pour l'expression de SsnA sauvage et mutée de <i>N. meningitidis</i>
<i>E. coli</i> BL21(DE3) <i>ssnA</i> _Nelon	Pour l'expression de SsnA (WT) de <i>N. elongata</i>
<i>E. coli</i> BL21(DE3) <i>ssnA</i> _Nwad	Pour l'expression de SsnA (WT) de <i>N. wadsworthii</i> (Chapitre IV)
<i>E. coli</i> BL21(DE3) <i>ssnA</i> _Dlykan	Pour l'expression de SsnA (WT) de <i>D. lykanthroporeppelens</i> (Chapitre IV)
<i>E. coli</i> BL21(DE3) <i>ssnA</i> _Vvulni	Pour l'expression de SsnA (WT) de <i>V. vulnificus</i> (Chapitre IV)

Tableau 6 Concentration en antibiotiques des milieux de culture

Antibiotique	Concentration (µg/ml)	Antibiotique	Concentration (µg/ml)
Ampicilline	<i>Ec</i> : 100	Streptomycine (Sm)	<i>Nm</i> : 200
Érythromycine (Ery)	<i>Ec</i> : 300 <i>Nm</i> : 3	Spectinomycine (Sp)	<i>Ec</i> : 50 <i>Nm</i> : 75
Kanamycine (Km)	<i>Ec</i> : 50 <i>Nm</i> : 100	Acide nalidixique (Nal)	<i>Nm</i> : 5
Chloramphénicol (Cm)	<i>Ec</i> : 25 <i>Nm, Nm_{us}</i> : 5		

Ec : *E. coli*, *Nm* : *N. meningitidis*, *Nmus* : *N. musculi*

2.2 Clonage, expression et purification de SsnA

Le **Tableau 7** détaille les plasmides d'expression utilisés dans cette étude pour produire les différentes protéines analysées.

Tableau 7 Liste des plasmides d'expression utilisés dans cette étude

Plasmides	Description	Source
pET15-MHL	Vecteur d'expression avec un his-tag en N-terminal	Stock du labo
pET-15b(+)	Vecteur d'expression avec un his-tag en N-terminal	Bio Basic
pET15-MHL_NMV0044	Pour purification de His-SsnA WT de <i>N. meningitidis</i>	Cette étude
pET15-MHL_NMV0044_Y6A	Pour purification de His-SsnA Y6A de <i>N. meningitidis</i>	Cette étude
pET15-MHL_NMV0044_Y17A	Pour purification de His-SsnA Y17A de <i>N. meningitidis</i>	Cette étude
pET15-MHL_NMV0044_E64A	Pour purification de His-SsnA E64A de <i>N. meningitidis</i>	Cette étude
pGEX-4T-1	Vecteur d'expression avec un GST-tag en N-terminal	Amersham
pGEX_NMV0044	Pour purification de GST-SsnA WT de <i>N. meningitidis</i>	Cette étude
pET-15b(+)_ssnA_Nelon	Pour purification de His-SsnA WT de <i>N. elongata</i>	(Bio Basic)
pET-15b(+)_ssnA_Nwad	Pour purification de His-SsnA WT de <i>N. wadsworthii</i>	(GenScript)
pET-15b(+)_ssnA_Dlykan	Pour purification de His-SsnA WT de <i>D. lykanthroporeppelens</i>	(GenScript)
pET-15b(+)_ssnA_Vvulni	Pour purification de His-SsnA WT de <i>V. vulnificus</i>	(GenScript)

Les sources entre parenthèses indiquent les constructions qui ont été synthétisées pour cette étude, autrement elles ont été générées par ligation de produits de PCR des gènes d'intérêt dans les vecteurs d'expression.

2.2.1 Génération des plasmides d'expression

Le gène *NMV_0044* de *N. meningitidis* 8013 a été amplifié par PCR, puis cloné dans le vecteur d'expression pET15-MHL dans les sites de restriction NdeI et XhoI pour générer le plasmide pET15-MHL_NMV0044. Ce vecteur a été choisi car il permet d'obtenir des protéines recombinantes avec un tag polyhistidine (6x-his) en N-terminal. Mesurant 18 aa, ce tag est l'un des plus petits sur le marché, ce qui est souhaitable afin de minimiser le risque de perturber l'activité enzymatique de SsnA, qui est particulièrement petite (96 aa). Pour générer les protéines mutées SsnA Y6A, Y17A et E64A, une mutagenèse ciblée a été réalisée par OE-PCR (*overlap extension PCR*) avec des amorces introduisant les mutations souhaitées. Les deux premières mutations ciblent les tyrosines hautement conservées des motifs GIY et YIG, tandis que la troisième mutation cible l'acide glutamique prédit comme étant le site de liaison d'un cofacteur métallique. Ces substitutions sont reconnues dans la littérature comme abolissant complètement l'activité nucléase de plusieurs autres protéines de la superfamille des GIY-YIG (Ibryashkina *et al.*, 2007; Kowalski *et al.*, 1999; Liu *et al.*, 2013; Mak *et al.*, 2010).

Pour générer la protéine recombinante GST-SsnA de *N. meningitidis*, le gène *NMV_0044* de *N. meningitidis* 8013 a été amplifié par PCR, puis cloné dans le vecteur pGEX-4T-1 pour générer le plasmide pGEX_NMV0044. Ce vecteur sert de contrôle supplémentaire pour garantir que la présence de tags en N-terminal n'affecte pas l'activité catalytique de SsnA. En effet, le tag GST mesure 226 aa, soit plus du double de la taille de SsnA.

Pour générer la protéine recombinante His-SsnA de l'espèce commensale *N. elongata*, le gène *NELON_RS07125* a directement été synthétisé dans le vecteur pET-15b(+) par BioBasic. Ce vecteur est similaire à pET15-MHL, avec l'expression du tag polyhistidine en N-terminal.

Les plasmides d'expression ont été amplifiés et vérifiés dans *E. coli* DH5 α , puis clonés dans la souche d'expression BL21(DE3). Avant de procéder aux purifications, l'expression et la solubilité de chaque protéine recombinante ont dû être optimisées en faisant varier la quantité d'IPTG servant à l'induction du promoteur ainsi que la durée et la température d'induction.

2.2.2 Expression et purification

Pour les protéines avec un his-tag, les purifications ont été faites par chromatographie d'affinité sur colonne avec une résine d'agarose couplée au nickel Ni-NTA (Thermo). Pour His-SsnA WT, 1.6 L de culture a été induit avec 1 mM d'IPTG en début de phase exponentielle ($DO_{600}=0.6-0.8$), et ce pendant 4 h à 37°C avec agitation à 250 rpm. Après centrifugation, le culot a été resuspendu

dans 40 ml de tampon de liaison (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8). Une première étape de lyse enzymatique a été faite par addition de 1 mg/ml de lysozyme (Thermo) et de 1 mM d'inhibiteur de protéases PMSF (Thermo), suivi d'une incubation sur glace de 30 min. Une seconde étape de lyse a ensuite été réalisée à l'aide d'un sonicateur Branson 250 sur glace pendant cinq rondes de 3 min à puissance 8 avec un cycle de travail de 50%. Le lysat obtenu a été clarifié par centrifugation 45 min à 15 000 rpm à 4°C. Après avoir équilibré 4 ml de résine Ni-NTA avec le tampon de liaison, celle-ci a été incubée 30 min à 4°C avec le surnageant du lysat sous faible agitation. Une fois dans la colonne, la résine chargée a été lavée successivement avec 40 ml de tampon de liaison supplémenté avec 20, 40, 80, puis 150 mM d'imidazole. Enfin, His-SsnA a été élué en plusieurs fractions de 10 ml avec 300 mM d'imidazole à pH 8.

Pour purifier His-SsnA de *N. elongata*, des cultures de 1 L ont été induites avec 0.85 mM d'IPTG pendant 22 h à 18°C avec agitation à 160 rpm. Après culottage, les cellules ont été resuspendues dans un tampon de liaison différent du précédent (10 mM Tris, 250 mM NaCl, 2 mM MgCl₂, 5% glycérol, pH 8) avant d'être lysées par sonication. Le reste du protocole a été suivi.

Pour les mutants His-SsnA Y6A, Y17A et E64A, nous avons été confrontés à d'importants problèmes de solubilité forçant la purification de la fraction insoluble en conditions dénaturantes. En effet, plus de 95% des protéines induites étaient systématiquement retrouvées dans les corps d'inclusions, et ce peu importe les conditions testées. Pour ces trois mutants, la fraction insoluble obtenue après la lyse et la centrifugation a été resuspendue dans un tampon de liaison contenant 8 M d'urée et le reste du protocole a été suivi, toujours avec de l'urée dans les tampons de lavage et d'éluion.

Pour la purification de GST-SsnA, une culture de 400 ml a été induite avec 0.8 mM d'IPTG pendant 4 h à 18°C. Le culot a été resuspendu dans 20 ml de PBS avec 1 mM de PMSF, lysé par sonication 15 x 30s à puissance 8 par cycles de 50% on/off sur glace, puis clarifié par centrifugation. La fraction soluble a ensuite été mélangée à 500 µL de résine au glutathion HiCap (Qiagen) pendant 2 heures à 4°C. La résine chargée a par la suite été lavée 3x avec du PBS, puis la protéine d'intérêt a été éluée avec une solution fraîchement préparée de 10 mM de glutathion réduit (Sigma) dans du PBS.

SDS-PAGE et Bradford

La pureté des fractions d'éluion a été analysée par SDS-PAGE sur gels de Bis-Tris 12% dans du tampon MES. Après 1h30 de migration à 120 V, les gels ont été colorés au bleu de Coomassie, puis imagés sur un appareil ImageQuant LAS 500 (GE). La taille de la protéine obtenue a été

confirmée par spectrométrie de masse MALDI-TOF (Ab Sciex) sans digestion en suivant les protocoles standards. La concentration a ensuite été calculée par test de Bradford à 595 nm avec un spectrophotomètre SmartSpec Plus (Bio-Rad) et une gamme de BSA.

Concentration et dialyse

Les fractions de His-SsnA WT de *N. meningitidis* ont été traitées par dialyse dans un tampon de stockage sans imidazole (200 mM NaCl, 50 mM Tris, 50% glycérol, pH 8), 2 x 2 h à 4°C. La protéine de *N. elongata* a quant à elle été dialysée dans 300 mM d'arginine, 300 mM (NH₄)₂SO₄, 20 mM Tris, 3 mM EDTA, pendant 16 h à température pièce, puis concentrée à l'aide d'un dispositif Amicon avec un seuil à 3 kDa (Sigma). Afin d'enlever l'urée et de replier correctement les protéines mutées His-SsnA Y6A, Y17A et E64A, celles-ci ont été dialysées 3 h dans 500 mM NaCl, 100 mM arginine, 10 mM Tris, et 5% glycérol. À titre de contrôle supplémentaire, la protéine WT a également été soumise au même protocole de purification de fraction insoluble puis de repliement. Son activité enzymatique était identique à la protéine purifiée de la fraction soluble, confirmant son repliement correct (données non montrées).

Ces différentes conditions d'expression et tampons de dialyse utilisés témoignent des importants problèmes de solubilités que nous avons rencontrés et sont le fruit d'une longue optimisation impliquant plusieurs personnes du laboratoire. En effet, même lorsque des protéines étaient purifiées avec succès de la fraction soluble, ces dernières avaient tendance à précipiter de façon drastique dès que leur tampon était changé ou lors de tentatives de concentration.

2.3 Interactions protéines-ADN

2.3.1 Oligonucléotides fluorescents

Pour étudier l'activité catalytique des différentes SsnA purifiées, de nombreux substrats d'ADN fluorescents ont été synthétisés (Sigma). La plupart de ces derniers sont marqués en 5' à la fluorescéine, leur donnant une fluorescence stable et indépendante de la taille des molécules testées. Ces oligonucléotides sont détaillés dans le **Tableau 8**. Lorsque nécessaire, des oligonucléotides complémentaires ont été hybridés pour former de l'ADN double-brin en les mélangeant à équimolarité dans un tampon d'hybridation (10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA), puis en chauffant le mélange 5 minutes à 95°C et en le laissant refroidir lentement.

Tableau 8 Oligonucléotides fluorescents

Nom	Séquence (5'-3')
RSE100_F	[6-fam]AATTCAAAAATATT <u>CCGTCATTCCCGCGAAAGCGGGAATCTAGAACGTAAAATCTAAAGAAACCGTTTTATCCGATAAGTTTCCGC-ACCGACAGACCTGG</u>
RSE100_R	[6-fam]CCAGGTCTGTCCGGTGCGGAAACTTATCGGATAAAACGGTTTTCTTTAGATTTTACGTTCTAGATTCCCGCTTTCCGGGGAATGACCG-AAATTTTTGAATT
RSE75_F	[6-fam]TTTTTATCAAATTCAAAAATATTCCGTCATTCCCGCGAAAGCGGGAATCTAGAACGTAAAATCTAAAGAAACCGT [Cy5]TTTTTATCAAATTCAAAAATATTCCGTCATTCCCGCGAAAGCGGGAATCTAGAACGTAAAATCTAAAGAAACCGT
RSE37	[6-fam]AAAAATATTCCGTCATTCCCGCGAAAGCGGGAATCTA
RSE28	[6-fam]CCGTCATTCCCGCGAAAGCGGGAATCTA
RNA37	[6-fam]AAAAUAUUCCGUCAUUCGCCGAAAGCGGGAUUCUA
RSE18-25	[6-fam]TTTTTATCAAATTCAAAA-A-T-A-T-T-C-C
NTS1	[6-fam]GAGGTTTCGGGCAACTTCTAAAC <u>CCGTCATTCCCGCGCAGGCGGGAATCCAGACCATTGGACAGCAGCAATATTCAAAGATTATCT-GAAAGTCCGGGATTC</u>
NTS2	[6-fam]GTCATTCCCACGAACCTACATTCCGTCATTCCCACGAAAGTGGGAATCTAGAACTCTCTAAAGCTTCAGCTAACCTTTGAATATTGCT-GTTGTTCTAAGGT
NTS3	[6-fam]ATCCCGTCATTCCCGCCACTTTCCGTCATTCCCGCGAAAGCGGGAATCTAGAACTCTCGGACTTTAGATAATCTTTGAATATTGCT-GTTGTTCTAAGGTC
NTS4	[6-fam]TAAATGACCCCAACTTAAAATCCCGTCATTCCCGCGCAGGCGGGAATCTAGGTCTGTCCGGTGC GGAAACTTATCGGGTAAAACGG-TTTCTTGAGATTTTG
NTS5	[6-fam]GTCATTCCCACGAACCTACATTCCGTCATTCCCACGAAAGTGGGAATCCAGTGCGTTGAGTTTCAGTCATTTCCAATAAATTGCCTT-AGTATTGAATGTC
RSE75mut	[6-fam]TTTTTATCAAATTCAAAA ATATTCCGTCATTCCCGCGAAAGCGGGAATCT AGAACGTAAAATCTAAAGAAACCGT
RSE75	[Cy5]TTTTTATCAAATTCAAAAATATTCCGTCATTCCCGCGAAAGCGGGAATCTAGAACGTAAAATCTAAAGAAACCGT
RSE75 G26A	[Cy5]TTTTTATCAAATTCAAAAATATT CA TATTCCCGCGAAAGCGGGAATCTAGAACGTAAAATCTAAAGAAACCGT
RSE75 C49A	[Cy5]TTTTTATCAAATTCAAAAATATTCCGTCATTCCCGCGAAAGCGGGAAT A TAGAACGTAAAATCTAAAGAAACCGT

La NTS est soulignée pour chaque oligonucléotide. Les nucléotides mutés sont indiqués en gras. RSE75mut correspond à 33 oligos indépendants mutés à une seule position à la fois, chacun composé d'un mix des trois autres nucléotides (A→C/G/T). RSE18-25 correspond à 8 oligos de même séquence mais de différentes longueurs, tronqués en 3'. [6-fam] : 6-carboxyfluorescéine, [Cy5] : cyanine 5, utilisée pour la thermophorèse uniquement.

2.4 Tests de liaison à l'ADN

2.4.1 Retards sur gel (EMSA)

Les tests de liaison à l'ADN par retard sur gel ont été réalisés en conditions natives pour ne pas perturber l'association protéine-ADN, avec une plus forte concentration en protéine que d'ADN pour être en conditions saturantes. Afin d'éviter le clivage de l'ADN, médié par un cofacteur métallique, ces expériences ont été réalisées en absence de métaux dans le tampon de réaction. Ces tests consistent à mélanger 1 μM de protéines purifiées avec 0.1 μM de substrat d'ADN fluorescent dans un tampon de réaction sans magnésium (50 mM NaCl, 10 mM Tris-HCl, pH 7.9). Le mix est incubé 30 minutes à température ambiante, après quoi un tampon de chargement natif est ajouté (6x : 10 mM Tris-HCl pH 7.5, 50% glycérol, 0.03% bleu de bromophénol). Comme contrôle négatif, une réaction est réalisée avec de l'eau à la place de la protéine, mais soumise aux mêmes traitements que les autres pour s'assurer qu'il n'y ait pas de dégradation due à la présence de nucléases contaminantes dans le tampon de réaction ou les substrats d'ADN. Les produits de réaction sont alors migrés sur un gel natif de polyacrylamide à 10% (1x TBE, 10% acrylamide/bisacrylamide 19:1, APS et TEMED) avec du TBE comme tampon de migration, pendant 40 min à 200 V. Les gels sont par la suite imagés sur un scanner Typhoon FLA 9500 (GE) avec les paramètres suivants : filtres pour Cy2, résolution à 50 μm et puissance à 900 PMT (tubes photomultiplicateurs).

2.4.2 Thermophorèse

Les expériences de thermophorèse (MST) ont été réalisées avec 2 nM (*Nm*) ou 20 nM (*Ne*) d'oligonucléotides marqués en 5' à la cyanine 5 (**Tableau 8**). Les protéines ont été diluées en série dans un tampon composé de 10 mM Tris, 100 mM arginine, 250 mM $(\text{NH}_4)_2\text{SO}_4$, 3 mM EDTA, 0.05% Tween-20. Les concentrations optimales de protéines et d'oligos ont été déterminées empiriquement de façon à obtenir des spectres de fluorescence uniformes et comparables entre les capillaires, à l'aide d'un appareil Monolith NT.115Pico (NanoTemper), de capillaires Monolith standards, et du logiciel intégré MO.AffinityAnalysis v2.3. La puissance d'excitation a été réglée entre 10 et 30% et l'acquisition à 40%. Entre trois et cinq répétitions de chaque condition ont été réalisées. Les constantes d'affinité (K_D) ont été calculées par le logiciel intégré, puis validées systématiquement à l'aide du logiciel GraphPad Prism 7.00 en utilisant l'algorithme de régression non linéaire « Liaison spécifique - un seul site ».

2.5 Tests de nucléase

Pour étudier l'activité nucléase des protéines purifiées, des tests similaires aux retards sur gel ont été réalisés, mais avec plusieurs modifications essentielles. Parmi ces modifications, du magnésium a été ajouté au tampon de réaction pour permettre l'activité catalytique métal-dépendante des enzymes testées, et les produits de digestions ont été résolus sur gels dénaturants afin de séparer les brins complémentaires et les fragments générés.

2.5.1 Sur gels d'agarose

Avant d'adopter et d'optimiser une méthodologie plus sensible et résolutive sur gels d'acrylamide avec des oligonucléotides fluorescents, divers tests de nucléase ont été résolus sur gels d'agarose. Pour évaluer l'activité nucléase de SsnA sur de l'ADN double-brin, trois substrats ont été utilisés : de l'ADN génomique total linéaire de phage lambda non méthylé (NEB), du plasmide circulaire sans NTS pGEM::Km, ainsi que du plasmide circulaire avec NTS pKOYebN::Sp. 200 ng de chaque substrat a été incubé en présence de 2 μ M de protéine purifiée His-SsnA WT pendant 1h30 à 37°C dans le tampon NEBuffer 2.1 (NEB) contenant du magnésium. Comme contrôle, la protéine a été substituée par de l'eau ou encore inactivée à la chaleur pendant 10 minutes à 95°C avant d'être ajoutée aux réactions. Les mélanges ont alors été résolus sur gels de 1% d'agarose dans du TBE pendant 40 min à 120 V, puis imagés sur l'appareil Gel Doc XR+ (Bio-Rad).

2.5.2 Sur gels d'acrylamide

Produits de PCR

Plusieurs tests de nucléase ont été réalisés sur des produits de PCR avant d'utiliser directement de longs oligonucléotides fluorescents. Pour ces tests, l'ADN génomique total de *N. meningitidis* MC58 a servi de matrice pour amplifier par PCR une région de 397 pb à proximité du gène *yebN* contenant plusieurs NTS. Différentes combinaisons d'amorces avec et sans marqueur fluorescent (6-carboxyfluorescéine en 5', Sigma) ont été utilisées pour générer des produits de PCR marqués sur le brin matrice, le brin complémentaire, ou les deux brins. Ces produits d'amplification ont par la suite été purifiés sur gel. 50 ng de produit de PCR fluorescent ont ensuite été incubés en présence de 10 μ M de protéine purifiée His-SsnA dans le tampon NEBuffer 2.1 pendant 1 h à 37°C. Pour stopper les réactions, un tampon de chargement dénaturant contenant de l'EDTA a été ajouté (2x : 95% formamide, 20 mM EDTA, 0.2% orange G). Les produits ont ensuite été dénaturés à 95°C 3 min pour séparer les brins d'ADN, puis résolus sur gels de

polyacrylamide dénaturants à 10% (1x TBE, 8 M urée, 10% acrylamide/bisacrylamide 19:1, APS et TEMED), pendant 1h30 à 200 V dans du TBE. L'acquisition d'image a été faite sur un appareil Typhoon FLA 9500 avec les paramètres suivants ; filtres pour Cy2, résolution de 50 μM et puissance du laser à 600 PMT. Des tests identiques ont été réalisés en substituant le tampon NEBuffer 2.1 par d'autres tampons de réaction commerciaux (NEB). Comme contrôle, de la protéine purifiée inactivée à la chaleur a été utilisée, ou 10 mM d'EDTA ont été ajoutés aux mélanges de réaction pour prévenir l'activité catalytique métal-dépendante. La protéine mutée His-SsnA E64A a également été testée dans les mêmes conditions comme contrôle supplémentaire.

Oligonucléotides fluorescents

Les tests de nucléase optimisés pour de l'ADN simple-brin ont été réalisés de façon similaire aux précédents, mais en utilisant comme substrat de longs oligonucléotides fluorescents (Sigma) plutôt que des produits de PCR, de l'ADN génomique ou des plasmides. Brièvement, 0.8 μM de chaque oligo ou paires d'oligo complémentaires hybridés sont mélangés à 1 μM de protéines purifiées dans du tampon NEBuffer 2.1, puis incubés 1h30 à 37°C. Les réactions sont encore une fois stoppées par l'ajout de tampon de chargement dénaturant, puis résolues sur gels dénaturant de 17.5% d'acrylamide et imagés au scanner Typhoon.

Pour les tests de cofacteurs, les réactions ont été faites dans un tampon sans magnésium supplémenté avec 10 mM de différents métaux. Pour la détermination de la température optimale, les réactions ont été incubées dans des thermocycleurs avec gradient de température. Pour le test de sensibilité, la protéine a été diluée en série dans le tampon Diluent A (NEB) avant d'être ajouté à l'ADN. Pour les cinétiques de clivage, différentes concentrations de substrat d'ADN ont été testées, et les réactions ont été stoppées aux temps indiqués par l'ajout du tampon de chargement contenant de l'EDTA. Pour la détermination du site de clivage, 0.2 μM des oligos RSE18-25 ont été résolus sur un gel de 20% à côté du produit de clivage de RSE75_F.

Quantifications

Lorsque nécessaire, les bandes correspondant au substrat (oligo intact) et au produit de clivage ont été quantifiées à l'aide du logiciel ImageJ (Schneider *et al.*, 2012). Brièvement, des rectangles de tailles identiques ont servi à mesurer l'intensité moyenne de chaque bande, dont le bruit de fond a été soustrait. Les quantifications ont par la suite été normalisées par rapport à leurs contrôles respectifs, soit les réactions sans nucléase (aucune digestion spécifique) ou encore sans cofacteur métallique. Pour les cinétiques de clivage, ces quantifications ont été rapportées

à la concentration initiale de substrat et analysées avec le logiciel GraphPad Prism 7.00. Les courbes tracées proviennent de régressions non linéaires à partir de l'équation Michaelis-Menten, qui a également servi à l'inférence des constantes catalytiques K_m et k_{cat} . Les vitesses initiales de réaction ont été calculées en mesurant la pente correspondant à la vitesse de réaction au premier point pour chaque substrat, soit entre 0 et 2 minutes. Toutes les quantifications de même que l'ensemble des expériences décrites proviennent d'au moins trois expériences indépendantes.

RÉSULTATS ET DISCUSSION

2.6 Purifications de protéines

La première étape dans la caractérisation biochimique de SsnA consistait à obtenir sous forme pure, soluble, et catalytiquement active des quantités suffisantes de protéine. À notre grande surprise, la purification de ces toutes petites protéines, pourtant prédites comme étant solubles, s'est révélée être un véritable défi qui a nécessité énormément de temps et de ressources. En effet, peu importe la condition testée, près de 95% de la protéine induite a été retrouvée dans la fraction insoluble du lysat, suggérant une certaine toxicité ou une très faible solubilité chez *E. coli* (données non montrées). Cette problématique s'est d'ailleurs empirée lors de nos tentatives de purification de mutants ou d'homologues, donnant des rendements particulièrement faibles ou des protéines impures.

Dans un premier temps, cinq protéines recombinantes ont été générées, soit celle de *N. meningitidis* (WT), trois variantes mutées (Y6A, Y17A, E64A), de même que la protéine homologue provenant de l'espèce commensale *N. elongata* (**Figure 23**). Ces cinq protéines ont été exprimées et purifiées avec un tag polyhistidine (His), un des plus petits sur le marché permettant de mitiger l'impact sur ces minuscules protéines. Les sites mutés (Y6A, Y17A, E64A) ont été sélectionnés à la suite d'une analyse bibliographique d'autres membres de la superfamille des endonucléases GIY-YIG montrant leur importance dans l'activité catalytique (Ibryashkina *et al.*, 2007; Kowalski *et al.*, 1999; Liu *et al.*, 2013; Mak *et al.*, 2010). En effet, l'acide glutamique en position E64 est prédit comme étant le site de liaison d'un cofacteur métallique, probablement du magnésium, essentiel à l'activité nucléase. Les tyrosines en position Y6 et Y17 quant à elles, font partie intégrante du motif conservé « GIY-YIG » également essentiel au clivage d'acides nucléiques. Afin d'exclure la possibilité que le His-tag influence l'activité catalytique de SsnA, une seconde protéine recombinante de *N. meningitidis* a été générée, cette fois avec un tag Glutathion S-transférase (GST). Toutes ces protéines ont été purifiées de façon satisfaisante par chromatographie d'affinité et visualisées par SDS-PAGE (**Figure 23B**). En plus des gels, une analyse par spectrométrie de masse (MALDI-TOF) a été réalisée sur deux d'entre elles pour confirmer leur pureté et leur taille exacte (**Figure 23B**). Selon ces expériences et en absence de cystéines pour former des ponts disulfure, SsnA est monomérique.

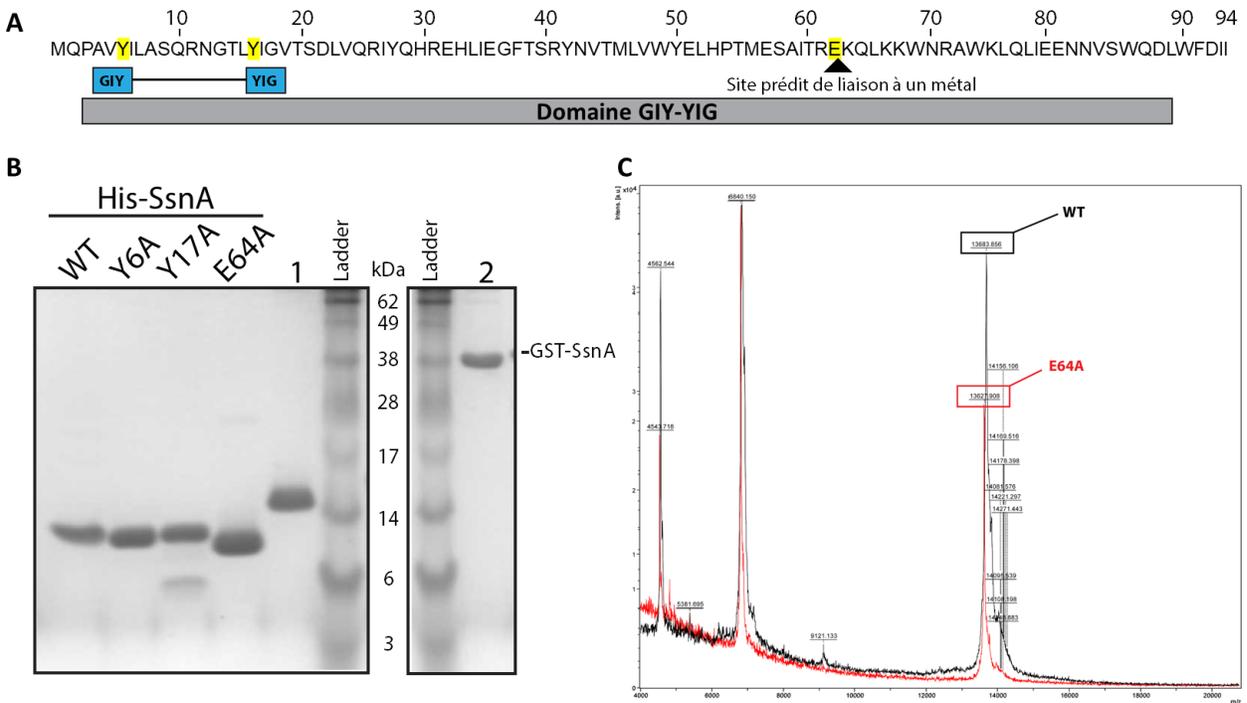


Figure 23 Purification de SsnA sauvage et mutée

A. Séquence protéique de SsnA (*N. meningitidis*), avec les composantes du domaine GIY-YIG annotées. Les résidus mutés sont surlignés. **B.** SDS-PAGE des fractions d'éluion utilisées pour chaque protéine purifiée. WT, Y6A, Y17A, E64A- His-SsnA de *N. meningitidis* 1- His-SsnA de *N. elongata*, 2- GST-SsnA de *N. meningitidis*. **C.** Analyse MALDI-TOF de His-SsnA WT et E64A montrant la pureté et la taille attendue à 13.6 kDa, sans indication d'oligomérisation.

2.7 Activité enzymatique

Une fois purifiées, les protéines ont été soumises à une panoplie de tests de liaison (EMSA) et de clivage d'acides nucléiques pour vérifier que SsnA est bien une endonucléase comme son domaine fonctionnel GIY-YIG le suggère.

2.7.1 Détermination du substrat de SsnA

En premier lieu, SsnA a été incubé en présence d'ADN double-brin linéaire et circulaire, puis le mélange a été résolu par électrophorèse sur gel d'agarose (**Figure 24**). À titre de contrôle supplémentaire, une fraction de SsnA a été inactivée à la chaleur quelques minutes immédiatement avant l'expérience. Dans l'optique de démontrer une interaction physique entre SsnA et les séquences répétées NTS, un plasmide contenant plusieurs de ces répétitions a également été testé. Ces expériences démontrent que SsnA n'est pas une DNase ou une exonucléase non spécifique, et suggèrent également qu'elle est incapable de cliver la NTS dans

les conditions testées. Cependant, la sensibilité des marqueurs d'ADN intercalant ne permet pas de détecter des fragments peu concentrés ou de petite taille.

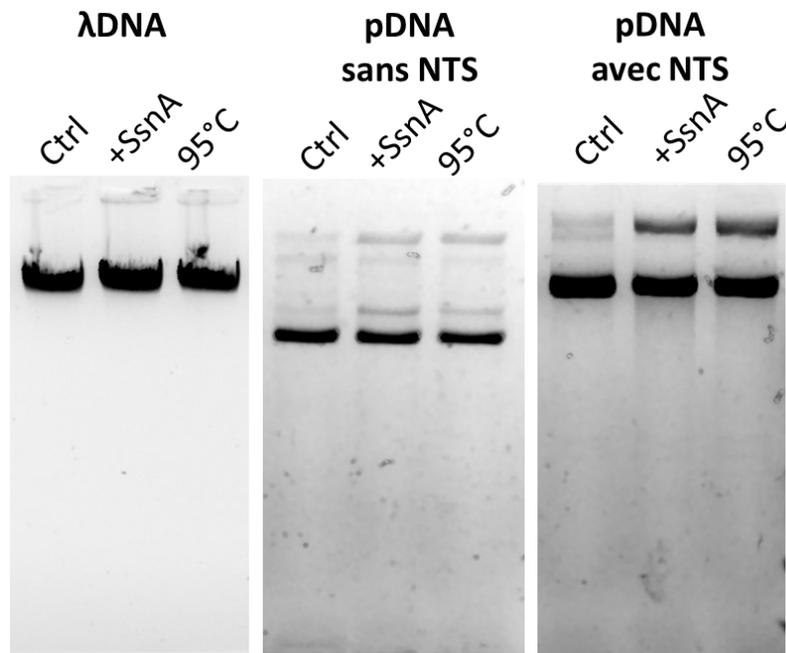


Figure 24 Tests d'activité nucléase résolus sur gel d'agarose

De l'ADN de phage lambda non méthylé (ADN λ double-brin linéaire) ainsi que deux plasmides (ADNp double-brin circulaire, pGEM et pKOYebN) ont été incubés à 37°C en absence (Ctrl) et en présence de SsnA (+SsnA), puis résolus par électrophorèse sur gel d'agarose. Les mêmes réactions ont aussi été réalisées après inactivation à la chaleur de SsnA (95°C).

Pour améliorer la sensibilité de nos tests, de même que la séparation et la résolution des acides nucléiques testés, nous avons tenté des expériences similaires, mais cette fois à partir de produits de PCR générés avec des amorces fluorescentes, puis résolus sur gels de polyacrylamide dénaturants (**Figure 25**). Cette approche, contrairement à la précédente, permet d'observer des produits de clivage en très faible concentration, sans restriction de taille puisque chaque amplicon est marqué en 5' avec un seul fluorophore. En utilisant un produit de PCR contenant une NTS, nous avons observé un très léger produit de clivage obtenu uniquement en présence de SsnA WT (**Figure 25A**). Ce fragment est obtenu avec divers tampons de réaction, mais n'apparaît pas en présence de protéine inactivée à la chaleur ou de protéine mutée (E64A). L'apparition d'une seule bande précise démontre que SsnA est une endonucléase spécifique et non une exonucléase ou une endonucléase non spécifique, autrement on observerait plutôt une traînée diffuse. Lorsqu'un chélateur de métaux (EDTA) est ajouté à la réaction, aucun produit de clivage n'est obtenu (**Figure 25B**). Puisque le seul métal présent dans le tampon 2.1 est le magnésium, ces résultats indiquent que SsnA est une nucléase magnésium-dépendante, et supportent

l'hypothèse que l'acide glutamique en position 64 est essentiel à l'interaction avec ce cofacteur métallique. De plus, le fragment d'ADN généré par SsnA mesure une trentaine de nucléotides, suggérant que l'enzyme ne coupe pas dans la NTS mais plutôt en amont de celle-ci.

Pour faire suite à ces résultats prometteurs, nous avons généré par PCR le même substrat d'ADN, cette fois marqué en 5' uniquement sur le brin matrice comme dans les tests précédents, uniquement sur le brin complémentaire, ou sur les deux brins simultanément (**Figure 25C**). À notre grande surprise, plutôt que d'obtenir deux produits de clivage avec l'ADN marqué sur les deux brins, ce qui correspondrait à une coupure double-brin, un seul fragment a été obtenu, tandis qu'aucun fragment n'a été produit avec l'ADN marqué uniquement sur le brin complémentaire. Ces résultats suggèrent que SsnA agit comme endonucléase de type nickase, capable de cliver un seul brin d'ADN sur un substrat double-brin. Cependant, cette activité nickase est infime et n'affecte qu'une fraction du substrat d'ADN, ce qui explique pourquoi les tests en gel d'agarose avec un marqueur intercalant n'avaient rien révélé. Cette très faible activité soulève deux hypothèses : soit les conditions de clivage sont sous-optimales, soit il s'agit d'activité non spécifique comparable à l'activité étoile (*star activity*) des enzymes de restriction. Il est important de noter ici que la quantité de protéine nécessaire pour obtenir cette infime activité nucléase est particulièrement élevée (10 μ M) par rapport à ce qui est d'usage courant en biologie moléculaire (généralement <1 μ M).

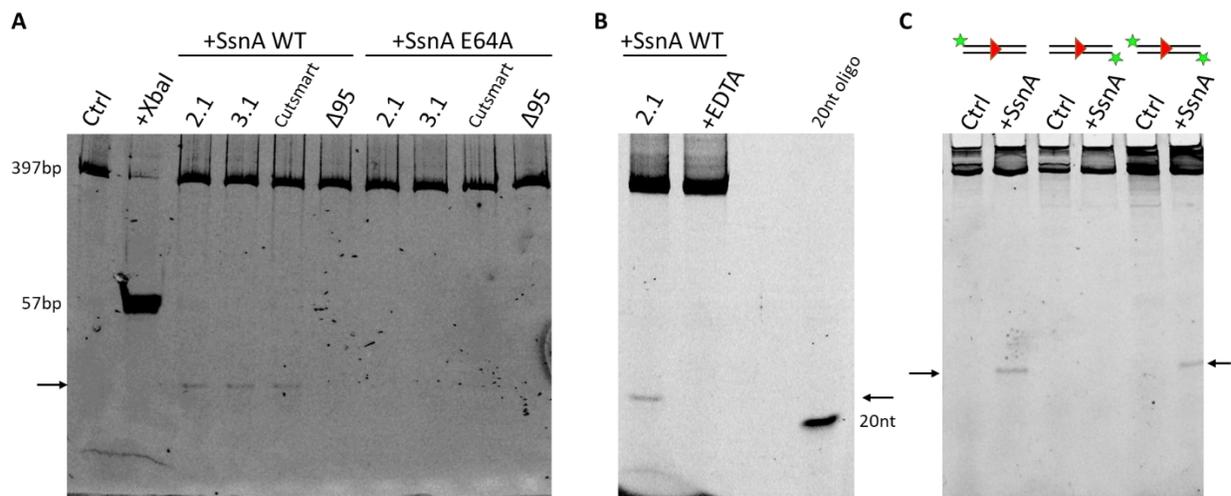


Figure 25 Tests d'endonucléase sur de l'ADN double-brin en gel de polyacrylamide

A. Clivage d'un produit de PCR fluorescent contenant une NTS. SsnA WT et E64A ont été incubés avec l'ADN en présence de différents tampons pour enzymes de restriction (NEBuffer 2.1, 3.1 et Cutsmart). Seul le brin matrice est marqué en fluorescence à son extrémité 5'. Le même ADN a également été digéré par l'enzyme commerciale XbaI générant un fragment de 57 pb. Les produits de clivages ont été résolus sur gel de polyacrylamide en conditions dénaturantes. **B.** Similaire en A, avec un contrôle supplémentaire où 25 mM d'EDTA ont été ajoutés dans le mix protéine-ADN. Un oligonucléotide fluorescent de 20nt a également été résolu afin de servir de marqueur de taille. **C.** Similaire aux tests précédents mais avec différentes combinaisons de fluorescence en 5' du brin matrice, du brin complémentaire, ou des deux brins. La flèche rouge indique la présence d'une NTS dans la séquence testée.

Puisque SsnA semble couper à très faible niveau un seul brin d'ADNdb contenant une NTS, et ce uniquement lorsque très concentrée, nous avons exploré l'idée que cette nucléase puisse agir sur l'ADN simple-brin (ADNsb). Pour ce faire, plutôt que d'utiliser des produits de PCR, nous avons fait synthétiser de longs oligonucléotides fluorescents complémentaires contenant une séquence similaire, que nous pouvions ensuite tester individuellement (ADNsb) ou hybridés entre eux (ADNdb). Après un long processus d'optimisation pour obtenir la meilleure sensibilité et résolution possible, nous sommes parvenus à des résultats qui ont complètement débloqué le projet et qui ont ouvert la porte à de nombreuses autres expériences. Nous avons notamment optimisé les ratios de protéine-ADN, la quantité de substrat fluorescent, le temps d'incubation, ainsi que le maillage des gels de polyacrylamide (voir section matériel et méthodes). Nos résultats démontrent que SsnA clive de façon spécifique et efficace de l'ADN simple-brin contenant la NTS dans une orientation précise, mais est incapable de cliver l'ADN double-brin ni même l'ADNsb avec la NTS dans l'orientation opposée (**Figure 26A**). Testée à bien plus faible concentration que dans les expériences précédentes, SsnA catalyse le clivage de plus de 95% de l'ADNsb contenant son site de reconnaissance, indiquant qu'il s'agit d'une ssDNase, soit une nucléase spécifique à l'ADN simple-brin. Cette activité est entièrement perdue lorsque la protéine est mutée en position Y6, Y17 et E64, confirmant que ces acides aminés font bien partie du site actif de l'enzyme. Un résultat identique est obtenu lorsque SsnA est purifiée et testée avec un tag GST plutôt qu'un his-tag, confirmant que la présence de ces marqueurs n'a pas d'impact significatif sur l'activité catalytique de la nucléase. L'absence d'activité sur l'ADNsb' (brin complémentaire) indique que SsnA possède une spécificité de séquence, capable de discriminer les deux brins d'ADN. Cette activité nucléase atypique a inspiré le nom que l'on a donné à cette protéine, soit SsnA pour *Specific single-strand nuclease*.

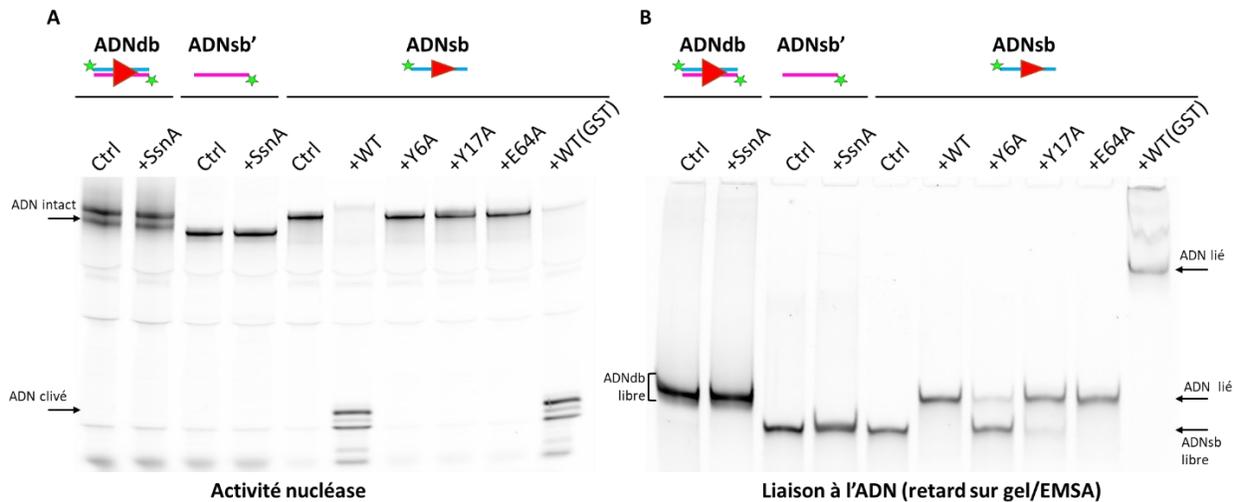


Figure 26 Activité catalytique de SsnA sur de l'ADN simple-brin

A. Test d'activité nucléase sur des oligonucléotides complémentaires, marqué avec un fluorophore en 5', contenant une NTS (RSE100_F/R). Les oligos ont été testés individuellement (ADNsb et ADNsb') ou hybridés ensemble avant l'expérience (ADNdb). Pour l'ADNsb, His-SsnA WT ainsi que les trois mutants ont été testés (Y6A, Y17A, E64A), en plus de GST-SsnA WT. Les réactions ont été résolues par PAGE en conditions dénaturantes. **B.** Test de liaison à l'ADN, aussi appelé EMSA ou retard sur gel. Réactions similaires en A, mais avec un tampon de réaction ne contenant pas de magnésium, et résolues par PAGE en conditions natives.

Suite à ces résultats prometteurs, nous nous sommes intéressés à l'activité de liaison à l'ADN de SsnA. En effet, le clivage spécifique d'acides nucléiques nécessite obligatoirement une étape transitoire de liaison entre la protéine et son substrat. Malgré cela, la spécificité de clivage n'est pas toujours la même que la spécificité de liaison, notamment dans le cas des endonucléases de type homing. Pour étudier ce phénomène, nous avons dû adapter la méthodologie utilisée pour les tests de clivage. Tout d'abord, nous avons enlevé le magnésium du tampon de réaction, cofacteur essentiel à l'activité de clivage de nombreuses nucléases dont SsnA. Ensuite, nous avons modifié les ratios ADN-protéine pour favoriser la saturation de l'ADN en protéine. Le temps et la température d'incubation ont également été modifiés pour favoriser la formation de structures secondaires dans l'ADN, et limiter la possibilité de dégradation non spécifique. Enfin, les complexes ADN-protéine ont été résolus sur gels de polyacrylamide natifs afin de ne pas les dissocier, par la méthode de retard sur gel aussi appelée EMSA (*electrophoretic mobility shift assay*). Logiquement, un ADN couplé à une protéine migre plus lentement dans un gel qu'un ADN seul, ce qui se résulte en un décalage ou un retard. Cette méthodologie a été appliquée aux mêmes échantillons que dans les précédents tests de nucléase (**Figure 26B**). Tant avec un his-tag qu'avec un tag GST, la nucléase se lie efficacement à l'ADNsb contenant son site de reconnaissance, ce qui est compatible avec les précédents tests de nucléase. Le retard est beaucoup plus prononcé avec le tag GST puisque ce dernier est nettement plus gros que le his-

tag et forme donc un complexe ADN-protéine plus imposant. Dans les deux cas, la bande retardée est nette et unique, ce qui est cohérent avec une activité sous forme monomérique uniquement. SsnA est incapable de lier l'ADNdb ni l'ADNsb' (complémentaire), ce qui explique pourquoi ces substrats n'étaient pas clivés dans les tests précédents. Les mutations Y17A et E64A permettent à SsnA de lier l'ADNsb mais pas de le cliver, confirmant que ces résidus sont essentiels à l'activité de clivage indépendamment de la liaison. À l'inverse, la mutation Y6A réduit considérablement le potentiel de liaison à son substrat, suggérant que cet acide aminé est impliqué dans cette étape plutôt que dans le clivage.

Ensemble, ces résultats confirment formellement que SsnA est une endonucléase d'ADN simple-brin avec une spécificité de séquence. Cette activité nucléase est uniquement observée lorsque le substrat d'ADNsb contient cet élément répété. En absence de magnésium ou lorsque son site catalytique est muté, l'enzyme est capable de lier spécifiquement son substrat mais est incapable de le cliver, suggérant que ces deux activités sont modulables. De plus, la protéine purifiée His-SsnA de *N. elongata* affiche une activité catalytique et une spécificité très similaire à celle de *N. meningitidis*, avec une séquence en acides aminés légèrement différente (82% d'identité) (voir Chapitre IV). Ces conclusions sont donc applicables aux protéines provenant de plusieurs espèces de *Neisseria*.

2.7.2 Conditions optimales de clivage

Une fois le bon substrat de clivage identifié, nous avons pu expérimenter dans quelles conditions l'enzyme est la plus active afin de procéder à la caractérisation de son activité nucléase plus en détail. Nous avons en premier lieu vérifié si le magnésium (Mg) est le seul cofacteur métallique divalent utilisable par SsnA (**Figure 27A**). Seuls deux autres métaux ont permis une certaine activité catalytique : le manganèse (Mn) et le nickel (Ni). Afin de s'assurer que l'activité liée au nickel ne soit pas due à la présence du his-tag sur la protéine recombinante, ce métal a également été testé avec la protéine marquée à la GST, donnant les mêmes résultats. Différentes concentrations de magnésium et de manganèse ont été testées, révélant que 10 mM de Mg permet une activité nucléase optimale, tandis que le manganèse ne permet qu'une fraction de cette activité, peu importe la concentration utilisée (**Figure 27B**). Le nickel n'a pas été testé davantage. L'activité nucléase de SsnA a également été évaluée à différentes températures (**Figure 27C**). Sans grande surprise, considérant que cette protéine est retrouvée principalement chez les *Neisseria*, des symbiotes de mammifères, cette enzyme est optimale à 37°C. Nous notons cependant qu'une activité non négligeable est détectée à une large gamme de

températures, allant de 15 à 54°C. Enfin, nous avons testé la sensibilité de cette enzyme en incubant des dilutions en série de cette dernière avec une concentration fixe de substrat (**Figure 27D**). Ces tests ont révélé une sensibilité remarquable, avec la conversion de substrat en produit détectable dès 64 pM de nucléase.

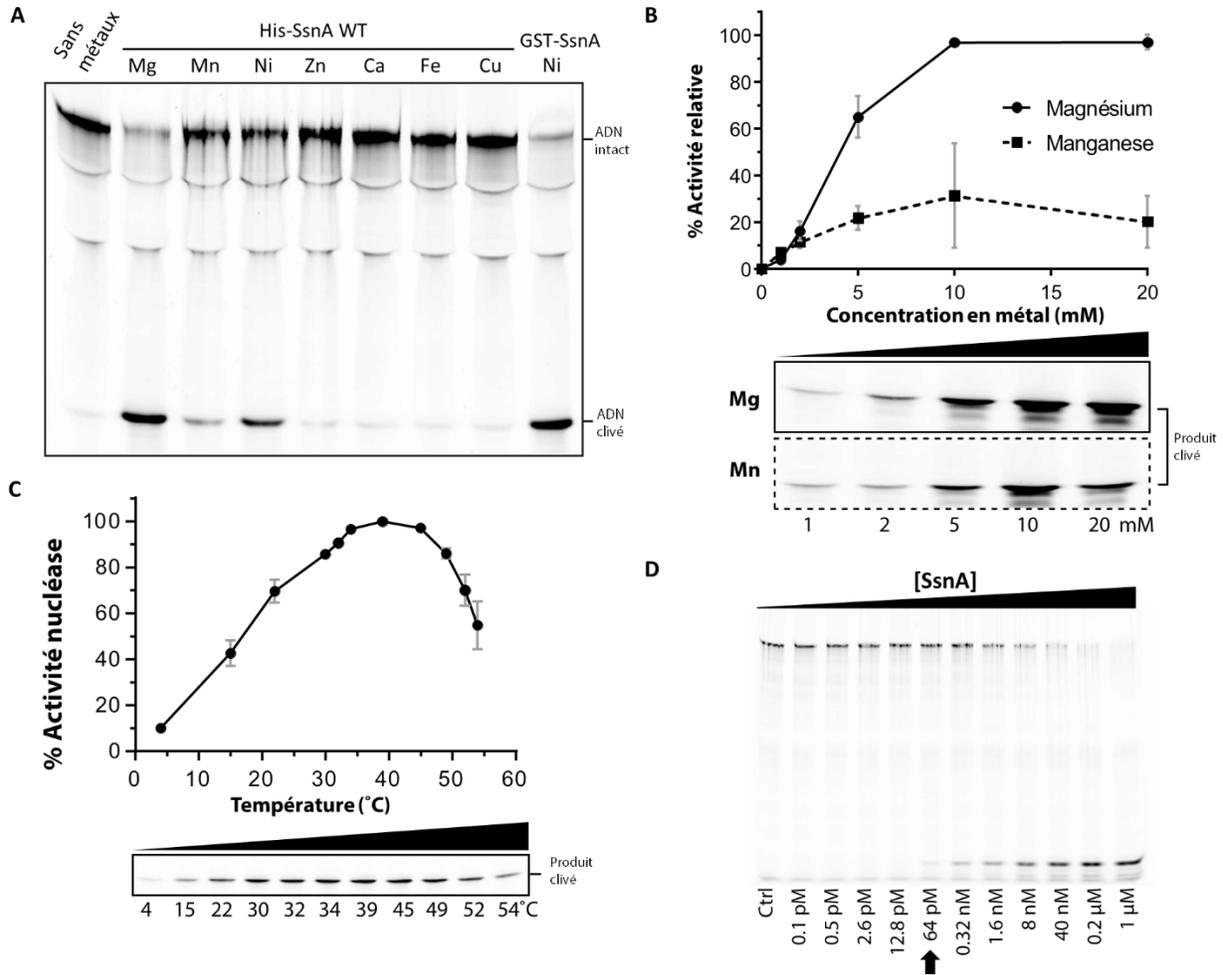


Figure 27 Optimisation des conditions de clivage de SsnA

A. Dépendance en cofacteur métallique de SsnA. Une concentration de 10 mM de chaque métal a été ajoutée à un tampon de réaction n'en contenant aucun. Étant donné que le his-tag interagit avec le nickel, ce métal a également été testé avec la protéine taguée à la GST. Mg magnésium, Mn manganèse, Ni nickel, Ca calcium, Fe fer, Cu cuivre. **B.** Tests de dose-réponse de magnésium et de manganèse sur l'activité de clivage de SsnA. Le produit de clivage a été quantifié de trois expériences indépendantes dont la moyenne et l'erreur type sont montrées. Un résultat représentatif résolu par PAGE est également montré en exemple. **C.** Effet de la température sur l'activité nucléase de SsnA. Trois expériences indépendantes ont été réalisées et quantifiées. La moyenne et l'erreur type sont montrées avec un exemple représentatif de résultat brut. **D.** Test de sensibilité de SsnA. L'enzyme a été diluée en série dans le tampon Diluant A (NEB) avant de l'incuber en présence de son substrat d'ADNsb. Une flèche indique la plus petite concentration où un produit de clivage est détecté. L'oligonucléotide fluorescent RSE100_F a été utilisé comme substrat pour toutes ces expériences.

2.7.3 Cinétiques de clivage

Aux vues de l'activité importante de SsnA sur de l'ADNsb, nous avons par la suite procédé à des cinétiques de clivage afin d'évaluer sa vitesse et de déterminer ses constantes catalytiques, qui permettent de comparer l'efficacité de différentes enzymes entre elles. Pour ce faire, nous avons réalisé des tests de nucléase en incubant une concentration fixe de protéine avec différentes concentrations de substrat, et ce pendant différents temps allant de deux minutes à 1h30. Pour stopper la réaction instantanément et ainsi garantir la précision de ces cinétiques, un tampon supplémenté en EDTA a été ajouté au mix à chaque intervalle de temps. En chélatant le magnésium, l'activité nucléase de SsnA est ainsi momentanément stoppée, et le mélange peut être préservé jusqu'à sa résolution par PAGE. Les résultats de ces cinétiques sont présentés en **Figure 28**. Dans les 15 premières minutes, plus de 90% du substrat est clivé par SsnA, démontrant une certaine rapidité (**Figure 28A-B**). Cette efficacité se traduit par une constante catalytique (k_{cat}) de $0.0019 \pm 0.002 \text{ s}^{-1}$ et une constante de Michaelis-Menten (K_m) de $4.68 \pm 1.45 \mu\text{M}$ (**Figure 28C**). k_{cat} désigne le taux de conversion du substrat en produit, soit le nombre de molécules catalysées par seconde, tandis que K_m désigne la concentration de substrat correspondant à la moitié de la vitesse maximale de l'enzyme. Ces chiffres seront discutés dans une prochaine section de ce chapitre.

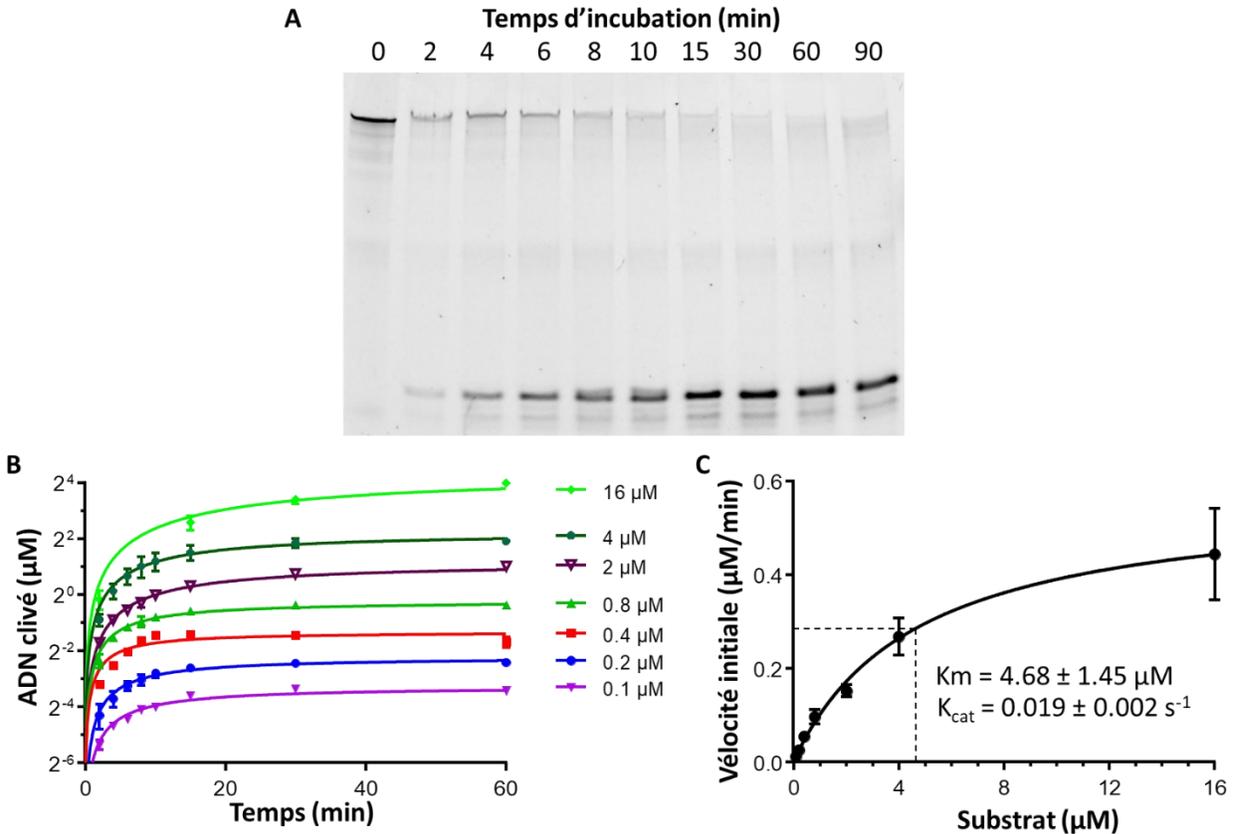


Figure 28 Détermination de la constante catalytique de SsnA

A. Exemple représentatif de cinétique de clivage, réalisé avec 0.1 μM de substrat. Les bandes correspondant aux substrats ainsi qu'aux produits de clivage ont été quantifiées pour chaque réplicat de chaque concentration testée. **B.** Compilation des cinétiques réalisées avec huit concentrations de substrat, chacune répétée au moins trois fois. Les moyennes et erreurs types sont présentées. **C.** Courbe de Michaelis-Menten montrant la vélocité initiale de SsnA en fonction de la concentration en substrat. Les vitesses initiales ont été calculées à partir du premier point de chaque cinétique, correspondant à la vitesse maximale de réaction à chaque concentration testée. Les moyennes et écarts-types sont présentés, ainsi que la valeur inférée de la constante catalytique K_m , égale à la concentration en substrat équivalent à la moitié de la vélocité maximale de l'enzyme, indiqué par des lignes pointillées. Toutes les expériences ont été réalisées avec l'oligonucléotide fluorescent RSE100_F.

2.7.4 Spécificité et site de clivage

Une fois l'activité catalytique de SsnA mieux définie et ses conditions optimales déterminées, nous avons voulu en savoir plus sur la spécificité de cette enzyme et son site de clivage précis. Pour ce faire, nous avons fait synthétiser des oligonucléotides fluorescents correspondant à la région clivée par SsnA, en conservant uniquement les 18 à 25 nucléotides en 5', afin de générer une échelle d'ADN simple-brin sans biais de séquence (**Figure 29A**). À l'aide d'un gel hautement résolutif (acrylamide à 20%), on observe que le fragment spécifiquement clivé par SsnA mesure exactement 20 nt. Ce site de clivage se situe quelques nucléotides en amont du début de la

séquence répétée NTS dans le substrat testé. Nous avons ensuite regardé l'interaction de SsnA avec la même séquence, mais raccourcie en 5' et en 3' de manière à ne conserver que la NTS sans ses régions flanquantes (**Figure 29B**). Ces tests révèlent que SsnA n'a pas besoin de son site de clivage pour se lier spécifiquement à l'ADNsb tant que la NTS est présente. En effet, une séquence de 28 nt contenant uniquement la NTS permet la liaison de l'enzyme, mais pas le clivage, alors qu'une séquence de 37 nt contenant la NTS et quelques nucléotides en amont permet autant la liaison que le clivage. Nous avons aussi testé l'activité de SsnA sur de l'ARN simple-brin de même longueur et de même séquence. Contrairement à l'ADNsb, aucune interaction n'est observée avec l'ARN. Ces résultats confirment que SsnA est une nucléase spécifique d'ADNsb incapable de cliver d'autres substrats. Cette enzyme se lie spécifiquement à la NTS et clive l'ADNsb quelques nucléotides en amont de celle-ci.

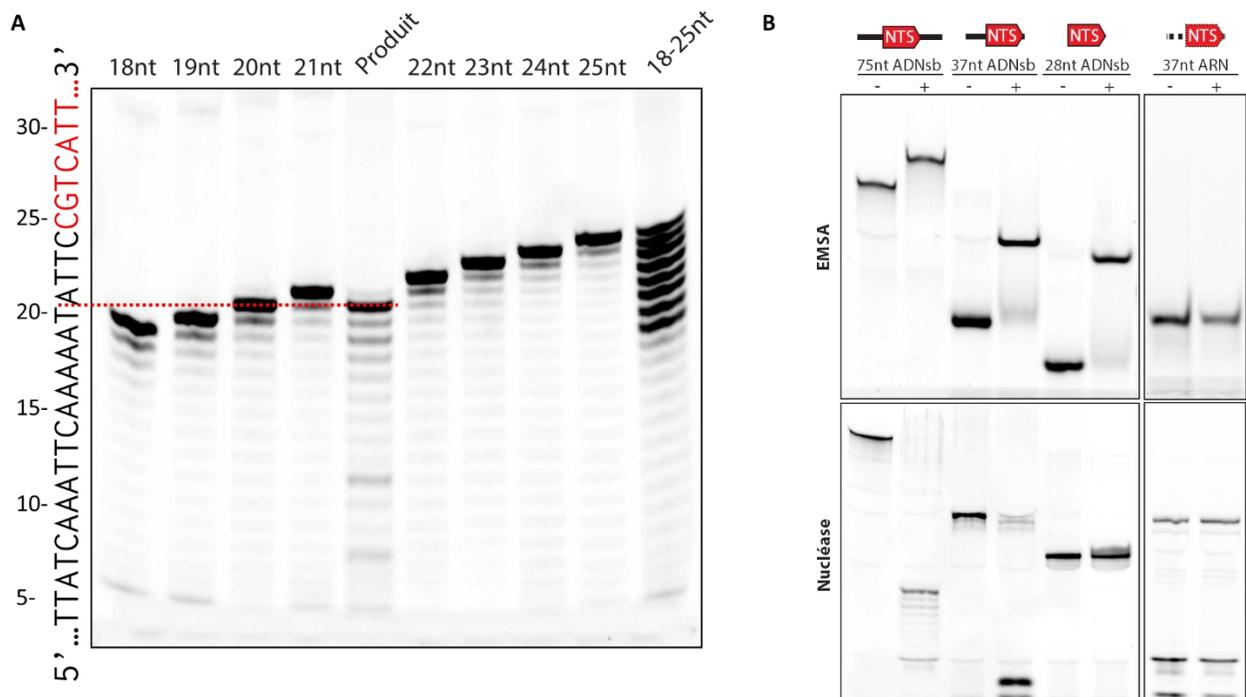


Figure 29 Détermination du site de clivage et de la taille minimale du substrat

A. Le produit de clivage de His-SsnA WT sur l'oligonucléotide fluorescent RSE75_F (au centre) a été résolu sur un gel de polyacrylamide dénaturant de 20% aux côtés d'oligos fluorescents de même séquence tronqués en 3', mesurant 18 à 25 nt. La séquence testée est affichée à gauche, alignée avec la position des bandes. Le début de la répétition NTS est illustré en rouge. La ligne pointillée montre la taille précise du fragment obtenu. **B.** Des oligos contenant une NTS flanquée des mêmes séquences mais de différentes longueurs (RSE75_F, RSE37, RSE28) ont été soumis à des tests de retard sur gel (haut) et d'activité nucléase (bas). Un oligo de séquence identique à RSE37 mais sous forme d'ARN a également été testé (RNA37). Les signes - et + désignent l'absence et la présence de His-SsnA WT dans la réaction.

Puisque la NTS est une séquence répétée relativement grande, comprenant une trentaine de nucléotides dont certains sont mieux conservés que d'autres, nous avons voulu explorer davantage l'interaction entre SsnA et cet élément fréquemment retrouvé dans le génome des *Neisseria*. Habituellement, les nucléases reconnaissant de grandes séquences tolèrent quelques mutations ou substitutions, contrairement aux enzymes de restriction dont le site de reconnaissance est très court mais invariable. Pour vérifier la rigueur de la spécificité de SsnA, nous avons testé son activité de liaison et de clivage sur plusieurs dizaines d'oligonucléotides, chacun possédant une seule substitution par rapport à la séquence d'origine (**Figure 30**). On constate que la capacité de liaison de la protéine à son substrat est presque entièrement perdue lorsque ce dernier est muté en position C49. Plusieurs autres mutations entre les positions 30 à 48 réduisent plus modestement cette activité, soit de 40 à 60% (**Figure 30A**). Lorsqu'on regarde uniquement le clivage de l'ADN, on constate que les nucléotides en position 26-28 et 30 ainsi que 47 et 49 sont essentiels à l'activité catalytique de SsnA (**Figure 30B**). Cependant, la liaison à l'ADN est une étape préalable essentielle au clivage. Les résultats de clivage ont donc été normalisés à ceux de liaison afin de dresser un portrait représentatif de l'importance de chaque nucléotide dans l'activité catalytique de SsnA envers son substrat (**Figure 30C**).

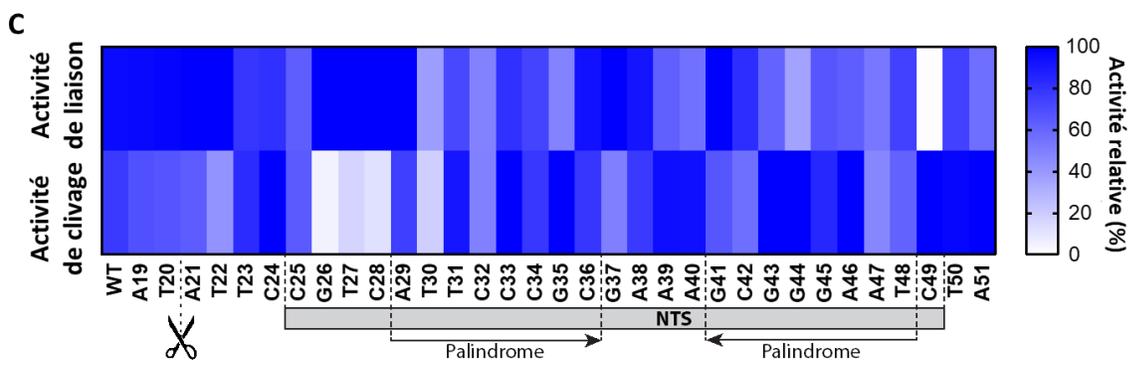
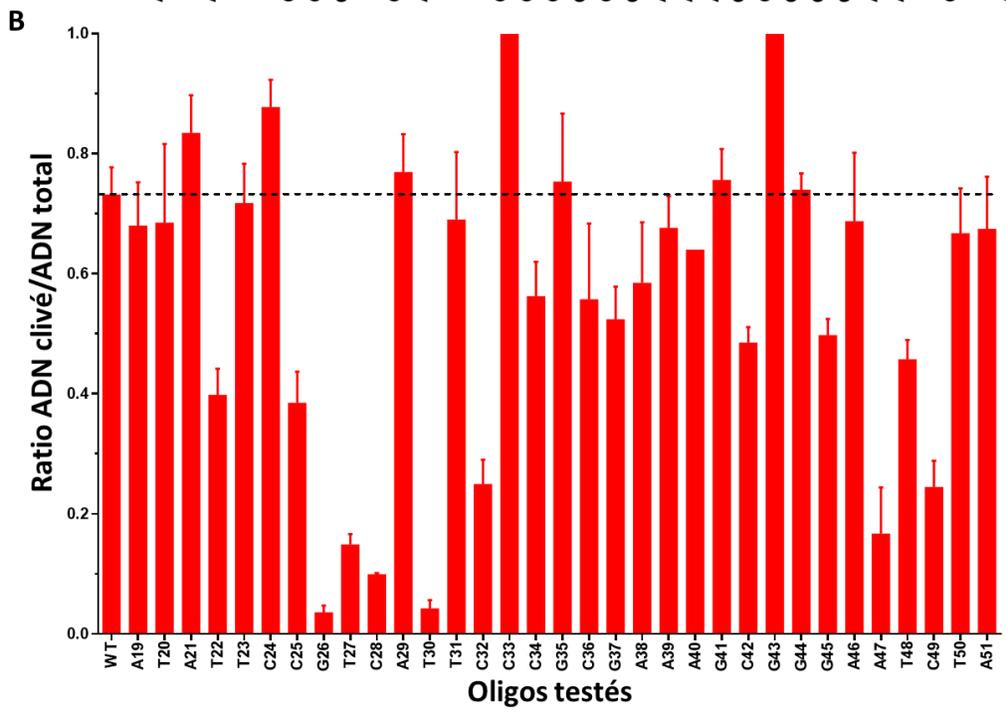
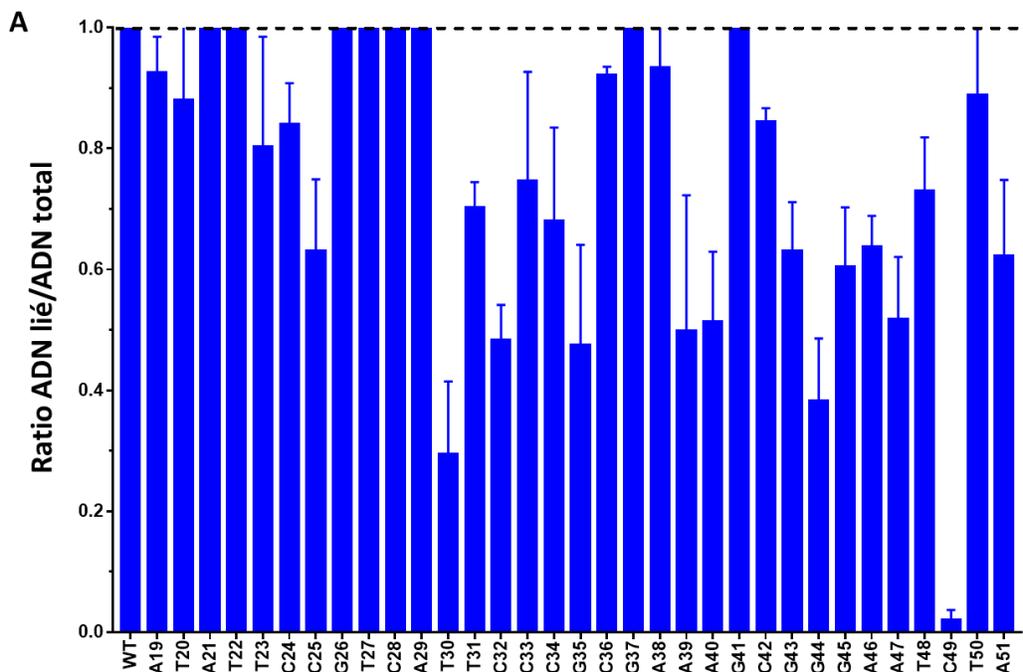


Figure 30 Cartographie de la spécificité de séquence de SsnA

A. Tests de liaison à l'ADN par EMSA de SsnA avec des oligonucléotides affichant chacun une substitution par rapport à la séquence d'origine (RSE75_F = WT). Chaque substitution consiste en un mélange des trois autres nucléotides possibles (ex : A→C/G/T). Les quantifications de l'ADN lié sur l'ADN total provenant d'au moins trois expériences indépendantes pour chaque oligo sont présentées, de même que la moyenne et l'erreur type. **B.** Tests de nucléase avec les mêmes substrats qu'en A. Quantifications brutes de l'ADN clivé sur l'ADN total d'au moins trois réplicats indépendants, avec la moyenne et l'erreur type. **C.** Compilation sous forme de carte de densité (*heatmap*) des résultats précédents, où l'activité nucléase a été normalisée en fonction de l'activité de liaison puisque l'enzyme ne peut théoriquement pas cliver son substrat sans d'abord s'y lier. L'intensité de la couleur bleue est proportionnelle à chaque type d'activité. La NTS, sa région palindromique ainsi que le site de clivage de SsnA sont indiqués.

Cette représentation simplifiée permet de constater que les nucléotides impliqués dans la liaison par SsnA diffèrent de ceux impliqués dans le clivage, suggérant une spécificité complexe. L'activité de SsnA est affectée par des substitutions sur une longue étendue de la séquence testée, qui couvrirait vraisemblablement un espace beaucoup plus large que la protéine si l'oligonucléotide n'adoptait pas de structure secondaire. Lorsque rapportés à la NTS, ces résultats démontrent que la liaison de SsnA dépend majoritairement de la séquence palindromique composant celle-ci, ainsi que de la cytosine immédiatement en aval du palindrome. Ceci suggère fortement que l'enzyme interagisse avec le pied de la tige formant la structure en épingle à cheveux donnée par la NTS. En effet, la formation de cette structure secondaire est la façon la plus probable et logique pour SsnA, une toute petite protéine, d'interagir avec autant de nucléotides à la fois. En contraste, le clivage de l'ADNsb par SsnA est plutôt médié par les trois nucléotides en amont du palindrome, qui sont donc immédiatement devant la tige de la structure en épingle à cheveux. De façon intrigante, la mutation des nucléotides correspondant au site de clivage précis de la nucléase affecte relativement peu son activité catalytique. Cette observation suggère que SsnA se lie spécifiquement au pied de la tige-boucle formée par la NTS, et clive à une distance fixe en amont de celle-ci. Ensemble, ces résultats confirment l'association physique entre la protéine SsnA et les éléments répétés NTS, en plus de prouver l'importance biologique des extensions hautement conservées autour du palindrome, qui différencient les NTS identifiées dans ce projet des séquences répétées dRS3 initialement décrites.

2.7.5 Clivage d'autres copies de la NTS

Dans tous les tests précédents, nous nous sommes focalisés sur une séquence précise contenant une NTS extraite du génome de *N. meningitidis*, et avons démontré que cette séquence interagit physiquement avec SsnA. Puisque plusieurs centaines de répétitions de la NTS sont retrouvées dans le même génome, nous avons élargi nos tests à des séquences contenant des NTS issues de différents loci génomiques (**Figure 31A**). Parmi celles-ci, on retrouve les séquences flanquant les gènes codant pour SsnA mais aussi les systèmes

d'acquisition du fer Tbp et Lbp, essentiels à la virulence. Toutes les constructions testées sont efficacement liées par SsnA, ce qui confirme que cette liaison est dépendante de la NTS et non de ses séquences avoisinantes. Cependant, l'activité nucléase de SsnA envers ces constructions semble être beaucoup plus variable. En effet, certaines ne sont pas du tout clivées tandis que d'autres le sont plus efficacement. Ces résultats indiquent que SsnA possède une certaine spécificité de séquence pour son activité nucléase, en plus de sa spécificité de liaison à la NTS. Cette activité variable est prédite comme étant particulièrement importante aux fonctions biologiques de SsnA chez *N. meningitidis* en particulier, dont plusieurs gènes importants sont entourés de nombreuses copies de la NTS.

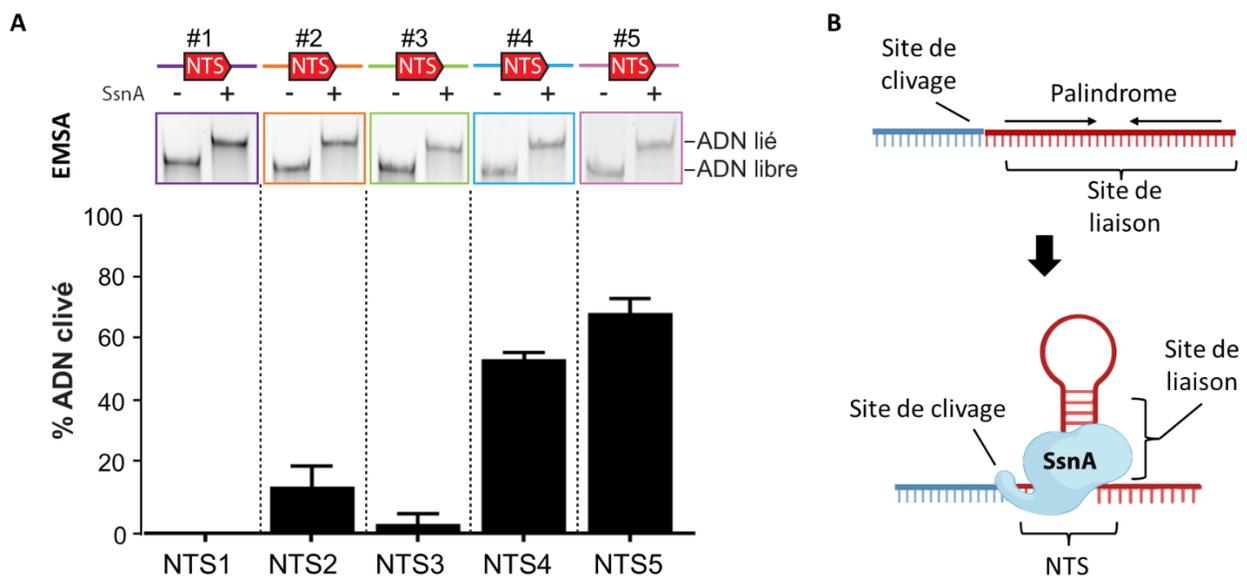


Figure 31 Interaction de SsnA avec d'autres NTS et modèle proposé

A. Tests de liaison et de clivage de séquences contenant des NTS provenant de diverses régions du génome de *N. meningitidis*. Un exemple représentatif de retard sur gel est illustré pour chaque construction testée, ainsi que les quantifications provenant d'au moins trois tests de nucléase indépendants, dont les moyennes et erreurs types sont présentées. **B.** Représentation graphique de l'interaction entre SsnA et l'ADNsb contenant son site de reconnaissance, faite avec BioRender. Sous forme d'ADNsb, la NTS forme une structure en épingle à cheveux en raison de sa séquence palindromique. SsnA lie de façon spécifique le pied de la tige de cette structure, et clive quelques nucléotides en amont de la NTS.

Combinée aux précédentes expériences, cette analyse nous permet de proposer un modèle d'interaction de SsnA avec la NTS qui expliquerait en partie comment une protéine de moins de 100 acides aminés peut interagir physiquement avec une séquence de cette taille (**Figure 31B**). Puisque la NTS abrite une longue région palindromique susceptible de former une structure en épingle à cheveux sous forme d'ADNsb, SsnA serait capable de reconnaître spécifiquement cette structure et de se fixer au pied de la tige pour couper non loin de là. Ce modèle théorique est supporté à échelle réelle par des modélisations *in silico* que nous avons réalisées (**Annexe I**).

Cependant, seules des études structurales expérimentales permettront d'inférer avec assurance le mécanisme de reconnaissance et de clivage de SsnA, qui nécessite probablement un changement de forme de la protéine et/ou de son substrat.

2.7.6 Affinité (K_D)

Dans une section précédente, nous avons déterminé les constantes catalytiques relatives à l'activité nucléase de SsnA. Afin de mieux évaluer cette activité, nous nous sommes également intéressés à son affinité de liaison pour son substrat. Des expériences de thermophorèse ont donc été réalisées avec les protéines de *N. meningitidis* et *N. elongata* sur plusieurs oligonucléotides.

Dans un premier temps, nous avons vérifié l'impact de mutations dans l'ADNsb précédemment identifiées comme importantes pour l'activité catalytique de SsnA (voir **Figure 30**). Comme suggéré par nos tests de retard sur gel, la mutation C49A diminue la liaison de SsnA à son substrat de plusieurs logarithmes (**Figure 31A**). La mutation G26A quant à elle, initialement identifiée comme essentielle au clivage de l'ADNsb mais pas à sa liaison, cause également une nette réduction d'affinité de SsnA. Nous avons ensuite comparé l'affinité de SsnA pour un substrat d'ADNsb comprenant uniquement son site de liaison (RSE28) à un autre contenant à la fois son site de liaison et son site de clivage (RSE75) (**Figure 32B**). De façon intéressante, la perte du site de clivage réduit considérablement l'affinité de liaison de SsnA pour son substrat, et ce autant pour l'enzyme de *N. meningitidis* que pour celle de *N. elongata*. Ces résultats indiquent que la présence du site de clivage stabilise considérablement la liaison de la protéine à son substrat, même si l'identité des bases à cette position n'a pas d'impact significatif sur l'activité de clivage. Concrètement, cela implique que SsnA interagit physiquement avec son site de liaison et de clivage de façon simultanée, ce qui supporte le modèle proposé d'interaction avec le pied de la tige-boucle formée par la NTS (**Figure 31B**).

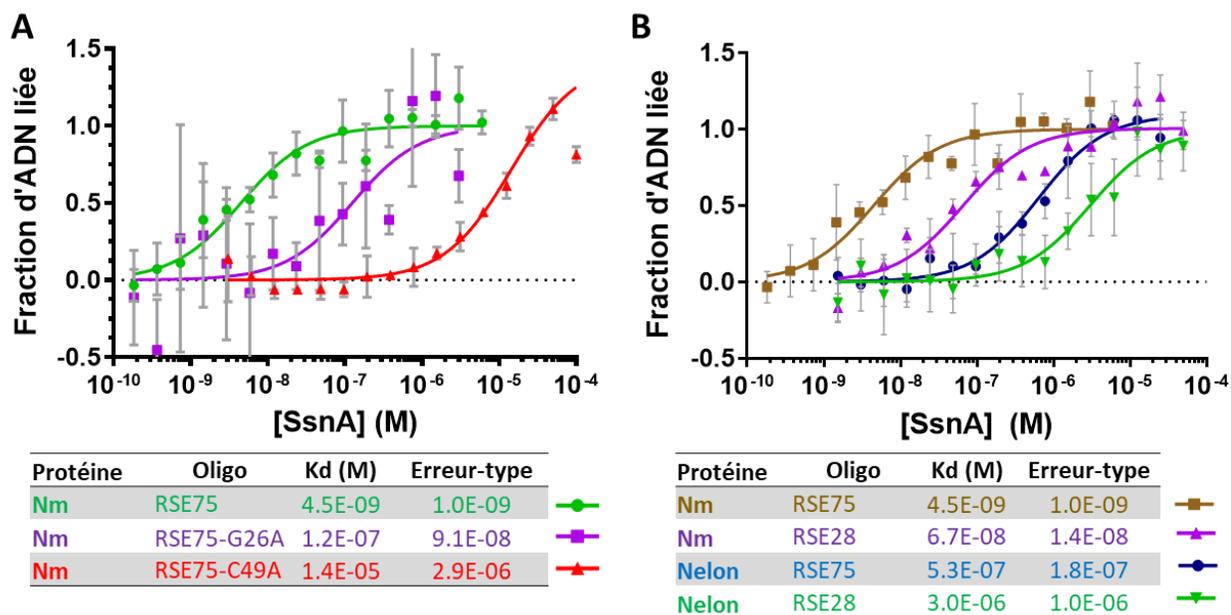


Figure 32 Détermination de l'affinité de SsnA pour son substrat

A. Thermophorèse (MST) d'oligonucléotides marqués à la cyanine-5 (cy5) en 5'. SsnA de *N. meningitidis* avec l'ADNsb RSE75 et deux variantes mutées en position G26 et C49. Les constantes d'affinité (K_D) sont indiquées dans le tableau en bas, dont les couleurs correspondent aux courbes présentées. La moyenne et l'erreur type d'au moins trois expériences indépendantes sont présentées. **B.** Expériences de thermophorèse avec les protéines de *N. meningitidis* (Nm) et de *N. elongata* (Nel) sur un oligonucléotide contenant la séquence de reconnaissance et de clivage (RSE75) et un autre contenant uniquement la séquence de reconnaissance (RSE28). La moyenne et l'erreur type d'au moins trois expériences indépendantes sont présentées.

À partir des constantes de dissociation calculées ici pour SsnA, il est possible de comparer son affinité et son efficacité à celle d'autres protéines interagissant uniquement avec l'ADNsb (**Tableau 9**). Selon nos données, SsnA possède une affinité particulièrement forte pour son substrat, caractérisée par une constante de dissociation (K_D) très faible. Cependant, sa constante de Michaelis-Menten (K_m) est plutôt élevée et sa constante catalytique (k_{cat}) est plutôt faible, ce qui suggère une efficacité (k_{cat}/K_m) relativement faible. Ce faible taux de conversion de substrat est cohérent avec la spécificité de SsnA pour une séquence étendue, et est typiquement retrouvé chez les endonucléases de type homing qui reconnaissent également de longues séquences d'ADN (McMurrough *et al.*, 2018; Turmel *et al.*, 1997). Concrètement, ces chiffres suggèrent que SsnA est capable de se fixer à son substrat d'ADNsb *in vivo* malgré l'activité protectrice d'autres protéines comme DprA et certaines Ssb.

Tableau 9 Constantes enzymatiques de protéines agissant sur l'ADNsb

Protéine	Espèce	Substrat	K_D (nM)	K_m	k_{cat} (s ⁻¹)	k_{cat}/K_m (M ⁻¹ s ⁻¹)	Source
SsnA	<i>N. meningitidis</i>	75 nt	4.5±1.0	4.68±1.45 μM	0.019±0.002	4084	Cette étude
DprA*	<i>H. pylori</i>	30 nt	370				(Wang <i>et al.</i> , 2014)
SsbA	<i>B. subtilis</i>	80 nt	0.2				(Yadav <i>et al.</i> , 2012)
SsbB	<i>N. gonorrhoeae</i>	90 nt dT ₉₀	100				(Jain <i>et al.</i> , 2012)
RecA	<i>E. coli</i>	13-63 nt	111-1008				(Warfel & LiCata, 2015)
ComFA	<i>S. pneumoniae</i>	50 nt	138				(Diallo <i>et al.</i> , 2017)
UvrA	<i>M. tuberculosis</i>	50 nt	~40				(Thakur <i>et al.</i> , 2020)
UvrB	<i>M. tuberculosis</i>	50 nt	550				(Thakur <i>et al.</i> , 2016)
UvrC	<i>M. tuberculosis</i>	50 nt	682				(Thakur <i>et al.</i> , 2021)
Cas12a	<i>Lachnospiraceae</i>	dC ₁₀₋₂₀		18-35 nM	2.3-2.8	0.8-1.3E08	(Nalefski <i>et al.</i> , 2021)
Nuc. S1	<i>A. oryzae</i>	ADNsb (tige-boucle)		6.8 μM	42	6.2E+06	(Li <i>et al.</i> , 2000)
RecJ	<i>D. radiodurans</i>	46 nt		90 nM	0.0258	2.86E+05	(Cheng <i>et al.</i> , 2016)

*DprA₅₋₂₁₇. À moins d'indication contraire, toutes les séquences testées sont de l'ADN simple-brin hétéropolymérique. Les constantes rapportées sont donc celles spécifiques à l'ADNsb.

CONCLUSIONS

Dans ce chapitre, la nucléase SsnA de *N. meningitidis* a été exprimée et purifiée sous forme recombinante avec deux tags différents, de même que trois mutants arborant des substitutions dans les sites actifs prédits. L'activité nucléase de ces protéines a été évaluée sur divers substrats, révélant que SsnA est une endonucléase spécifique d'ADN simple-brin. Cette enzyme n'a aucune ou très peu d'activité sur de l'ADN génomique ou plasmidique double-brin ainsi que sur de l'ARN. Elle est cependant capable de cliver spécifiquement de l'ADN simple-brin contenant sa séquence de reconnaissance, soit la NTS, d'où le nom SsnA (Specific Single-strand Nuclease). Confirmant son statut de ssDNase séquence-spécifique, SsnA est incapable de cliver de l'ADN simple-brin contenant une NTS en orientation inverse. Pour rappel, les NTS sont des éléments répétés palindromiques retrouvés presque exclusivement chez les *Neisseria*, dont l'abondance est fortement accrue chez l'espèce pathogène *N. meningitidis*.

De façon intrigante, bien que SsnA ait besoin d'une copie entière de la NTS sous forme simple-brin pour s'y fixer, son site de clivage se trouve quelques nucléotides en amont, suggérant un mécanisme de reconnaissance et de clivage complexe étant donné la minuscule taille de cette enzyme. SsnA dépend d'un cofacteur métallique tel que le magnésium, le nickel ou le manganèse pour exercer son activité nucléase, qui est optimale à 37°C. Nos analyses catalytiques révèlent que cette enzyme est très sensible et possède une forte affinité pour son substrat, mais son efficacité est plutôt faible. En comparaison, la protéine de *N. elongata* possède une plus faible affinité pour la NTS testée, ce qui est intrigant considérant que le génome de cette espèce abrite beaucoup moins de copies de cette séquence répétée. Il est possible qu'une séquence différente donne des résultats différents pour cette espèce.

3 CHAPITRE III - CARACTÉRISATION PHÉNOTYPIQUE DE SsnA

Les deux premiers chapitres de cette thèse démontrent que SsnA est une endonucléase d'ADN simple-brin clivant de façon spécifique à proximité de la séquence répétée NTS, retrouvée en de nombreuses copies dans le génome des *Neisseria*. La suite logique de ce projet est donc d'explorer les fonctions biologiques découlant de cette activité enzymatique. Est-ce que cette nucléase unique joue un rôle dans l'évolution du méningocoque, sa virulence, ou s'agit-il plutôt d'un élément génétique mobile sans utilité particulière pour l'hôte ? Quels processus biologiques font intervenir de l'ADN simple-brin, seul substrat de SsnA ? Ce troisième chapitre est dédié à la caractérisation fonctionnelle de SsnA chez *N. meningitidis*, ce qui implique la génération de plusieurs souches mutées ainsi qu'une panoplie de tests phénotypiques.

MATÉRIEL ET MÉTHODES

3.1 Souches de *Neisseria* et conditions de culture

Les souches de *Neisseria* utilisées dans cette étude sont détaillées dans le **Tableau 10**. *N. meningitidis* 8013 2C4.3 a servi de souche de référence pour la majorité de cette étude. Elle provient d'un isolat clinique appartenant au complexe clonal ST-18 du séro groupe C. Cette souche est encapsulée, adhérente et dotée de pili lui conférant des hauts niveaux de transformation par compétence naturelle (Nassif *et al.*, 1993). Dans plusieurs tests de transformation, la région *prlC-yebN-katA* a été substituée par celle de la souche MC58 afin d'augmenter l'homologie avec certains de nos plasmides intégratifs. Cette dernière n'a pas pu être utilisée directement puisqu'elle nécessite un vaccin supplémentaire.

Pour les tests d'infection *in vitro* et *in vivo*, les souches *Nm* LNP24198 et LNP20553 ont été utilisées. Ces dernières, issues du complexe clonal ST-11, ont été isolées respectivement du liquide cébrospinal et du sang de patients infectés, et sont considérées comme hyperinvasives en raison de leur taux élevé d'invasion, de leur virulence accrue chez la souris, et de leur association à un mauvais pronostic clinique (Deghmane *et al.*, 2009; Zarantonelli *et al.*, 2008).

Toutes les souches de *N. meningitidis* ainsi que *N. muscoli* ont été cultivées à 37°C en incubateur humide avec 5% de CO₂, sur milieu *Gonococcal Base* (GCB, Oxoid), complété après

autoclavage avec des suppléments Kellogg filtrés. En absence de milieu GCB commercial sans agar, la recette du fournisseur a été reproduite pour les milieux liquides.

Supplément Kellogg 1 (100x) : 40% glucose, 1% glutamine, et 0.002% thiamine pyrophosphate dans de l'eau, filtré sur 0.2 µm.

Supplément Kellogg 2 (1000x) : 0.05% Fe(NO₃)₃·9 H₂O dans de l'eau, filtré sur 0.2 µm.

Tableau 10 Souches de *Neisseria* utilisées

Espèce et souche	Mutants de <i>SsnA</i>	Description
<i>N. meningitidis</i> 8013 2C4.3	WT	WT : Souche sauvage.
	Δ <i>ssnA</i> ::Ery (KO)	KO : <i>ssnA</i> délété.
	<i>pilEp-ssnA</i> ::Km (OE)	OE : <i>ssnA</i> surexprimé par insertion ailleurs dans le génome sous un promoteur fort, sans enlever le gène natif.
	Δ <i>ssnA</i> ::Ery NMV_0632:: <i>pilEp-ssnA</i> ::Km (Compl)	Compl : <i>ssnA</i> délété puis réintroduit sous un promoteur fort ailleurs dans le génome (complémentée).
<i>Nm</i> 8013 2C4.3 <i>rpsL</i> *	WT, KO et Compl	Mutation K43R dans le gène <i>rpsL</i> , conférant une résistance à la streptomycine (Sm).
<i>Nm</i> 8013 2C4.3 <i>rpsL</i> * <i>yebOut_MC58</i>	WT, KO et Compl	La région <i>prlC-yebN-katA</i> est remplacée par la région équivalente de <i>Nm</i> MC58, avec une mutation ponctuelle dans <i>yebN</i> conférant une sensibilité au manganèse.
<i>Nm</i> 8013 2C4.3 <i>rpsL</i> * <i>yebOut_MC58</i> NTSmut (G26A)	WT, KO et Compl	Toutes les NTS de la région <i>prlC-yebN-katA</i> portent la mutation G26A.
<i>Nm</i> 8013 2C4.3 <i>rpsL</i> * <i>yebOut_MC58</i> NTSmut2 (G26A+C49A)	WT, KO et Compl	Toutes les NTS de la région <i>prlC-yebN-katA</i> portent la double mutation G26A et C49A.
<i>Nm</i> 8013 2C4.3 <i>rpsL</i> * <i>yebOut_MC58</i> Δ <i>yebN</i> ::Cm	WT, KO	<i>yebN</i> est remplacé par une cassette de résistance au chloramphénicol (Cm).
<i>Nm</i> 8013 2C4.3 <i>rpsL</i> * <i>yebOut_MC58</i> Δ <i>yebN</i> ::CmSTOP	WT, KO	<i>yebN</i> est remplacé par une cassette de résistance au chloramphénicol inactivée par un codon stop.
<i>Nm</i> 8013 2C4.3 <i>rpsL</i> * <i>yebOut_MC58</i> Δ <i>yebN</i> ::CmSTOP Km	WT, KO	<i>yebN</i> est remplacé par une cassette de résistance au chloramphénicol inactivée par un codon stop. Une résistance à la kanamycine est insérée ailleurs dans le génome.
<i>Nm</i> 8013 2C4.3 <i>gyrA</i> *		Mutation dans <i>gyrA</i> conférant une résistance à l'acide nalidixique (Nal).
<i>Nm</i> 8013 2C4.3 <i>recA6</i> ::Tet	WT, KO, Compl	Expression de <i>recA</i> inductible en présence d'IPTG.
<i>Nm</i> LNP24198 <i>lux_km</i>	WT, KO	Souche hyperinvasive luminescente exprimant l'opéron <i>lux</i> avec une résistance à la kanamycine (Km).
<i>Nm</i> LNP20553 <i>lux</i>	WT, KO, Compl	Souche hyperinvasive luminescente.
<i>N. musculi</i> DSM 101846 <i>yebOut_MC58</i> Δ <i>yebN</i> ::Cm		Espèce commensale. La région <i>prlC-yebN-katA</i> est remplacée par la région équivalente de <i>Nm</i> MC58. <i>yebN</i> est remplacé par une cassette de résistance au chloramphénicol (Cm).

Toutes les souches KO et Compl ont été générées avec les mêmes plasmides. *Nm* : *Neisseria meningitidis*

3.1.1 Génération des souches de *N. meningitidis* (Nm)

Ce projet de recherche a nécessité la génération de plusieurs dizaines de souches de *N. meningitidis* à l'aide de nombreux plasmides intégratifs. La plupart de ces plasmides ont été obtenus par clonage conventionnel impliquant l'amplification de régions génomiques par PCR, la génération d'extrémités compatibles par digestions enzymatiques, la ligation à divers vecteurs, et enfin la transformation dans *E. coli* DH5 α . Lorsque nécessaire, des modifications à ces plasmides ont été apportées par mutagenèse ciblée à l'aide de PCR par extension d'amorces (OE-PCR). Les constructions ont été vérifiées par PCR, par digestions enzymatiques et/ou par séquençage Sanger (Génome Québec). Pour certaines constructions plus complexes, des fragments d'ADN ont été synthétisés directement par BioBasic et GenScript, puis sous-clonés dans les vecteurs d'intérêt. Bien que la génération de ces constructions génétiques ait représenté un travail considérable et parfois un défi technique, les étapes menant à leur obtention ne sont pas détaillées afin d'alléger le contenu de cette thèse. Les amorces de clonage et de mutagenèse ainsi que les plasmides intermédiaires ne sont également pas indiqués pour les mêmes raisons.

Brièvement, les plasmides intégratifs ont été générés chez *E. coli* de façon à ce qu'ils contiennent des régions d'homologie au génome receveur d'au moins 500 pb de part et d'autre de toute insertion ou modification souhaitée. Lors de la transformation de *Neisseria* par compétence naturelle, ces régions guident l'intégration d'une partie du plasmide au chromosome de l'hôte par recombinaison homologue. Pour transformer les *Neisseria*, un stock congelé de bactéries est cultivé sur milieu solide, puis quelques colonies sont étalées sur une gélose contenant 10 mM de MgCl₂. Ensuite, 1 μ g ou 10 μ L de plasmide purifié sont ajoutés directement sur l'étalement, et le mélange bactéries-ADN est incubé environ 6 h. La croissance au niveau de la goutte d'ADN est alors prélevée, puis étalée sur une gélose contenant les marqueurs de sélection appropriés. Les colonies résistantes obtenues sont enfin repiquées, vérifiées par PCR, puis stockées dans du milieu contenant 15% de glycérol et conservées à -80°C.

Afin de générer les différents hybrides de *Nm* 8013 2C4.3 possédants la région *prlC-yebN-katA* de la souche MC58, une méthodologie de modification génétique développée et publiée récemment par mes collègues et moi-même a été utilisée (Nyongesa *et al.*, 2022a). D'abord, la mutation K43R a été introduite dans le gène *rpsL* par transformation avec de l'ADN génomique d'une autre souche, donnant une résistance à la streptomycine (Sm). Ensuite, le locus *prlC-yebN-katA* a été entièrement délété et remplacé par la cassette de sélection RPLK à l'aide du plasmide pYebOut::RPLK. La cassette RPLK a enfin été remplacée avec différentes constructions correspondant au locus natif de la souche MC58 avec diverses mutations ou marqueurs de

sélection. Une approche similaire a été utilisée pour générer la souche luminescente *Nm* LNP20553 lux.

Génération des mutants de *SsnA*

Dans une majorité des souches utilisées, des mutants d'expression de *ssnA* ont été générés. Les souches délétées (KO) ont été obtenues par transformation avec le plasmide pKO*ssnA*::Ery. Ce dernier remplace une partie centrale du gène *ssnA* par une cassette de résistance. Afin de s'assurer que les phénotypes obtenus avec les souches KO ne soient pas dus à des effets polaires associés à la délétion de *ssnA*, des souches complémentées ont été générées (Compl). Pour ce faire, le plasmide pOE*ssnA*::Km a été transformé dans les souches KO. Ce dernier réinsère *ssnA* ailleurs dans le génome sous le contrôle du promoteur constitutif fort de *pilE*. Ce plasmide a également été transformé dans certaines souches WT afin de générer des souches surexprimant *ssnA* (OE), qui encodent donc leur gène natif en plus de cette seconde copie surexprimée. Les souches ont systématiquement été vérifiées par PCR en amplifiant une région interne au gène *ssnA*, ainsi que par qPCR avec les mêmes amorces dans le cas des souches *N. meningitidis* 8013 2C4.3.

3.2 Courbes de croissance et transcriptomique

Courbes de croissance

Les souches *N. meningitidis* 8013 2C4.3 WT, KO et Compl ont été inoculées sur milieu GCB solide à partir de stocks congelés et incubés une nuit à 37°C avec 5% de CO₂. Chaque souche a ensuite été resuspendue à une DO₆₀₀ de 0.1 dans du milieu GCB liquide, et 2 ml de chaque suspension ont été transférés dans un tube de culture, puis incubés avec agitation à 300 rpm. La densité optique à 600 nm a alors été mesurée toutes les heures jusqu'à la phase stationnaire à l'aide du spectrophotomètre SmartSpec Plus (Bio-Rad).

RT-qPCR de *SsnA*

À partir de cultures fraîches des mêmes souches en triplicatas, des tapis de bactéries ont été étalés sur milieu GCB solide puis incubés exactement 6 heures afin de capturer l'expression génique en phase exponentielle de croissance. L'ARN total a alors été extrait de chaque culture à l'aide d'un kit RNeasy (Qiagen). Un µg d'ARN par échantillon a par la suite été traité à la DNase I (Invitrogen) pour éliminer l'ADN génomique contaminant, puis rétrotranscrit à l'aide d'amorces aléatoires (Invitrogen) et de la rétrotranscriptase RevertAid H Minus RT (Thermo). Des contrôles

sans RT ont été réalisés pour vérifier l'absence d'ADN génomique. Des qPCR ont enfin été réalisées sur 10 ng d'ADNc avec le kit PowerUp SYBR Green (Thermo) sur un appareil StepOne Plus (Applied Biosystems). Des amorces amplifiant une région de 175 pb au centre du gène *ssnA*, ainsi que des amorces ciblant l'endogène de référence *gyrA* ont été utilisées. La méthode du $\Delta\Delta CT$ a été utilisée pour calculer les changements d'ordre d'expression (*fold changes*).

RNA-seq

Cette analyse ayant été faite avant la génération de la souche complémentée de *N. meningitidis* 8013 2C4.3, elle a été réalisée sur les souches WT, KO et OE. Pour rappel, la souche OE exprime son gène *ssnA* natif en plus d'une seconde copie du même gène ailleurs dans le génome, cette fois sous un promoteur fort. Des extractions d'ARN ont été réalisées en triplicatas de la même façon que pour les qPCR, puis envoyées à Génome Québec pour séquençage par NovaSeq (Illumina). Entre 5 et 10 millions de lectures ont été obtenus par échantillon. Les séquences ont été traitées sur le serveur Galaxy (Afgan *et al.*, 2018). Brièvement, ces dernières ont été filtrées par l'outil Groomer v1.0.4, tronquées en 3' avec Trimmer v1.0.0, puis alignées sur le génome de référence avec Bowtie2 v2.2.6.2 et enfin compilées avec htseq-count v0.6.1 (Blankenberg *et al.*, 2010; Langmead & Salzberg, 2012). L'analyse d'expression différentielle de gènes a par la suite été réalisée à l'aide du serveur TCC-GUI par les algorithmes EdgeR et DESeq 2, donnant des résultats similaires (Love *et al.*, 2014; Robinson *et al.*, 2010; Su *et al.*, 2019).

3.3 Tests de transformation

3.3.1 Transformations avec de l'ADN purifié

Les tests de transformation avec de l'ADN purifié ont tous été réalisés avec le même protocole. Les plasmides utilisés pour ces tests ainsi que pour le reste de cette étude sont détaillés dans le **Tableau 11**. Ce tableau, non exhaustif, n'inclut pas toutes les constructions intermédiaires ayant servi aux nombreux clonages et à la génération de certaines souches, et se concentre sur ceux pertinents aux résultats montrés afin d'être plus concis. Pour rappel, la transformation des *Neisseria* se fait à l'aide de plasmides intégratifs qui recombinent au chromosome de l'hôte via la recombinaison RecA qui reconnaît et apparie les régions d'homologie.

Tableau 11 Liste des plasmides pour *Neisseria* utilisés dans cette étude

Usage	Plasmides	Description	Source
Génération de souches de <i>Nm</i>	pKO <i>ssnA</i> ::Ery	Pour déléter <i>ssnA</i> de <i>N. meningitidis</i> 8013.	Cette étude
	pOE <i>ssnA</i> ::Ery	Pour surexprimer ou compléter <i>ssnA</i> .	Cette étude
	pYebOut::RPLK	Permet de substituer la région entière <i>prc-yebN-katA</i> par la cassette de contre-sélection RPLK.	Cette étude
	pYebOut_MC58	Pour générer les souches <i>Nm</i> yebOut_MC58, avec une mutation dans <i>yebN</i> conférant une sensibilité au manganèse.	Cette étude
	pYebOut_KO <i>yebN</i> ::Cm	Pour générer <i>Nm</i> et <i>N. musculli</i> Δ <i>yebN</i> ::Cm utilisés dans les tests de transformation par coculture.	Cette étude
	pVD300 <i>recA6</i>	Pour générer une souche avec la recombinaise RecA inductible à l'IPTG.	Addgene #37062
Tests de nucléase	pGEM ::Km	Plasmide sans NTS utilisé pour les tests de nucléase en gel d'agarose.	Stock du labo
	pKOYebN ::Sp	Plasmide avec plusieurs NTS de la région <i>yebN</i> de <i>Nm</i> MC58, dont celle utilisée comme contrôle pour les tests de nucléase. Aussi utilisé pour les tests de transformation.	Stock du labo
Tests de transformation	pKO <i>opaA</i> ::Cm	Plasmide intégratif avec NTS, modifiant le locus <i>opaA</i> .	(GenScript)
	pKONMV_0504 ::Cm	Plasmide intégratif avec NTS, modifiant le locus <i>NMV_0504</i> .	(GenScript)
	pKO <i>bolA</i> ::Cm	Plasmide intégratif sans NTS, modifiant le locus <i>bolA</i> .	(GenScript)
	pKO <i>lgtFrfaK</i> ::Cm	Plasmide intégratif sans NTS, modifiant le locus <i>lgtF-rfaK</i> .	(GenScript)
	pOutNTSwt	Plasmide intégratif remplaçant la région riche en NTS du locus <i>prc-yebN-katA</i> par celle de la souche <i>Nm</i> MC58.	Cette étude
	pOutNTSmut	Même que le précédent, mais avec la mutation G26A dans chaque copie de la NTS de la région <i>prc-yebN-katA</i> .	(Bio Basic)
	pInNTSwt	Similaire à pOutNTSwt, mais les régions d'homologies sont limitées aux régions intergéniques entourant <i>yebN</i> contenant les NTS.	Cette étude
	pInNTSmut	Même que le précédent, avec la mutation G26A dans chaque NTS.	(Bio Basic)
	pInNTSmut2	Même que le précédent, avec la double-mutation G26A et C49A dans chaque NTS.	(Bio Basic)

Les sources entre parenthèses indiquent les constructions qui ont été synthétisées pour cette étude, autrement elles ont été générées par clonage conventionnel. pVD300*recA6* a été gracieusement fourni par Hank Seifert (Seifert, 1997). *Nm* : *N. meningitidis*.

Des cultures fraîches des différentes souches de *N. meningitidis* ont été resuspendues à une DO₆₀₀ de 1 (2.5×10^8 CFUs/ml) dans du milieu GCB supplémenté avec 10 mM de MgCl₂. Cent µl de ces suspensions ont été transférés dans une plaque 96-puits en plusieurs réplicats, auxquels ont été ajoutés 200 ng d'ADN génomique ou plasmidique, purifiés avec des kits QIAamp et QIAprep respectivement (Qiagen). Les plasmides ont été préalablement linéarisés avec une enzyme clivant dans la région du vecteur non homologue au chromosome afin de favoriser la recombinaison et l'intégration des régions homologues. Ces réactions de transformation ont ensuite été incubées 2 heures à 37°C sans agitation pour permettre à l'ADN transformant d'être capté par les pili, importé au cytoplasme puis intégré au chromosome, et donner le temps aux différents marqueurs de sélection d'être exprimés. Enfin, des dilutions en série de chaque réaction ont été réalisées dans du milieu GCB sans magnésium, après quoi 30 µl des dilutions appropriées ont été étalés sur géloses GCB avec et sans antibiotique, puis incubées 24 h. Les colonies (CFUs) ont alors été dénombrées, et le nombre de transformants ayant acquis la résistance voulue a été normalisé sur le nombre total de CFUs pour représenter le taux de transformation propre à chaque réaction. Les antibiotiques utilisés sont détaillés dans le (Chapitre II). Pour certaines transformations utilisant l'allèle natif *yebN* comme marqueur de sélection plutôt qu'une résistance à un antibiotique, la sélection des transformants a été faite sur milieu GCB supplémenté avec 1 mM de MnCl₂.

Pour chaque expérience, au minimum deux réactions parallèles de transformation ont été réalisées pour toutes les combinaisons de souches et d'ADN, et au moins trois répétitions indépendantes de chaque expérience ont été réalisées. Les données ont été analysées statistiquement et illustrées avec le logiciel GraphPad Prism 7.00. Les tests statistiques sont indiqués dans la légende de chaque figure.

3.3.2 Transformations par cocultures

***N. meningitidis* comme donneur**

Pour ces tests, les souches WT et SsnA KO de *Nm ΔyebN::Cm* ont servi de souches donneuses, et celles de *Nm ΔyebN::CmSTOP_Km* ont servi de souches receveuses. Donneurs et receveurs réfèrent ici au gène de résistance au chloramphénicol, qui est inséré à la place du locus *yebN* dans une région riche en NTS. Puisqu'une simple mutation non-sens différencie les allèles de résistance au chloramphénicol des souches donneuses et receveuses, la fréquence de cet évènement est statistiquement beaucoup plus élevée que la fréquence de transfert du gène entier

de résistance à la kanamycine de la souche receveuse vers la souche donneuse, qui nécessite une large insertion par double-recombinaison homologue. Cette méthodologie permet donc de visualiser l'échange unidirectionnel de l'allèle voulu (Cm).

Brièvement, des cultures fraîches des souches donneuses et receveuses ont été resuspendues à une DO₆₀₀ de 0.1 dans du GCB+10 mM MgCl₂, et 50 µl de chacune ont été mélangées en plusieurs réplicats dans une plaque 96-puits. Les mélanges ont été incubés 3 h à 37°C avec 5% de CO₂ sans agitation, puis dilués et étalés sur milieu avec kanamycine et chloramphénicol. Pour ces tests, les taux de transformations sont présentés par rapport au volume de réaction. Comme pour les tests précédents, plusieurs cocultures indépendantes ont été réalisées pour chaque combinaison donneur-receveur, et ce lors d'au moins trois expériences indépendantes.

***N. musculi* comme donneur**

Un protocole identique au précédent a été utilisé, mais cette fois en utilisant comme donneur *N. musculi* Δ*yebN*::Cm SsnA WT, puisque l'expression de *ssnA* dans la souche donneur n'a pas d'influence sur le taux de transformation de la souche receveuse. Cette souche donneuse a été générée par transformation avec le plasmide pYebOut_KO*yebN*::Cm, et possède donc des régions d'homologie au locus *prlC-katA* parfaitement identiques à la région correspondante des souches receveuses.

3.4 Variation antigénique de *pilE*

Pour ces tests, le plasmide pVD300*recA6* a été transformé dans *N. meningitidis* 8013 afin d'obtenir une expression inducible à l'IPTG de la recombinase RecA, essentielle au processus de variation antigénique de *pilE* par recombinaison homologue avec les copies silencieuses *pilS1-5*. Puisque RecA est également essentielle à l'intégration d'ADN transformant, les mutants SsnA KO et Compl ont ensuite été générés en présence d'IPTG lors de la transformation. Un contrôle de transformation sans IPTG a été inclus et n'a révélé aucun transformant, confirmant l'absence d'expression de RecA sans ajout d'inducteur.

Sans ADN transformant

Pour ce premier test, les souches WT, KO et Compl ont été inoculées sur GCB sans IPTG de façon à obtenir des colonies séparées, après quoi quatre colonies de chaque souche ont été diluées séparément dans du GCB liquide. Ces colonies ont servi de réplicats et ont été soumises aux mêmes manipulations de façon simultanée, afin d'évaluer les taux de variation antigénique

des cellules filles provenant de plusieurs clones parentaux. Après dilution, chaque suspension a été étalée sur milieu GCB supplémenté avec 1 mM d'IPTG de façon à obtenir entre 500 et 1000 colonies filles, puis incubée 16 heures à 37°C avec 5% de CO₂. La totalité des colonies obtenues, représentant une diversité aléatoire d'allèles *pilE*, a ensuite été regroupée par clone parental, et l'ADN génomique total de ces mélanges a été extrait à l'aide d'un kit QIAamp (Qiagen).

Une PCR haute-fidélité (Phusion, NEB) a ensuite été réalisée sur ces échantillons avec des amorces contenant une séquence adaptateur, qui se lie aux extrémités invariables de *pilE*. La pureté et l'intégrité des produits de PCR ont été vérifiées sur gel et ces derniers ont été purifiés sur colonne (kit QIAquick, Qiagen), puis quantifiés à l'aide d'un spectrophotomètre NanoDrop 2000 (Thermo). Grâce à la courte séquence adaptateur ajoutée aux extrémités des produits de PCR, des codes-barres uniques ont pu être ajoutés à 1 ng du produit de chaque clone parental grâce au *PCR Barcoding Kit* (Nanopore), par une seconde réaction de PCR utilisant cette fois la polymérase LongAmp Taq (NEB). Après vérification sur gel, chaque produit de PCR étiqueté a ensuite été purifié à l'aide de 0.8 volume de billes Sera-Mag Select (Cytiva) en suivant le protocole du fournisseur. Cette fois-ci, pour garantir la précision des mesures et une bonne normalisation des échantillons avant le séquençage, des quantifications ont été réalisées sur un appareil Qubit 4 avec le kit dsDNA HS (Invitrogen). L'ensemble des douze produits de PCR avec des codes-barres uniques a alors été regroupé de façon équimolaire puis attaché à un adaptateur de séquençage par ligation en suivant les recommandations du fournisseur.

Réduite ainsi à un seul échantillon par multiplexage, la librairie a été séquencée par MinION Mk1C (Nanopore) jusqu'à obtenir 2 millions de lectures. Ces lectures ont été déchiffrées à l'aide du réseau neuronal Guppy en mode haute-fidélité, divisées par étiquette (code-barre), puis tronquées pour supprimer les adaptateurs et les codes-barres.

Pour l'analyse des variants, nous avons été contraints de créer notre propre pipeline bio-informatique puisqu'une telle expérience n'avait encore jamais été réalisée par séquençage Nanopore. La variation antigénique modifie la séquence du gène *pilE* par recombinaison homologue avec diverses régions des copies silencieuses *pilS1-5*. Ces événements de recombinaison peuvent échanger aussi peu qu'un seul nucléotide, mais le haut taux d'erreur associé au séquençage par Nanopore rend la détection de ces microchangements ardue (Vink *et al.*, 2012). Afin de détecter des petits événements de recombinaison tout en limitant les faux-positifs dus aux erreurs de séquençage, une matrice a été générée en divisant les gènes *pilE* et *pilS1-5* en courtes séquences de 12 nucléotides (k-mers/12-mers). Les lectures associées à chaque clone parental ont ensuite été alignées à cette matrice avec un seuil d'identité à 100%.

Les séquences s'alignant sur des 12-mers de *pilS* mais pas de *pilE* correspondent à des variants. Cette approche permet d'assigner une grande diversité de variants possibles du gène *pilE*, en plus d'identifier précisément l'origine et la position des événements de recombinaisons avec les différents *pilS*.

Avec ADN transformant

Afin d'étudier l'impact du mécanisme de transformation par compétence naturelle sur la variation antigénique de *pilE*, une méthodologie semblable a été utilisée, avec quelques modifications. Cette fois-ci, quatre colonies parentales pour chaque souche n'ont pas été resuspendues et étalées immédiatement sur milieu avec IPTG, mais plutôt étalées de nouveau sur milieu sans IPTG, incubées 6 h pour atteindre la phase exponentielle, puis resuspendues à une DO_{600} de 1 dans du GCB supplémenté avec 1 mM d'IPTG, 10 mM de $MgCl_2$, et 1.3 μg d'ADN génomique total provenant de *N. meningitidis* 8013 2C4.3 WT. Ces réactions de transformation, supposées fournir aux souches receveuses des copies exogènes de *pilS1-5*, ont ensuite été incubées 1h à 37°C avec 5% de CO_2 sans agitation. Chaque réaction a enfin été diluée, puis étalée sur milieu sans IPTG pour prévenir la possibilité de variation antigénique post-transformation. Le reste du protocole est identique au précédent, mais 10 millions de lectures ont été obtenues plutôt que 2 millions en raison d'une plus faible perte de matériel lors des différentes étapes de purification et de préparation.

3.5 Tests de virulence

3.5.1 Infections *in vitro*

THP-1

La lignée de monocytes THP-1 (ATCC TIB-202) a été cultivée dans du RPMI 1640 GlutaMAX avec 10% de SVF de façon à avoir des monocouches en plaque 24-puits. 17.5 h avant l'expérience, les monocouches ont été différenciées en macrophages par l'ajout de 20 ng/ml de PMA. Des cultures fraîches de *N. meningitidis* LNP20553 lux SsnA WT, KO et Compl ont été resuspendues dans du RPMI+SVF, quantifiées et normalisées selon leur DO_{600} , puis ajoutées à chaque puits de façon à avoir cinq fois plus de bactéries que de cellules (multiplicité de l'infection MOI=5). Après 1h30 d'incubation à 37°C avec 5% de CO_2 , les puits ont été lavés au PBS, le milieu a été changé et la luminescence a été acquise à l'aide d'un lecteur de plaque Victor3 (Perkin-Elmer). Les puits dédiés aux tests d'invasion ont immédiatement après été traités avec

200 µg/ml de gentamicine pendant 30 min. Des mesures de luminescence ont été prises toutes les heures après avoir lavé les puits et changé le milieu, jusqu'à 8.5 heures post-infection (PI), après quoi l'expérience a été poursuivie jusqu'à 24 h PI sans lavages et changements de milieux, sauf avant l'ultime mesure. Pour les puits dédiés aux tests d'invasion, une concentration bactériostatique de gentamicine (10 µg/ml) a été ajoutée à 8.5 h PI pour limiter la croissance extracellulaire et la réinfection des cellules et ainsi observer la survie et la multiplication intracellulaire liée à l'infection primaire. Cette expérience a été réalisée en triplicatas à deux reprises indépendantes.

Calu-3

La lignée de cellules épithéliales pulmonaire Calu-3 (ATCC HTB-55) a été cultivée dans du DMEM GlutaMAX (GIBCO) avec 10% de SVF de façon à obtenir des monocouches en plaque 24-puits. Des cultures fraîches de *N. meningitidis* LNP20553 lux SsnA WT, KO et Compl ont été resuspendues dans du DMEM+SVF, quantifiées et normalisées, puis ajoutées à chaque puits de façon à avoir une MOI=40. Après 3 h d'incubation, les puits dédiés aux tests d'adhésion ont été lavés trois fois avec du PBS, puis traités avec 1% de saponine dans du DMEM+SVF pendant 10 min. Après 6 h et 24 h, les puits dédiés aux tests d'invasion ont été lavés au PBS, traités avec 200 µg/ml de gentamicine pendant 30 min à 37°C, puis lavés encore et traités à la saponine. Les suspensions ont alors été diluées en série puis étalées sur milieu GCB et incubées 16 heures pour procéder au décompte des colonies (CFUs) totales associées aux cellules (adhésion) et des colonies intracellulaires (invasion). Cette expérience a été réalisée en triplicatas à deux reprises indépendantes.

3.5.2 Infections de souris

Les expériences animales ont été réalisées en conformité avec les recommandations du Centre national de biologie expérimentale dans une pièce en niveau de confinement 2, après validation éthique des protocoles utilisés. Les souris proviennent de Charles River.

Des souris BALB/c femelles de six semaines ont été préinfectées avec 250 unités de formation de plaque (PFU) du virus de l'Influenza A par voie intranasale afin de les rendre susceptibles à l'infection invasive par *N. meningitidis*, qui est spécifique à l'humain (Alonso *et al.*, 2003). Sept jours plus tard, des cultures fraîches de *N. meningitidis* LNP24198 lux_{km} (SsnA WT et KO) et LNP20553 lux (SsnA WT, KO et Compl) ont été resuspendues à une DO₆₀₀ de 0.05 dans du PBS supplémenté avec 20 mg/ml de transferrine humaine chargée en fer (Sigma). 100 µl de ces suspensions ont ensuite été injectés par voie intrapéritonéale, causant des infections

systemiques. La luminescence associée aux bactéries a été mesurée simultanément dans les souris immédiatement après l'infection et à 6, 12, 24 et 32 h PI, grâce à l'appareil Ivis Lumina III (Perkin-Elmer). Le signal luminescent associé aux souris entières a été quantifié et normalisé pour chaque mesure via le logiciel LivingImage v4.7.3 (Perkin-Elmer). Pour les souches LNP24198 lux_Km, un total de dix souris ont été infectées par souche au cours de deux expériences indépendantes. Pour les souches LNP20553 lux, cinq souris ont été infectées par souche. Sur toutes ces expériences, deux souris provenant de différentes souches ont été exclues des analyses car l'infection n'a pas été productive due à une mauvaise injection, se traduisant par l'absence totale de signal lumineux à 6 h PI.

3.6 Tests statistiques

Les tests statistiques sont décrits dans la légende de chaque figure concernée, et proviennent systématiquement d'au minimum trois expériences indépendantes. Ces analyses ont été sélectionnées en fonction du nombre de points, de la variance, de la distribution normale et du nombre de conditions comparées pour chaque expérience. Tous les tests statistiques ont été réalisés avec le logiciel GraphPad Prism 7.00.

RÉSULTATS ET DISCUSSION

3.7 Génération et validation des mutants

Dans le but de tester l'importance de la protéine SsnA chez *N. meningitidis*, nous avons généré des mutants dans plusieurs souches de cette espèce pathogène, dont la souche de référence *Nm* 8013 2C4.3. Celle-ci provient d'un isolat clinique encapsulé et adhérent doté de pili lui conférant des hauts niveaux de transformation par compétence naturelle (Nassif *et al.*, 1993). À l'aide de cette compétence naturelle, nous avons supprimé le gène *ssnA*, générant des souches knock-out (KO), puis l'avons réintroduit ailleurs dans le chromosome sous un promoteur fort pour générer des souches complémentées (Compl). Les mutants ainsi générés ont été vérifiés par PCR ainsi que par RT-qPCR (**Figure 33A-B**). L'ajout d'un promoteur fort devant le gène *ssnA* dans la souche complémentée se traduit par une importante surexpression, validant cette construction et suggérant que ce gène est faiblement exprimé dans la souche sauvage. L'influence de ces modifications génétiques sur la croissance en conditions optimales de *N. meningitidis* a également été évaluée (**Figure 33C**). Ces vérifications indiquent que SsnA n'est pas une protéine essentielle et que sa délétion et surexpression n'a aucun effet sur la croissance *in vitro*, ce qui facilite la suite des tests phénotypiques.

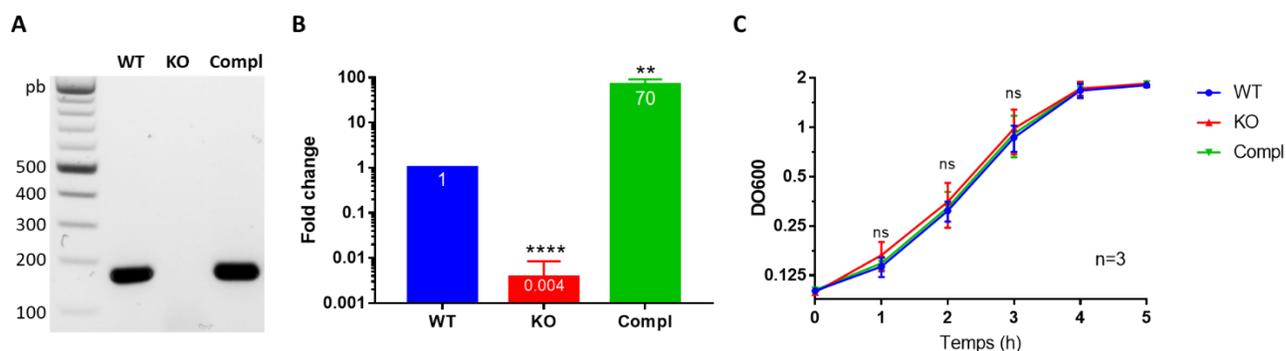


Figure 33 Validation des mutants de SsnA chez *N. meningitidis* 8013

A. PCR résolue sur gel d'agarose des mutants de SsnA chez *Nm* 8013 2C4.3, réalisée avec les amorces SsnA_F/R qui amplifient une courte région à l'intérieur du gène. **B.** RT-qPCR réalisée avec les mêmes amorces à partir d'ARN extrait des souches précédentes en phase exponentielle de croissance. La moyenne et l'erreur type de l'expression relative de *ssnA* provenant de trois expériences indépendantes sont montrées et comparées à l'expression chez la souche WT par des tests de Student non-pairés. **C.** Courbes de croissance des trois souches de *Nm*, réalisées par mesure de la densité optique en milieu liquide avec agitation. Les moyennes et erreur-types de trois expériences indépendantes sont présentées, et comparées statistiquement par ANOVA unilatérale avec comparaison multiple de Tukey. Ns non significatif, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Au cours de ce projet, des mutants similaires de SsnA ont été générés dans de nombreuses autres souches dérivées de *Nm* 8013 2C4.3, de même que dans d'autres souches comme *Nm* LNP20553 et *Nm* LNP24198, ainsi que dans d'autres espèces comme *N. musculli*. Toutes ces constructions ont été vérifiées *a minima* par PCR, mais ces vérifications ne sont pas incluses dans cet ouvrage afin d'éviter les redondances.

3.8 Étude transcriptomique

Nous avons démontré par des tests biochimiques que la protéine purifiée SsnA interagit de façon spécifique avec la NTS. Or, ces éléments répétés sont retrouvés exclusivement dans les régions intergéniques, bien souvent immédiatement en amont et en aval de gènes, et affichent une structure en épingle à cheveux. Cette structure et distribution particulière suggère que ces NTS puissent potentiellement modifier l'expression génique en agissant comme terminateurs transcriptionnels ou en influençant les régions promotrices. Puisque de telles fonctions sont déjà associées à plusieurs éléments répétés des *Neisseria*, nous avons voulu vérifier cette possibilité. En plus de procéder à des tests phénotypiques plus ciblés, nous avons donc réalisé une étude exploratoire de transcriptomique par séquençage d'ARN (**Figure 34**). Pour ces analyses, nous avons inclus une souche surexprimant SsnA (OE) à la place de la souche complémentée. La seule différence entre celles-ci est que le gène natif n'a pas été supprimé dans la souche OE, elle exprime donc deux copies de celui-ci, dont une avec un promoteur fort et l'autre avec son promoteur intrinsèque.

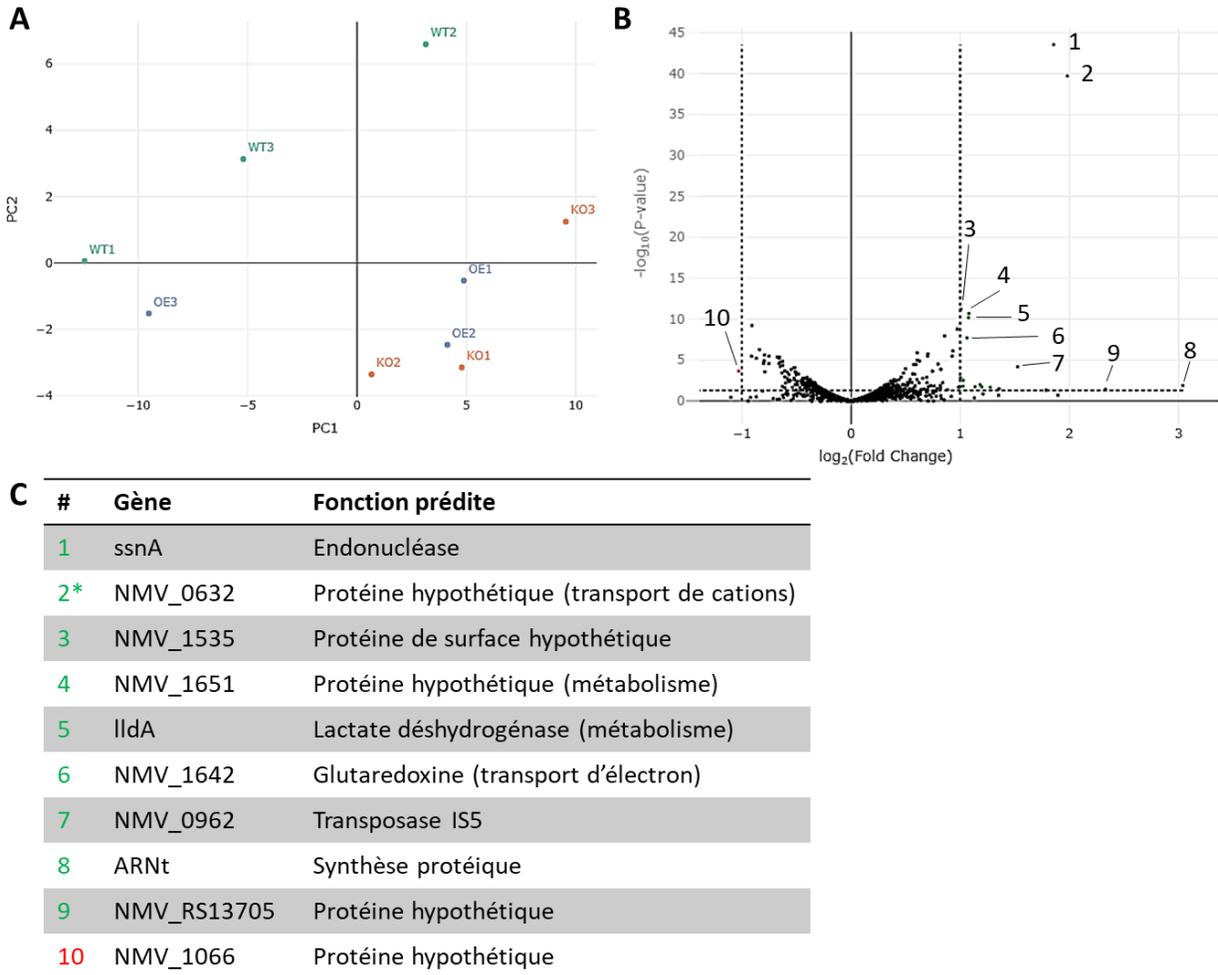


Figure 34 Analyse transcriptomique de mutants de *N. meningitidis*

A. Analyse par composante principale (PCA) des neuf transcriptomes, où OE désigne la souche surexprimant *ssnA* et KO désigne la souche n'exprimant pas ce gène. Trois extractions indépendantes d'ARNm pour chacune des souches ont été analysées. **B.** Graphique en volcan montrant les changements d'expression et les p-values des gènes lorsque les souches OE sont comparées aux souches KO. Les dix points affichant la plus grande différence significative sont numérotés et identifiés. Des lignes pointillées délimitent les seuils généralement acceptés comme significatifs d'un point biologique (*fold change* ≥ 2) et statistique ($p\text{-value} \leq 0.05$). **C.** Tableau rapportant les dix gènes les plus différenciellement exprimés entre *Nm* OE et *Nm* KO, avec leurs fonctions prédites. *NMV_0632 correspond au locus dans lequel *SsnA* a été introduit avec un promoteur fort. Sa surexpression similaire à celle de *SsnA* est donc tout à fait attendue. Vert : Surexpression, Rouge : Sous-expression.

Cette étude transcriptomique n'a détecté que très peu de différences d'expression entre les trois souches testées. Une analyse par composante principale révèle d'ailleurs une bonne dispersion des isolats d'ARNm, même au sein d'une seule souche, suggérant que les différences obtenues viennent principalement de variabilité biologique plutôt que de réels phénotypes reproductibles associés à *SsnA* (**Figure 34A**). Même en comparant les deux souches avec la plus grande différence d'expression de *ssnA* (OE vs KO), très peu de gènes sont identifiés comme significativement modulés d'un point de vue biologique et statistique (**Figure 34B**). Parmi les dix

candidats dont l'expression varie le plus, neuf sont surexprimés dans la souche OE. Les deux se démarquant le plus parmi ceux-ci correspondent à *ssnA* et à NMV_0632, ce qui était attendu. En effet, le locus NMV_0632 dans la souche OE est le site d'insertion de la copie de *ssnA* combinée à un promoteur fort. Parmi les autres gènes modulés, on retrouve principalement des protéines hypothétiques, une transposase et un ARN de transfert (**Figure 34C**). Cependant, leur relativement faible score statistique ou faible changement d'expression sont insuffisants pour suggérer un réel lien avec la protéine SsnA. Concrètement, ces résultats démontrent que SsnA n'agit pas sur la transcription d'autres gènes dans les conditions testées, malgré son association fonctionnelle avec la séquence répétée NTS.

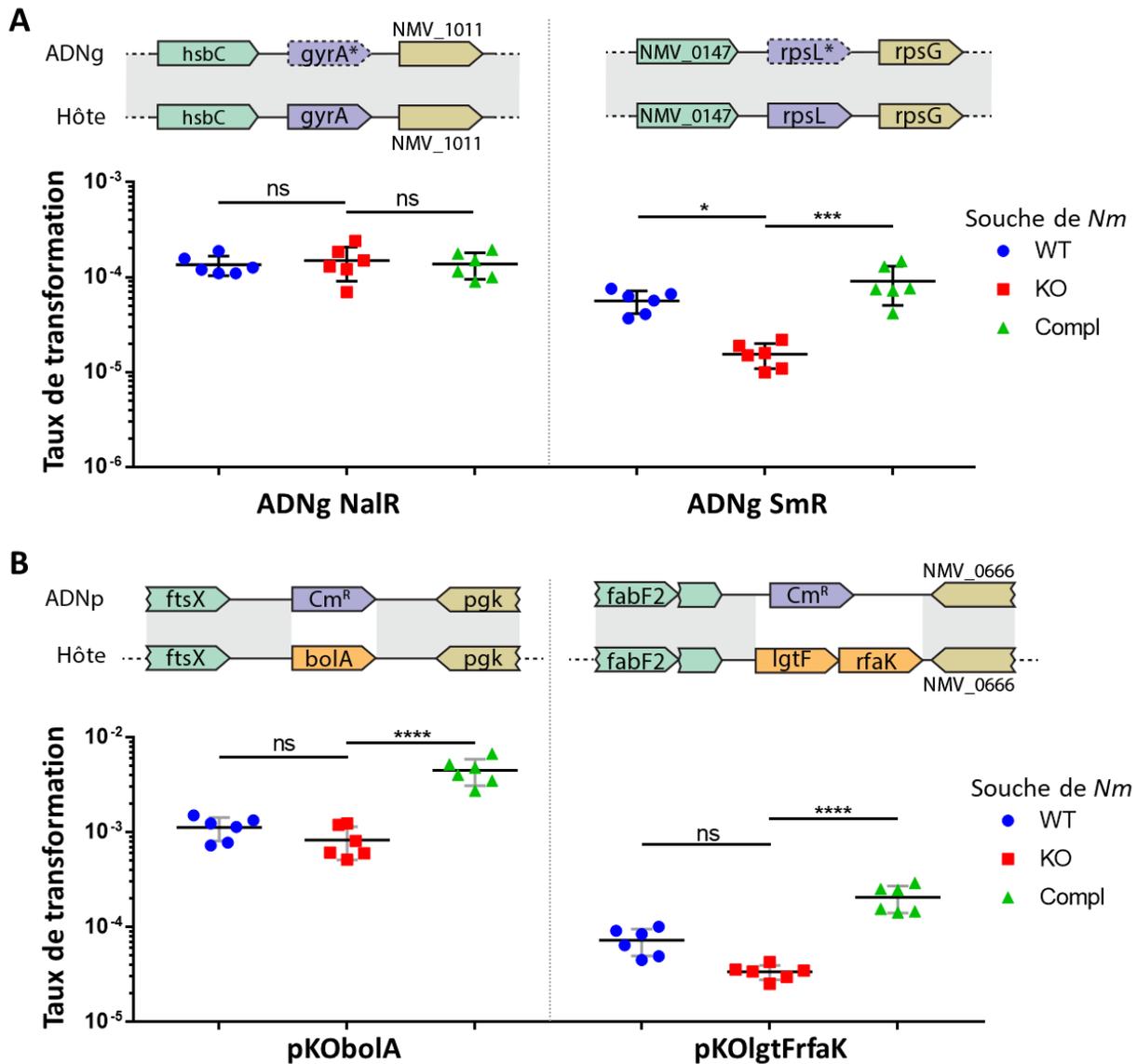
3.9 Rôle dans la transformation par compétence naturelle

Les processus biologiques faisant intervenir de l'ADN simple-brin sont relativement peu nombreux. On y retrouve les trois R de l'ADN, soit la réplication, la réparation et la recombinaison, ainsi que les trois mécanismes majeurs de transferts horizontaux de gènes (HGT), soit la conjugaison, la transduction et la transformation. Nous nous sommes donc entre autres penchés sur la transformation et la recombinaison. Grâce à la forte compétence naturelle constitutive de ce genre bactérien, la transformation est reconnue comme le principal mécanisme de diversité génétique chez ces espèces (Rotman & Seifert, 2014). Bien que l'ADN environnemental soit majoritairement sous forme double-brin, la machinerie de transformation des *Neisseria* ne transporte qu'un seul brin à son cytoplasme, qui est ensuite recombiné au chromosome de l'hôte s'il partage une homologie suffisante (Hamilton & Dillard, 2006). Plusieurs protéines de ces deux mécanismes interagissent spécifiquement avec l'ADNsb, que ce soit pour le protéger des nucléases cellulaires ou encore pour recruter d'autres effecteurs.

Dans le but de tester simultanément le rôle de SsnA dans la compétence naturelle et la recombinaison, nous avons soumis diverses souches de *N. meningitidis* à de nombreux tests de transformation quantitatifs. Ces derniers mesurent la fréquence d'intégration au chromosome de gènes de résistance aux antibiotiques agissant comme marqueurs de sélection. Ces marqueurs s'intègrent de façon ciblée au génome de l'hôte par le biais de plasmides intégratifs composés de régions homologues au chromosome. En changeant de locus cible ou en introduisant des mutations dans les régions d'homologie, nous sommes en mesure d'étudier l'impact des séquences répétées NTS sur la transformation et la recombinaison.

3.9.1 Transformation d'ADN sans NTS dans les zones de recombinaison

Afin de juger de l'influence de SsnA sur la transformation de façon générale indépendamment de la présence de NTS, nous avons débuté nos tests en utilisant comme substrat de l'ADN génomique total (ADNg) d'une souche similaire (**Figure 35A**). Ces substrats proviennent de clones de *N. meningitidis* 8013 résistants à l'acide nalidixique (Nal) ou à la streptomycine (Sm), dus à des mutations dans les gènes *gyrA* et *rpsL* respectivement. Leur homologie au génome receveur est donc parfaite excepté pour ces mutations, ce qui permet leur intégration par une seule recombinaison homologue. Aucune NTS n'est retrouvée à proximité de ces régions. Lorsque les mutants de SsnA sont transformés avec l'ADNg conférant une résistance au Nal, aucun phénotype n'est observé. Cependant, des différences significatives sont obtenues lorsque ces mêmes souches sont transformées avec le second ADNg (Sm^R). En effet, on détecte une faible diminution du taux de transformation dans la souche n'exprimant pas SsnA (KO) et une augmentation de ce taux dans la souche complétée (Compl). Seuls, ces résultats ambigus ne permettent pas de conclure quant à l'influence de SsnA sur la transformation, mais valident toutefois le potentiel de la méthodologie utilisée.



Afin de mieux contrôler les régions d'homologie déterminant l'intégration de l'ADN donneur au génome de l'hôte, nous avons ensuite procédé à des tests de transformation similaires, mais cette fois en utilisant des plasmides intégratifs plutôt que de l'ADN génomique (**Figure 35B**). Ces derniers possèdent une cassette de résistance au chloramphénicol (Cm), flanquée de séquences de 500 pb parfaitement homologues au génome receveur, qui dictent son site d'intégration. Une double recombinaison homologue est donc nécessaire à l'intégration du gène de résistance. Aucune NTS n'est retrouvée dans ces constructions. Bien qu'ils s'intègrent à des régions différentes du génome, les deux plasmides testés donnent des résultats similaires. Pour le mutant de délétion de *ssnA*, nous obtenons peu ou pas de diminution du taux de transformation. Cependant, nous obtenons une augmentation plus marquée dans la souche surexprimant *SsnA*. Ces résultats sont similaires à un des tests effectués avec de l'ADNg. Ce phénotype, inattendu aux vues de nos tests biochimiques qui n'ont pas révélé d'interactions autres qu'avec la NTS, suggère que cette nucléase pourrait favoriser légèrement l'intégration d'ADN sans NTS par transformation. Il est possible que *SsnA*, lorsque surexprimée, interagisse avec d'autres effecteurs de la transformation ou de la recombinaison, ou directement avec l'ADN de façon non spécifique en conditions physiologiques. Comme ce phénotype semblait non relié aux NTS et peut-être indirect, nous n'avons pas exploré davantage cette avenue.

3.9.2 Transformation d'ADN avec NTS dans les zones de recombinaison

Pour faire suite à nos résultats sur l'activité enzymatique de *SsnA*, nous avons réalisé d'autres tests de transformation, mais cette fois avec des plasmides contenant un nombre et une disposition variée de NTS (**Figure 36**). Ces derniers s'intègrent à différents endroits du génome de l'hôte par double recombinaison homologue des régions contenant les NTS. Lorsque transformés dans nos souches de *Nm*, ces trois plasmides révèlent un important phénotype. En absence de *SsnA*, ces constructions sont intégrées beaucoup plus fréquemment au génome que dans la souche sauvage. À l'inverse, dans la souche complémentée qui surexprime *ssnA*, une diminution drastique du taux de transformation est observée. Il est à noter que ce phénotype est l'inverse de celui observé avec les constructions sans NTS (**Figure 35**), ce qui renforce ces différences. Le type de marqueur de sélection utilisé n'affecte pas ces résultats, puisque deux constructions similaires avec des gènes de résistance différents donnent exactement la même chose (données non montrées). Concrètement, ces résultats indiquent que *SsnA* réduit considérablement l'intégration d'ADN transformant contenant des NTS. *SsnA* semble donc jouer un rôle de médiateur de la transformation ou de la recombinaison. L'interaction physique entre la nucléase et ces éléments répétés, observée lors de nos tests biochimiques, semble également

confirmée *in vivo* chez le méningocoque, bien que le mécanisme sous-jacent soit encore obscur. Vraisemblablement, SsnA exerce son activité nucléase sur l'ADN transformant une fois présent dans le cytoplasme sous forme simple-brin, juste avant sa recombinaison au chromosome.

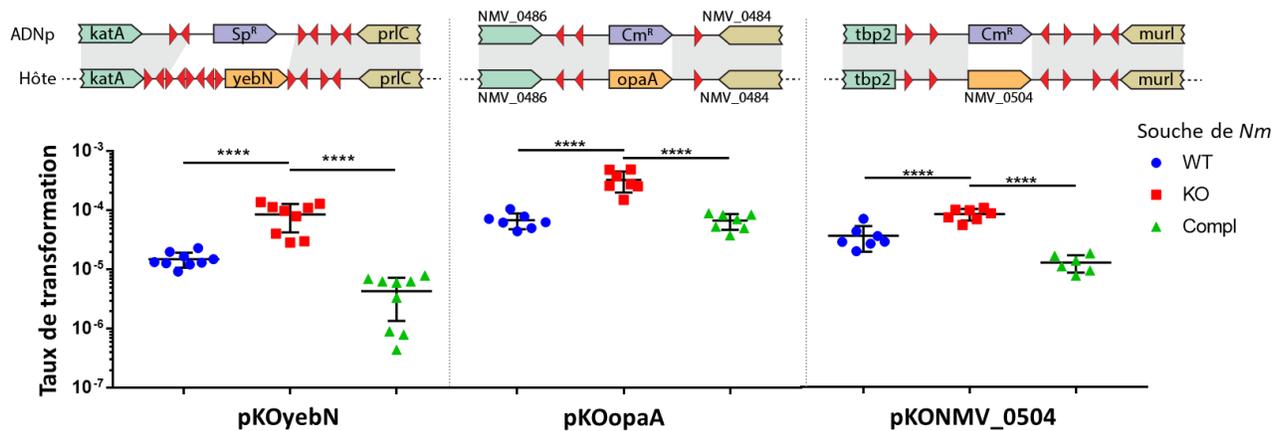


Figure 36 Transformation de *Nm* avec des plasmides contenant la NTS

Trois plasmides intégratifs ont été utilisés comme substrats pour des tests de transformation chez les mutants de *Nm*. Chacun contient un marqueur de sélection flanqué de séquences homologues au génome de l'hôte contenant plusieurs copies de la NTS. Ces régions homologues sont ombragées. Les moyennes et écarts-types d'au moins trois expériences indépendantes composées de plusieurs réactions distinctes de transformation sont affichés. Des ANOVA unilatérales avec comparaisons multiples de Tukey ont été utilisées comme tests statistiques. Ns non significatif, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Les précédents tests montrent une forte corrélation entre deux phénotypes opposés liant SsnA à la présence ou l'absence de NTS dans l'ADN transformant. Cependant, malgré la confirmation de ces phénotypes dans diverses régions du génome du méningocoque, ces derniers ne sont pas une preuve directe de l'interaction SsnA-NTS *in vivo*.

Il est à noter qu'un des plasmides avec NTS (pKOYebN) a été construit à la base pour transformer la souche *N. meningitidis* MC58 du sérogroupe B et non la souche 8013 du sérogroupe C, la première étant peu manipulée au laboratoire pour des raisons de biosécurité. Ce plasmide possède par conséquent des régions d'homologie différentes du génome de 8013. Afin de supprimer ce biais de séquence, nous avons généré une souche hybride de 8013 dans laquelle nous avons remplacé la région entière entre les gènes *katA* et *prlC* par celle de MC58. Nous avons également introduit une mutation ponctuelle dans le gène *yebN* situé au milieu de cette région riche en NTS, conférant une sensibilité au manganèse qui peut être exploitée comme marqueur sélection. Cette souche, nommée *Nm* 8013 2C4.3 yebOut_MC58, a été générée par une méthodologie de manipulation génétique développée par mes pairs et moi-même, publiée très récemment (Nyongesa *et al.*, 2022a). De cette souche mère, divers variants ont été générés

pour être utilisés dans des tests de transformation. Pour toutes ces différentes souches, les mutants d'expression de *ssnA* (KO et Compl) ont aussi été générés.

3.9.3 Transformations de plasmides avec des NTS mutées

Afin de prouver de façon plus directe que le phénotype observé précédemment est bien dû à la présence de NTS dans l'ADN transformant, nous avons procédé à la génération de deux plasmides presque identiques correspondant à la région *katA-prlC*. Dans un de ces plasmides, la première guanine de toutes les copies de la NTS a été substituée par une adénine. Cette substitution correspond à la mutation G26A sur la séquence RSE75, révélée dans nos tests biochimiques comme empêchant presque totalement le clivage par SsnA (voir **Figure 30**, Chapitre II). Le gène *yebN* sauvage de ces constructions permet la sélection des transformants sur milieu avec manganèse. En évitant l'insertion d'une cassette entière de résistance, un seul évènement de recombinaison sur toute la longueur de la séquence permet l'intégration de l'ADN transformant. Pour que les mutations introduites dans les NTS n'affectent pas le degré d'homologie entre l'ADN transformant et le génome receveur, nous avons généré des souches de *Nm* portant exactement les mêmes mutations, appelées *Nm* NTSmut(G26A). Chaque plasmide a été transformé dans les souches de *Nm* correspondantes (**Figure 37**). Lorsque les NTS sont intactes, le phénotype précédemment observé est reproduit, ce qui confirme l'effet inhibiteur de SsnA sur l'intégration d'ADN transformant contenant ces séquences répétées. Cependant, ce phénotype est perdu lorsque les NTS sont mutées. Une seule substitution dans chaque copie de la NTS suffit donc à prévenir l'action inhibitrice de SsnA. Ces tests confirment davantage l'interaction entre la nucléase et les NTS chez le méningocoque, d'où le nom « *Neisseria Transformation Sequences* » que nous avons donné à ces éléments répétés.

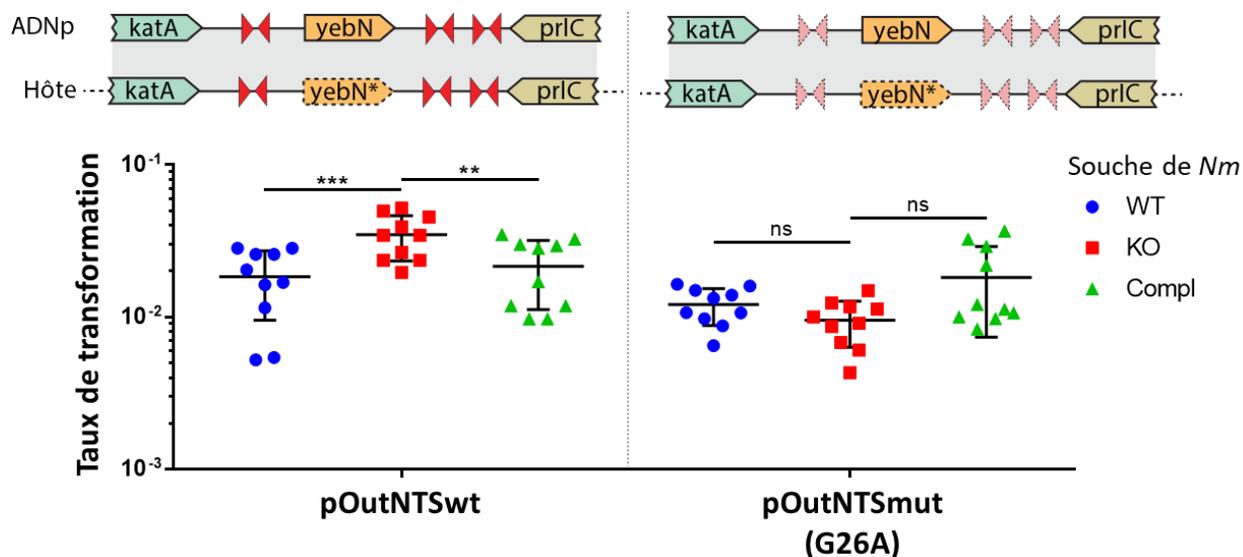


Figure 37 Transformation de plasmides avec des NTS intactes et mutées

Dans le panneau de gauche, le plasmide pOutNTSwt contenant des NTS intactes a été transformé dans *Nm* yebOut_MC58. À droite, le plasmide pOutNTSmut(G26A) a été transformé dans *Nm* NTSmut(G26A). Les régions homologues sont ombragées. Les moyennes et écarts-types d'au moins trois expériences indépendantes composées de plusieurs réactions distinctes de transformation sont affichés. Des ANOVA unilatérales avec comparaisons multiples de Tukey ont été utilisées comme tests statistiques. Ns non significatif, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

3.9.4 Autres tests de transformation

Intrigués par ces découvertes, nous avons ensuite tenté de modifier plusieurs paramètres dans nos tests de transformation dans le but de renforcer nos conclusions et de mieux comprendre comment SsnA influence l'intégration d'ADN transformant. En effet, à ce stade-ci, il n'est toujours pas clair si cette nucléase agit réellement dans la compétence naturelle des *Neisseria*, ou bien seulement dans la recombinaison, qui représente la dernière étape de la transformation mais est également impliquée dans des mécanismes de réparation d'ADN et de variabilité génétique. Nous avons donc généré une nouvelle série de plasmides intégratifs et de souches assez similaires aux précédents. Puisque notre intérêt se porte principalement sur les régions contenant les NTS et leur interaction avec SsnA, nous avons fait en sorte que l'homologie entre chaque plasmide et le génome receveur soit limitée à ces régions, dans l'espoir d'accentuer les phénotypes précédemment observés (**Figure 38**). Encore une fois, lorsque les NTS sont intactes, la présence de SsnA réduit le taux d'intégration de telles constructions génétiques. Un phénotype identique est obtenu avec de l'ADN génomique total, suggérant que la longueur de l'homologie à l'extérieur des régions riches en NTS n'a aucune influence sur l'action de SsnA (**Figure 38A**). Nous avons par la suite testé des plasmides possédant les mêmes caractéristiques, mais abritant une ou deux substitutions dans chaque NTS : G26A ou G26A+C49A. Selon nos tests biochimiques, la

première mutation empêche le clivage par SsnA mais pas sa liaison à la NTS, tandis que la seconde mutation, correspondant au dernier nucléotide de ces séquences répétées, empêche la liaison de l'enzyme. En combinant ces deux mutations, nous avons espoir d'inverser complètement le phénotype obtenu en présence de NTS, pour se rapprocher de celui obtenu avec les constructions sans NTS (voir **Figure 35**). Peu importe la mutation utilisée, la transformation de ces constructions donne les mêmes résultats, soit une perte d'effet lié à la présence ou l'absence de SsnA (**Figure 38B**). En effet, malgré de légères différences lorsque la mutation G26A est utilisée seule, la surexpression de la nucléase n'inverse par l'effet observé pour la souche KO. Ceci indique encore une fois que la longueur des séquences d'homologie en amont et en aval des NTS a peu d'impact sur l'influence de SsnA, alors que la modification des NTS abroge le phénotype observé.

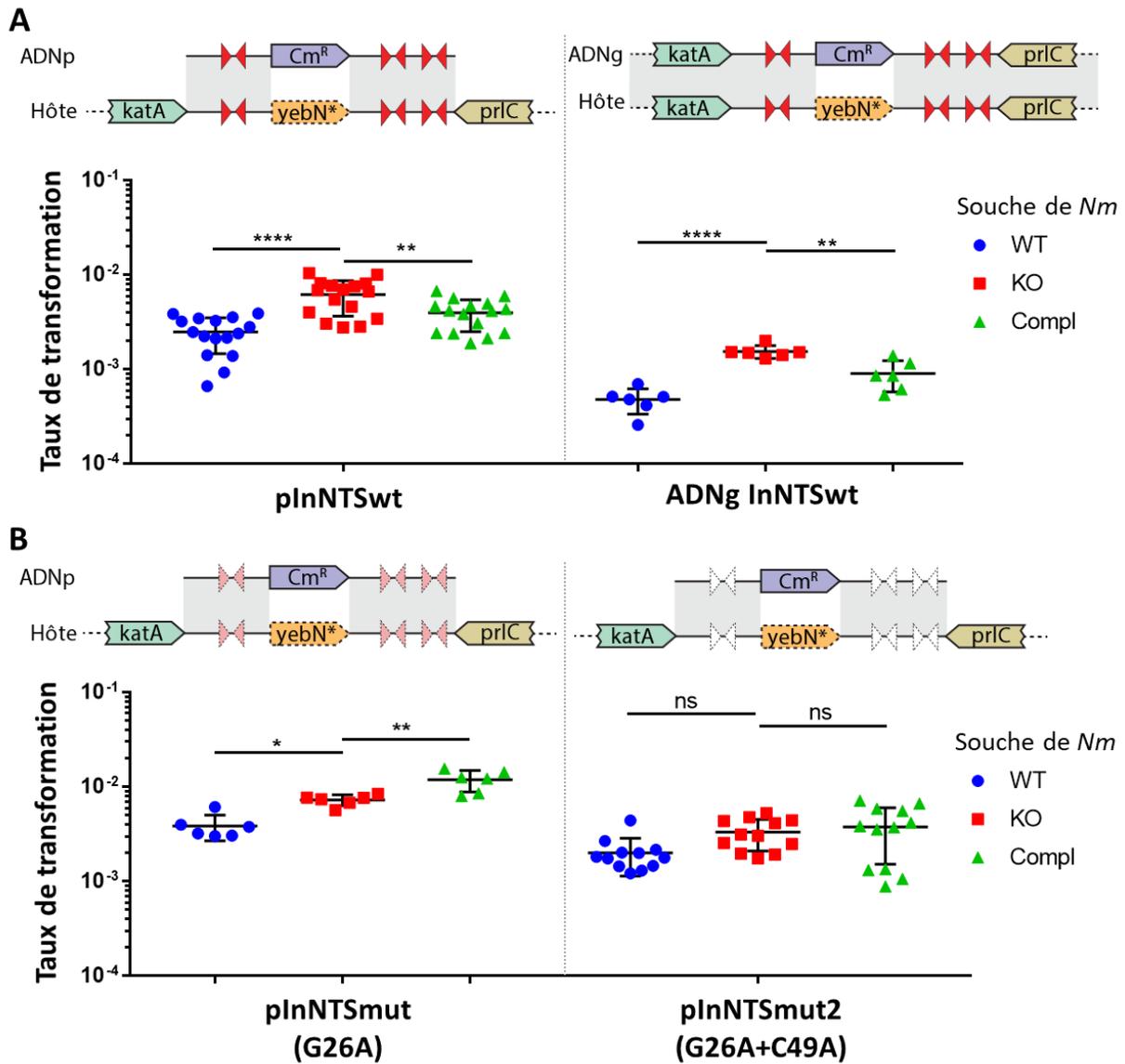


Figure 38 Transformations d'autres substrats avec NTS

A. Tests de transformations avec un plasmide (gauche) ou de l'ADN génomique total (droite) contenant un gène de résistance au chloramphénicol et des NTS intactes. Pour le plasmide, l'homologie permettant la double recombinaison est limitée aux régions intergéniques abritant les NTS. **B.** Tests de transformation avec des plasmides similaires, mais dont les NTS abritent toutes une substitution (gauche) ou deux (droite) réduisant le clivage et/ou la liaison de *SsnA in vitro*. Comme pour la figure précédente, les souches receveuses portent les mêmes mutations pour minimiser les biais d'homologie. Les régions homologues sont ombragées. Les moyennes et écarts-types d'au moins trois expériences indépendantes composées de plusieurs réactions distinctes de transformation sont affichés. Des ANOVA unilatérales avec comparaisons multiples de Tukey ont été utilisées comme tests statistiques. ns non significatif, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

3.9.5 Transformations par co-culture

Tous les précédents tests de transformation ont été réalisés avec de l'ADN génomique ou plasmidique pur comme substrat. Or, en conditions biologiques de colonisation ou d'infection par *N. meningitidis*, l'ADN transformant provient d'autres *Neisseria* qui larguent leur contenu cellulaire à leur mort ou qui sécrètent activement leur ADN. Afin de se rapprocher de ces conditions biologiques, mais aussi dans le but d'étudier le rôle de SsnA dans le don d'ADN, nous avons à nouveau généré une panoplie de souches dont la nucléase est intacte ou délétée. Les souches receveuses, nommées *Nm ΔyebN::CmSTOP Km*, possèdent à la place du gène *yebN* une cassette de résistance au chloramphénicol inactivé par une mutation ponctuelle générant un codon stop, en plus d'un gène de résistance à la kanamycine inséré ailleurs dans le génome. Les souches donneuses quant à elles, possèdent au même endroit une copie intacte du gène de résistance au Cm. À l'aide de cocultures de souches donneuses et receveuses, nous pouvions donc mesurer les transferts horizontaux de gènes (HGT) de façon unilatérale. En effet, il est attendu que la simple recombinaison homologe nécessaire à l'échange des allèles de résistance au Cm soit beaucoup plus fréquente que le transfert de la résistance à la Km de la souche receveuse à la souche donneuse, qui nécessite deux recombinaisons homologues simultanées puisqu'il s'agit d'une grande insertion.

Lorsque *N. meningitidis* agit comme donneur, on peut observer une forte augmentation des taux de transformation chez la souche receveuse n'exprimant pas SsnA, tout comme lors des tests précédents avec de l'ADN purifié (**Figure 39A**). Cet effet est indépendant de la présence de SsnA dans la souche donneuse, puisqu'aucune différence n'est observée lorsque celle-ci n'exprime pas la nucléase. Cela signifie que SsnA agit uniquement sur le traitement ou la recombinaison d'ADN transformant une fois dans la cellule receveuse, et non sur la sécrétion ou la modification de l'ADN transmis par la cellule donneuse. Afin d'élargir ces résultats à l'ensemble des HGT ayant lieu entre les *Neisseria*, nous avons réalisé une seconde série d'expériences quasi identiques, mais cette fois en utilisant l'espèce commensale *N. muscili* comme donneuse d'ADN, dans laquelle nous avons introduit les mêmes modifications génétiques (**Figure 39B**). Encore une fois, l'absence de SsnA dans la souche receveuse de *Nm* se traduit par une augmentation significative de l'intégration chromosomique du marqueur dans la région ciblée.

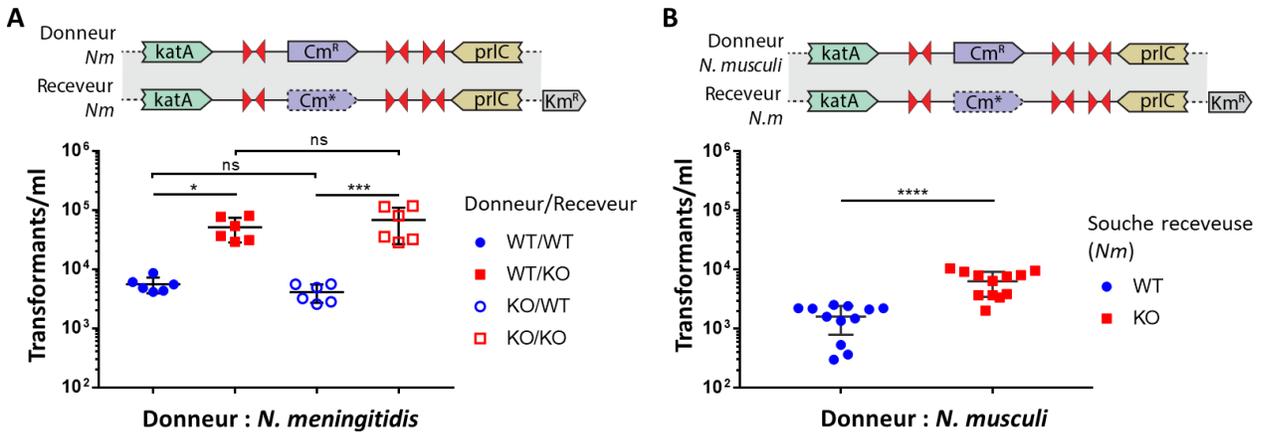


Figure 39 Transformation de *Nm* par coculture avec une souche donneuse

A. Tests de transformation avec *Nm* $\Delta yebN::Cm$ comme souche donneuse et *Nm* $\Delta yebN::CmSTOP$ Km comme souche receveuse. Les transformants ont été sélectionnés sur milieu avec Km+Cm, quantifiés par rapport au volume de réaction, et comparés à l'aide d'une ANOVA unilatérale avec comparaison multiple de Tukey. **B.** Tests de transformation similaires en A, mais utilisant *N. muscili* $\Delta yebN::Cm$ comme souche donneuse. Puisque la présence ou l'absence de SsnA dans la souche donneuse n'a pas d'influence, seule la souche SsnA WT a été utilisée comme donneuse. Différences comparées à l'aide d'un test de Student non-pairé. Pour A et B, les régions homologues sont ombragées. Les moyennes et écarts-types d'au moins trois expériences indépendantes composées de plusieurs réactions distinctes de transformations sont affichés. ns non significatif, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Ensemble, ces expériences de HGT par cocultures supportent davantage le rôle important que joue SsnA dans l'intégration d'ADN contenant des NTS au génome, et suggèrent que ce mécanisme est partagé entre les *Neisseria* encodant ce gène. Ces fonctions biologiques soulèvent l'importance de cette petite nucléase dans l'évolution et l'adaptation des *Neisseria*, tout particulièrement de *N. meningitidis* qui possède des centaines de copies de la NTS. Bien que le mécanisme exact médié par SsnA de même que sa place dans la machinerie complexe de compétence naturelle et de recombinaison soient encore méconnus, les informations révélées par nos tests biochimiques et phénotypiques servent déjà de bases solides qui permettent d'orienter plusieurs projets de recherche sous-jacents et de soulever plusieurs hypothèses. L'hypothèse la plus concordante avec toutes ces données est que SsnA exerce un mécanisme de restriction d'ADN simple-brin qui clive à proximité des NTS quand elles sont présentes dans l'ADN transformant. Cette restriction génère des fragments dont l'homologie écourtée avec le chromosome de l'hôte réduit leur taux d'intégration par recombinaison homologue. Ce modèle d'action proposé ainsi que les conséquences d'un tel mécanisme sont davantage détaillés dans la section discussion de cette thèse.

3.10 Impact sur la variation antigénique de PilE

Le rôle de SsnA dans l'intégration d'ADN transformant via la compétence naturelle est important, mais nous soupçonnons que ce n'est pas la seule fonction biologique de cette enzyme. En effet, d'autres processus cellulaires chez les *Neisseria* font intervenir de l'ADN simple-brin par recombinaison homologue, dont la variation antigénique des pili. Ce mécanisme permet la diversification génétique d'un gène précis, celui de la piline majeure *pilE*, par recombinaison aléatoire de pseudogènes appelés *pilS*, présents en plusieurs copies uniques possédant chacune diverses régions d'homologie avec *pilE*. Or, *pilE* et chacune des copies de *pilS* sont entourées de nombreuses copies de la NTS, cible de l'activité catalytique de SsnA. Cette coïncidence frappante soulève la possibilité que cette nucléase module le taux de variation antigénique de *pilE* par un mécanisme similaire à celui observé lors de la transformation d'ADN exogène. Nous avons donc voulu tester quantitativement ce phénomène dans des mutants d'expression de SsnA.

Pour étudier la variation antigénique induite par SsnA, nous avons dû générer des souches dans lesquelles le gène *recA* est inductible. En effet, la recombinase RecA encodée par ce gène est cruciale à la recombinaison homologue, mais son activité constitutive pose problème quand on veut mesurer ce mécanisme à un moment précis ou dans des conditions spécifiques. Nous avons donc généré la souche *Nm* 8013 2C4.3 *recA6::Tet* dans laquelle le *recA* natif a été remplacé par un allèle sous le contrôle du promoteur Lac et couplée à un gène de résistance à la tétracycline (Tet). En présence d'IPTG induisant l'expression de *recA*, nous avons ensuite procédé à la délétion et à la complémentation de SsnA dans cette nouvelle souche. Pour quantifier et identifier précisément les événements de recombinaison ayant lieu lors de la variation antigénique, nous avons amélioré les méthodologies existantes en introduisant le séquençage par Nanopore MinION et un pipeline d'analyse unique (**Figure 40**).

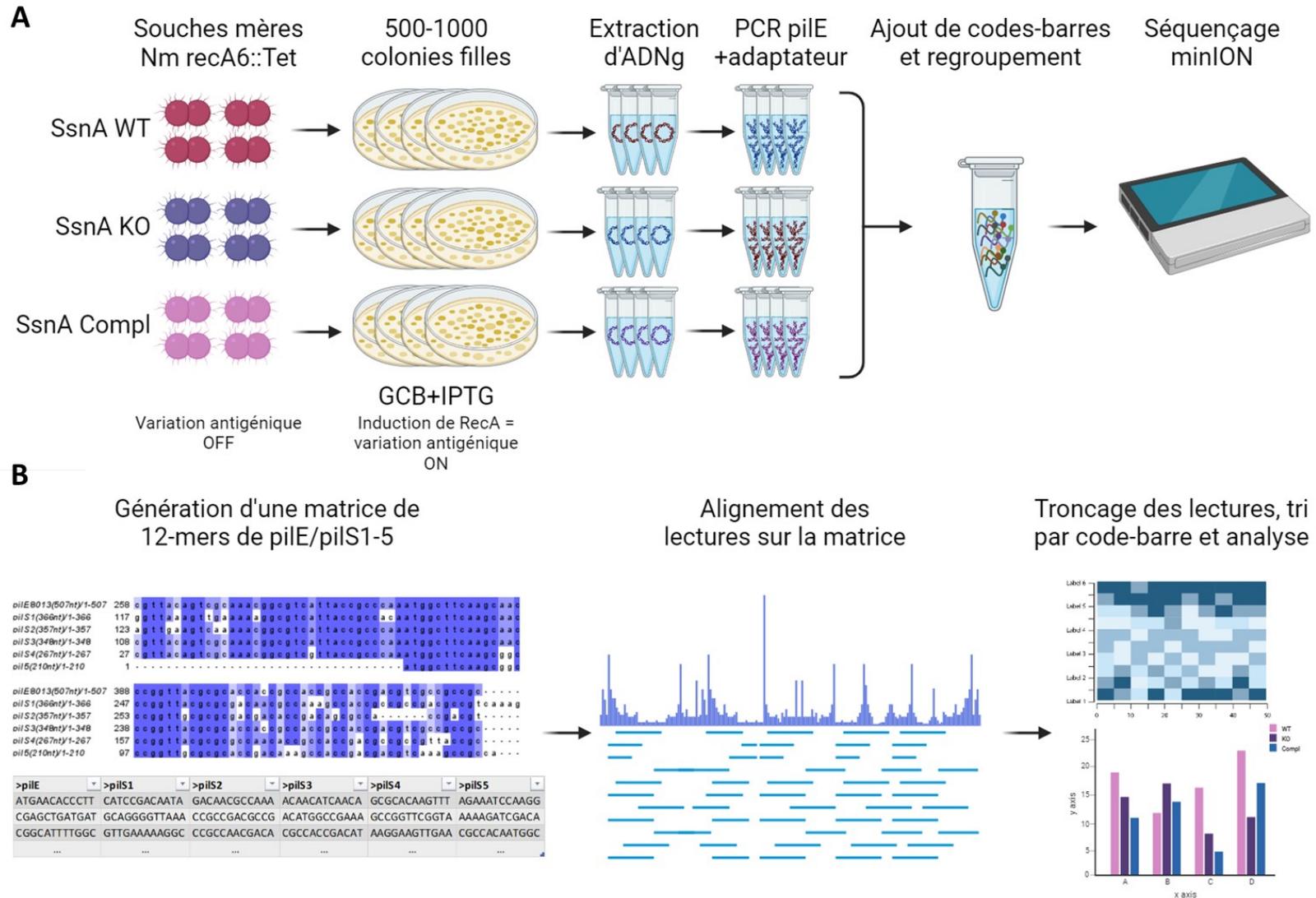


Figure 40 Méthodologie des tests de variation antigénique sans ADN transformant

A. Schéma expérimental. La souche *Nm recA6::Tet*, capable de recombinaison uniquement en présence d'IPTG, a été utilisée comme souche parentale pour chaque mutant de SsnA. **B.** Pipeline d'analyse bio-informatique. Les détails expérimentaux et d'analyse sont présentés dans le texte.

3.10.1 Variation antigénique en absence d'ADN transformant

Dans un premier temps, nous avons isolé quatre colonies de chaque souche en absence d'IPTG (WT, KO, Compl) (**Figure 40A**). Chacune de ces colonies parentales a ensuite été diluée puis étalée sur milieu avec IPTG de façon à obtenir plusieurs centaines de colonies filles capables de variation antigénique. Chaque série de colonies filles a ensuite été groupée, puis leur gène *pilE* a été amplifié par PCR et séquencé par MinION. Afin d'amplifier l'ensemble des variants obtenus à partir de chaque colonie parentale, les amorces utilisées se fixent dans les régions invariables de ce gène entourant les zones d'homologie et de recombinaison avec les copies *pilS*. Afin de permettre une analyse à haut débit et à faible coût en une seule réaction de séquençage, chaque échantillon correspondant aux différentes souches parentales a été marqué par une séquence unique agissant comme étiquette pour le séquençage. Pour l'analyse des variants, les séquences obtenues ont été séparées selon ces étiquettes, puis alignées contre chacune des copies de *pilS* et de *pilE* afin d'identifier l'origine et la nature de chaque événement de recombinaison. Pour détecter les microrecombinaisons et pas seulement les larges réarrangements, nous avons généré une matrice de segments de 12 nucléotides (12-mers) qui a servi aux alignements des lectures de séquençage (**Figure 40B**).

En conditions normales de culture, la fréquence d'un variant donné de *pilE* tourne entre 0.02% et 2.5% (**Figure 41**). Comme contrôle, nous avons également aligné les lectures de séquençage à des régions invariables de *pilE*, ce qui a confirmé que la grande majorité des fragments analysés appartiennent bien à cette région génétique. Ces chiffres sont comparables à ceux répertoriés dans la littérature, validant davantage notre méthodologie et notre analyse unique (Davies et al., 2014; Helm & Seifert, 2010; Rotman et al., 2016). Malheureusement, le taux d'erreur de séquençage généré par MinION ne permet pas d'inférer la fréquence totale de variation de *pilE*, puisque les événements de recombinaison de cette région peuvent échanger aussi peu qu'un seul nucléotide (Criss et al., 2005). Aucune différence significative n'a été détectée entre les souches WT, KO et Compl, suggérant que la nucléase SsnA n'est pas impliquée dans ce processus ou que notre méthode n'est pas assez résolutive.

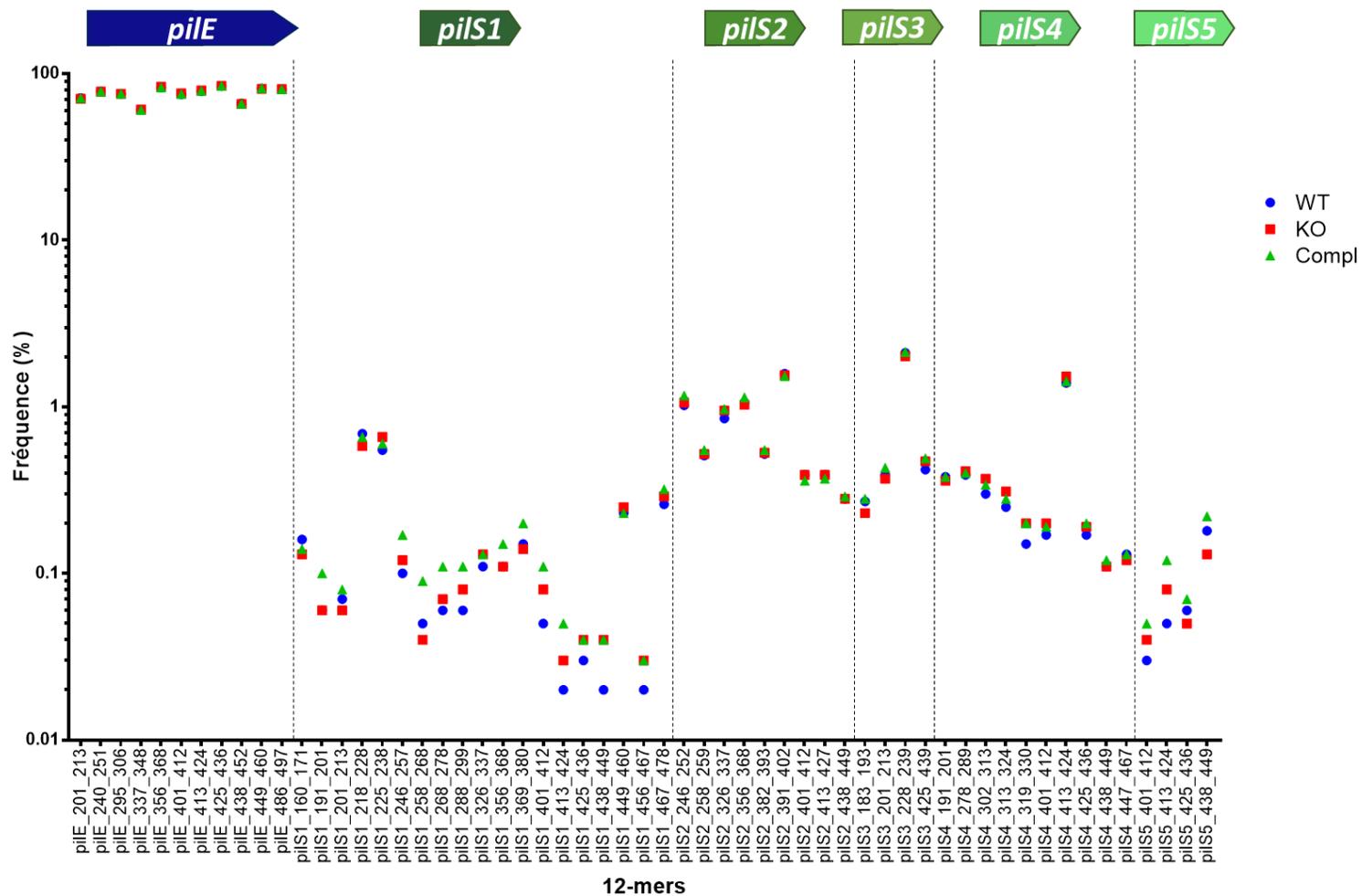


Figure 41 Variation antigénique de *Nm* en absence d'ADN transformant

Graphique illustrant la fréquence et l'identité des recombinaisons entre le gène *pilE* et les pseudogènes *pilS1-5*. Plusieurs centaines de colonies provenant de quatre clones parentaux de chaque souche de *Nm* 8013 *recA6::Tet* sont regroupées dans cette analyse de variants par séquençage du gène *pilE*. Toutes les lectures séquencées ont été alignées sur une matrice représentant des segments de 12 nt (12-mers) couvrant diverses régions de *pilE* et des *pilS*, représentés en abscisse. À des fins de clarté, seule la moyenne des quatre réplicats est affichée. Aucune différence statistiquement significative n'a été trouvée entre les souches testées par ANOVA unilatérale sur mesures répétées avec comparaison multiple de Tukey.

3.10.2 Variation antigénique en présence d'ADN transformant

L'origine des copies de *pilS1* utilisées comme substrats de recombinaison pour moduler la séquence de *pilE* est sujette à débat. En effet, des études suggèrent que ce mécanisme est principalement intracellulaire grâce à la nature polyploïde des *Neisseria* pathogènes, tandis que d'autres suggèrent que la transformation d'ADN d'autres cellules par compétence naturelle est la principale source d'allèles *pilS1* (Vink *et al.*, 2012). Pour vérifier l'influence de la transformation dans la variation antigénique en relation avec la présence de SsnA, nous avons réalisé des tests très similaires aux précédents. Cette fois-ci, chaque souche parentale a été incubée avec de l'ADN génomique total de *N. meningitidis* 8013 en présence d'IPTG avant d'étaler ces souches sur milieu sans IPTG. Cette méthodologie vise à exprimer RecA, essentielle à la variation antigénique, uniquement lorsque de l'ADN exogène est activement importé par compétence naturelle. Puisque les résultats de cette analyse sont redondants avec la précédente, seul le variant le plus fréquent provenant de chaque *pilS* est illustré en **Figure 42**. Malgré la présence d'ADN transformant, aucune différence n'est observée une nouvelle fois entre les souches testées. De plus, les taux de variants spécifiques sont identiques à ceux obtenus en absence d'ADN transformant. Concrètement, ces résultats indiquent que la variation antigénique est bien un processus intracellulaire indépendant de la transformation par compétence naturelle, en plus de confirmer que SsnA ne joue aucun rôle dans ce mécanisme spécifique aux espèces pathogènes de *Neisseria*. Néanmoins, ces résultats négatifs donnent une information cruciale sur la relation entre SsnA et le mécanisme de recombinaison homologue, dernière étape de la transformation nécessaire à l'intégration de l'ADN transformant au chromosome de l'hôte. En effet, l'absence de modulation de la variation antigénique en présence et en absence de SsnA suggère que cette nucléase n'est pas impliquée dans la recombinaison homologue, et donc qu'elle agit bien dans le processus de transformation, avant l'interaction de l'ADN simple-brin avec RecA.

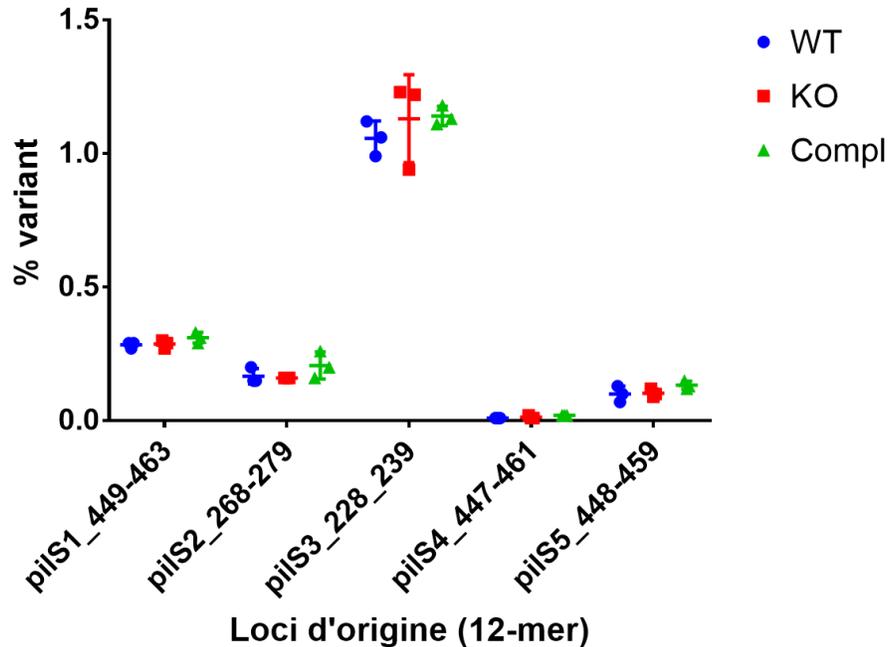


Figure 42 Variation antigénique de *Nm* en présence d'ADN transformant

Graphique illustrant la fréquence et la nature des recombinaisons entre le gène *pilE* et les pseudogènes *pilS1-5* en conditions favorisant l'import d'ADN exogène par transformation. Analyse identique à la précédente, mais dont seul le variant le plus fréquent provenant de chacune des copies de *pilS* est illustré. Les moyennes et écarts-types provenant de trois clones parentaux indépendants de chaque souche sont présentés. Aucune différence significative n'a été détectée par ANOVA unilatérale sur mesures répétées avec comparaison multiple de Tukey.

3.11 Implications dans la virulence de *N. meningitidis*

Pour rappel, les NTS sont beaucoup plus nombreuses dans le génome des *Neisseria* pathogènes que chez les espèces commensales, et ce tout particulièrement pour *N. meningitidis* qui en possède plusieurs centaines. De plus, ces éléments répétés sont distribués de façon non aléatoire dans le génome, formant des clusters importants autour de plusieurs gènes codant pour des protéines de surface impliquées dans la virulence et l'évasion immunitaire. Cette forte corrélation entre la virulence et la présence de NTS suggère que ces dernières sont impliquées dans la pathogénicité des *Neisseria*. Nous avons démontré lors de tests biochimiques que la nucléase SsnA se fixe spécifiquement à ces NTS et clive l'ADN simple-brin à proximité. Cette activité enzymatique se traduit chez la bactérie par une modulation importante de l'intégration d'ADN transformant contenant ces séquences répétées. La compétence naturelle de ces espèces est considérée comme un atout considérable qui permet la génération de diversité génétique, l'acquisition de gènes ou encore la réparation d'ADN. Puisque SsnA est impliquée dans la

transformation de *N. meningitidis* via son interaction avec les NTS, nous avons voulu explorer son ou ses potentiels rôles dans la virulence du méningocoque.

3.11.1 Infections *in vitro*

Un facteur de virulence peut être impliqué à plusieurs étapes de la pathogénicité comme l'adhésion aux cellules hôtes, leur invasion, la multiplication intracellulaire ou encore l'évasion immunitaire. Nous avons exploré plusieurs de ces avenues, en utilisant d'abord des modèles d'infection *in vitro*. Deux lignées cellulaires ont été choisies : la lignée de monocytes humains THP-1, des cellules phagocytaires pouvant être différenciées en macrophages, ainsi que la lignée de cellules épithéliales pulmonaires Calu-3. Ces dernières, représentant différents types cellulaires auxquels les *Neisseria* pathogènes sont confrontés lors d'une infection, sont fréquemment utilisées dans l'étude de leur potentiel d'adhésion et d'invasion (Colicchio *et al.*, 2015; Sutherland *et al.*, 2010). Dans le but de progresser par la suite sur des modèles d'infection de souris, nous avons dû changer de souches de référence afin d'augmenter leur potentiel pathogène. En effet, le méningocoque étant un pathogène humain obligatoire qui interagit spécifiquement avec des récepteurs retrouvés uniquement sur nos cellules, il est préférable d'utiliser des souches qualifiées d'hyperinvasives dans des modèles murins d'infection. Nous avons donc de nouveau généré des mutants de SsnA KO et Compl dans la souche hyperinvasive *Nm* LNP20553, dans laquelle nous avons préalablement introduit l'opéron Lux sous un promoteur fort. Cette modification confère à ces souches une luminescence constitutive qui permet de réaliser des mesures d'intensité lumineuse à l'aide d'un lecteur de plaque, représentant un avantage de temps énorme par rapport à la culture et au décompte de colonies traditionnellement utilisé (CFUs).

Dans un premier temps, nous avons testé le potentiel d'adhésion et d'invasion des souches de *Nm* LNP20553 lux dans la lignée de monocytes THP-1 (**Figure 43**). Brièvement, ces cellules ont été différenciées en macrophages avant l'infection via l'ajout de PMA (phorbol myristate acetate) de façon à obtenir une monocouche adhérente dans chaque puits de plaques de cultures. Pour les tests d'adhésion, les monocouches ont été incubées en présence d'une quantité définie de bactéries luminescentes avant de débiter les acquisitions ponctuelles du signal lumineux, systématiquement précédées de lavages pour enlever les bactéries non adhérentes (**Figure 43A**). Après les 8.5 premières heures, l'expérience a été prolongée jusqu'à 24 h post-infection (PI), en lavant les puits uniquement avant l'acquisition finale. On peut observer que la quantité de bactéries adhérentes augmente lentement tout au long de cette expérience et de façon plus

prononcée en absence de lavages entre 8.5 et 24 h post-infection. Cependant, aucune différence significative n'est observée entre les souches utilisées suggérant que SsnA n'influence pas l'adhérence cellulaire dans ces conditions. Une expérience complémentaire indique que l'augmentation après 8.5 h est probablement due à la croissance de *Nm* dans le milieu de culture cellulaire utilisé, en absence de lavages fréquents (données non montrées).

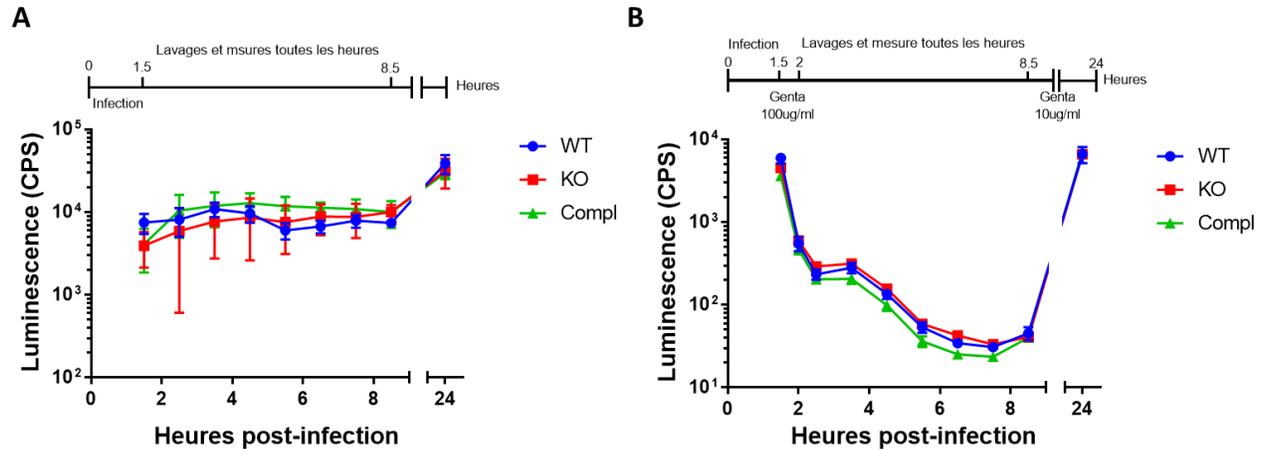


Figure 43 Adhésion et invasion de macrophages THP-1

Des monocouches de monocytes THP-1 différenciés en macrophages ont été incubées en présence de souches hyperinvasives luminescentes de *Nm* 20553 exprimant ou pas SsnA. **A.** Test d'adhésion mesurant la quantité totale de bactéries vivantes associées aux cellules par mesure de luminescence. Avant chaque acquisition, les monocouches ont été lavées et leur milieu changé pour éliminer les bactéries non adhérentes. Aucun lavage ni mesure n'a pas été effectué entre 8.5 h et 24 h post-infection (PI). **B.** Test d'invasion mesurant la luminescence associée aux bactéries intracellulaires après un traitement à la gentamicine (genta). Acquisitions réalisées immédiatement avant et après le traitement antibiotique ainsi qu'à toutes les heures post-infection après lavages et changement de milieu. Pour inhiber la croissance extracellulaire en absence de lavages entre 8.5 et 24 h PI, une concentration plus faible de gentamicine a été utilisée. En A comme en B, les moyennes et écarts-types provenant de triplicatas représentatifs de deux expériences indépendantes sont présentés. Aucune différence statistique n'est détectée entre les souches par test de Friedman (ANOVA non-paramétrique).

Pour les tests d'invasion, un traitement à la gentamicine a été utilisé immédiatement après l'infection pour éliminer les bactéries adhérentes et ne conserver que celles intracellulaires (**Figure 43B**). Pour limiter la croissance extracellulaire entre 8.5 h et 24 h PI en absence de lavages fréquents, une concentration bactériostatique de gentamicine a été utilisée, bien plus faible que le premier traitement bactéricide. Des mesures de luminescence avant et après le premier traitement antibiotique montrent que ce dernier est efficace, puisqu'il réduit la charge bactérienne d'un logarithme. Entre 2.5 h et 7.5 h PI, la charge bactérienne intracellulaire continue de diminuer, suggérant que *Nm* ne parvient pas à se multiplier dans les macrophages et est progressivement éliminée. Cependant, entre 7.5 h et 8.5 h PI, on constate un changement de

tendance vers la hausse, qui semble se prolonger jusqu'à 24 h PI pour atteindre des niveaux de luminescence comparables à ceux immédiatement après l'infection. Ces résultats suggèrent que *Nm* peut se multiplier dans les macrophages et se propager après une période d'adaptation de quelques heures, ce qui concorde avec d'autres études similaires (Spinosa *et al.*, 2007). Encore une fois, aucune différence significative n'est constatée entre les mutants de SsnA, indiquant que cette nucléase n'agit pas non plus sur l'invasion, la survie et la multiplication intracellulaire dans la lignée et les conditions testées.

Malgré ces premiers résultats négatifs, nous avons reproduit ces expériences avec la lignée de cellules épithéliales pulmonaires Calu-3, avec quelques modifications au protocole (**Figure 44**). Plutôt que d'utiliser la luminescence, nous avons procédé aux quantifications de la charge bactérienne par décompte de colonies (CFUs) à des temps précis puisque nous voulions voir des différences entre les souches sans nécessairement étudier la cinétique d'infection générale de *Nm*. Seules les souches SsnA WT et KO ont été utilisées cette fois. Pour les tests d'adhésion, un traitement à la saponine a été ajouté après l'infection pour éliminer les cellules eucaryotes et détacher les bactéries adhérentes (**Figure 44A**). Les tests d'invasions quant à eux, ont été réalisés pendant 6 h et 24 h (**Figure 44B**). Dans les deux cas, aucune différence significative n'est détectée entre les souches testées, indiquant encore une fois que SsnA ne joue aucun rôle dans l'adhésion et l'invasion dans ces conditions.

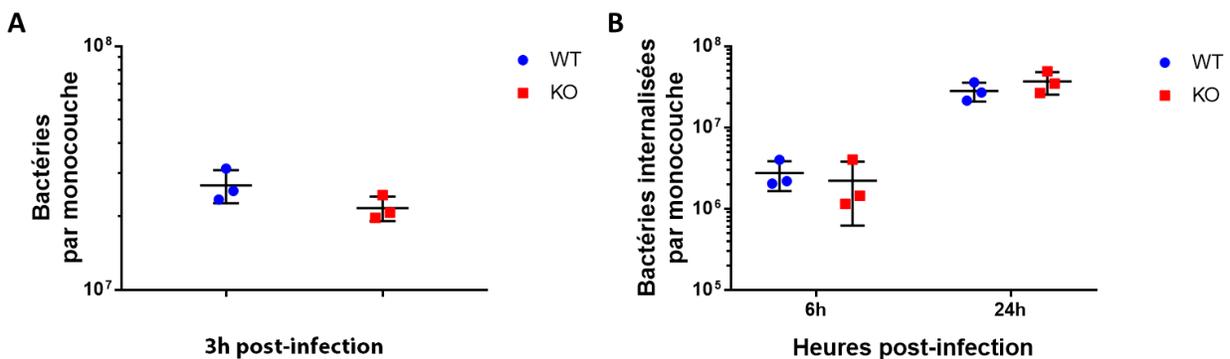


Figure 44 Adhésion et invasion de cellules épithéliales Calu-3

Les souches SsnA WT et KO de *Nm* 20553 lux ont été incubées sur des monocouches de Calu-3. **A.** Test d'adhésion mesurant le nombre de bactéries (CFUs) associées aux cellules 3 h post-infection (PI), à la suite d'un traitement à la saponine. **B.** Test d'invasion mesurant le nombre de bactéries intracellulaires 6 h et 24 h PI à la suite d'un traitement à la gentamicine. En A comme en B, les moyennes et écarts-types provenant de triplicatas représentatifs de deux expériences indépendantes sont présentés. Aucune différence statistique n'est détectée entre les souches par tests de Student non-pairés.

Ces précédents résultats ne sont pas très surprenants, mais étaient toutefois nécessaires pour éliminer de potentielles fonctions d'adhésine ou d'invasine de SsnA indépendamment du système immunitaire. En rétrospective, d'autres lignées cellulaires auraient peut-être été plus pertinentes pour étudier nos souches de *N. meningitidis*, comme la lignée de cellules épithéliales du pharynx Detroit 562 ou encore des lignées endothéliales du cerveau mimiquant la barrière hématoencéphalique. Ces dernières sont en effet plus représentatives des interactions hôte-pathogène du méningocoque (Seib & Peak, 2019).

3.11.2 Infection systémique de souris

Notre laboratoire a développé ou amélioré dans les dernières années plusieurs modèles d'infection de souris par différentes *Neisseria*, dont *N. meningitidis*. Afin d'obtenir une infection systémique similaire à la méningococcie chez l'humain sans utiliser de souris humanisées ou transgéniques, nous infectons des souris normales avec une dose sous-létale de virus influenza une semaine avant de procéder à la surinfection avec *N. meningitidis*. Cette préinfection à la grippe rend les souris susceptibles à l'invasion par *N. meningitidis* via l'expression d'une cascade de cytokines pro et anti-inflammatoires (Alonso *et al.*, 2003). De plus, afin de fournir la source de fer spécifiquement retrouvée chez l'humain et absente des souris, nous supplémentons l'inoculum bactérien de transferrine humaine chargée en fer au moment de l'infection. Enfin, nous utilisons avec ces modèles animaux des souches hyperinvasives luminescentes de *N. meningitidis*, permettant le suivi en temps réel de l'infection de façon très précise, tout en réduisant le nombre de souris utilisées puisque des mesures répétées peuvent être réalisées dans les mêmes sujets sans les sacrifier. Afin d'étudier le potentiel rôle de SsnA dans la virulence du méningocoque dans des conditions biologiques plus représentatives que par tests *in vitro*, faisant intervenir toute la complexité du système immunitaire, nous avons utilisé ces modèles murins.

Dans un premier temps, le potentiel pathogène des souches *Nm* LNP20553 lux SsnA WT, KO et Compl, utilisées auparavant dans les tests d'adhésion et d'invasion *in vitro*, a été évalué (**Figure 45**). Malgré un début d'infection similaire dans toutes les souris, indiquant une bonne normalisation des inocula de départ, les trois souches utilisées affichent de grosses différences de signal lumineux, représentatif de la charge bactérienne, à peine 24 h après l'infection (**Figure 45A**). En effet, avant même de procéder aux quantifications, on constate de façon très évidente une diminution drastique de la luminescence dans les souris infectées avec la souche SsnA KO par rapport à la souche sauvage. Ce phénotype est bien complétement par la réintroduction du

gène *ssnA* sous un promoteur fort (souche Compl), confirmant que ces effets sont bien dus à la présence ou l'absence de SsnA chez *N. meningitidis*. En s'attardant aux quantifications du signal lumineux de chaque souris à chaque temps d'acquisition, on constate que ces différences sont significatives dès la douzième heure après l'infection (**Figure 45B**). Indépendamment de la souche utilisée, on observe une augmentation rapide de la charge bactérienne et une dissémination du péritoine vers toute la circulation sanguine des souris aux premières heures de l'infection, indicateurs d'infection systémique réussie accompagnée d'une multiplication bactérienne active dans le sang et les tissus. La charge bactérienne semble ensuite diminuer dans le temps, probablement à la suite de la réaction immunitaire de l'hôte. Au-delà de 24 h PI, l'infection avec la souche KO est presque complètement éliminée, tandis que celles avec les souches WT et Compl font encore rage et s'accompagnent de symptômes cliniques importants atteignant des points critiques (données non montrées). Des courbes de survie viennent confirmer ces observations, même si l'expérience a dû être abrégée après 32 h pour des raisons éthiques (**Figure 45C**).

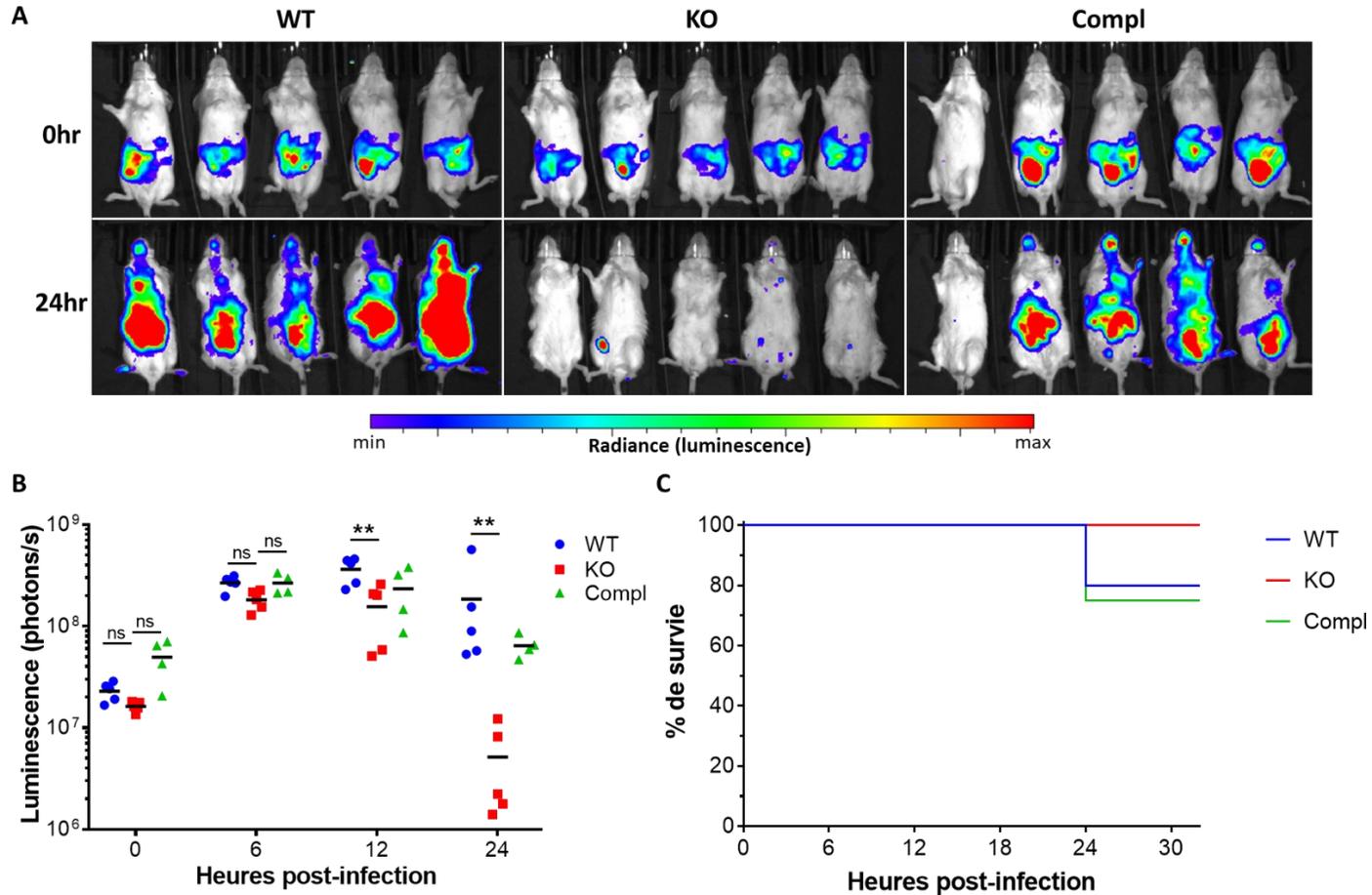


Figure 45 Infections systémiques de souris avec *Nm* LNP20553 lux

Après une préinfection à l'influenza, des souris BALB/c ont été infectées par injection intrapéritonéale de *Nm* LNP20553 lux SsnA WT, KO et Compl. Immédiatement après l'infection, puis 6, 12, 24 et 32 h post-infection (PI), les souris ont été imagées et le signal luminescent représentant la charge bactérienne active a été quantifié. Cinq souris par souche ont été infectées. **A.** Superposition des photographies et des acquisitions de luminescence de toutes les souris infectées aux temps 0 et 24 h PI. L'échelle de signal lumineux est identique dans toutes les images, permettant de les comparer visuellement en plus de procéder aux quantifications. La souris infectée avec la souche Compl la plus à gauche n'a pas été correctement injectée, et a donc été retirée des analyses quantitatives puisque le signal luminescent y était nul même à T0. **B.** Quantifications normalisées de la luminescence totale associée à chaque souris à chaque temps d'acquisition. Les moyennes sont présentées, et les trois souches ont été comparées statistiquement à chaque temps à l'aide d'ANOVA unilatérale avec comparaison multiple de Tukey. ns non significatif, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. **C.** Courbes de survie des souris infectées. Pour des raisons éthiques, l'expérience a été interrompue à 32 h PI car les souris infectées avec les souches WT et Compl ont atteint un point critique, alors que celles infectées avec la souche KO se portaient bien.

Ces résultats, très contrastés par rapport à l'absence de phénotype d'adhésion et d'invasion, suggèrent que SsnA est un effecteur important de la pathogénicité de *N. meningitidis*. En effet, l'absence de cette nucléase est associée à une élimination presque complète de l'infection en moins de 32 h, alors que sa présence ou sa surexpression mène à des infections létales. Le fait qu'aucune différence n'est observée en présence ou absence de SsnA au pic de l'infection, soit après 6 h, pourrait suggérer un rôle dans l'évasion immunitaire et le maintien de l'infection plutôt qu'un rôle dans son établissement et la multiplication bactérienne. Ces fonctions biologiques d'adaptation sont fort probablement reliées à son interaction avec les NTS, qui flanquent de nombreux gènes de virulence.

Étant donné l'importance de ces découvertes et l'impact qu'elles représentent dans les études sur *Neisseria meningitidis*, un pathogène redoutable avec un lourd bilan en santé publique, nous avons voulu valider le rôle de SsnA dans la virulence du méningocoque à l'aide d'une deuxième souche hyperinvasive. Cette validation est d'autant plus importante puisque nous n'avons pas été en mesure d'identifier le mécanisme sous-jacent à ces effets. Nous avons donc généré un mutant SsnA KO de la souche *Nm* LNP24198 lux_Km. Cette fois-ci, la souche complémentée n'a pas été générée en raison d'un problème de redondance des marqueurs de sélection utilisés. Encore une fois, on observe une évolution drastiquement différente de l'infection entre la souche WT et KO (**Figure 46A**). Tout comme pour les précédents tests, la charge bactérienne entre les deux souches est identique dans les premières heures de l'infection, avant de chuter pour la souche KO à partir de 12 h PI (**Figure 46B**). Cette fois-ci, les quantifications sont montrées jusqu'à 32 h PI, mais il est important de noter que plusieurs souris infectées avec la souche WT ont succombé à l'infection avant cette acquisition et n'ont donc pas pu être imagées et quantifiées. Les courbes de survie témoignent encore des différences cliniques qu'occasionnent ces souches (**Figure 46C**). L'effet important de la présence ou l'absence de SsnA dans la pathogénicité de *N. meningitidis* est donc parfaitement reproduit et confirmé.

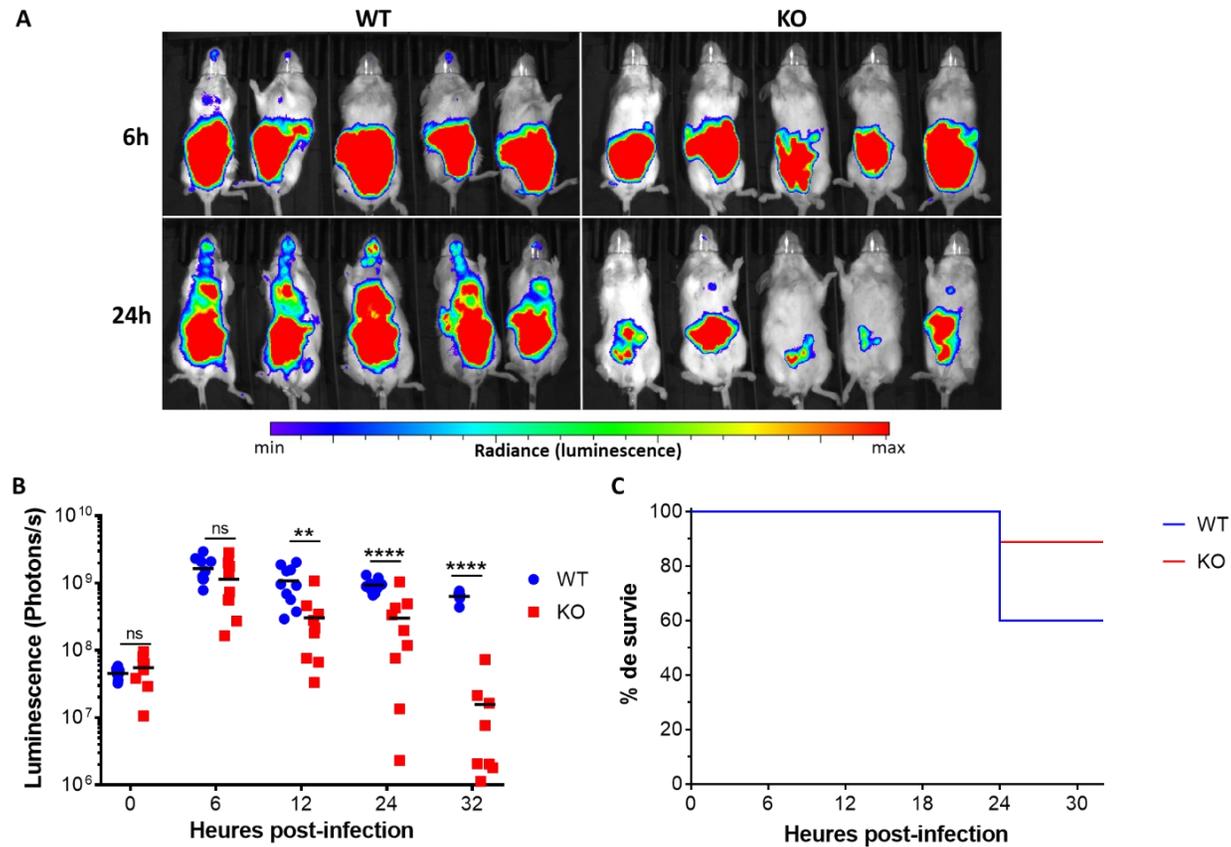


Figure 46 Infections systémiques de souris avec *Nm* LNP24198 lux_Km

Sept jours après une préinfection à l'influenza, des souris BALB/c ont été infectées par injection intrapéritonéale de *Nm* LNP24198 lux_Km SsnA WT et KO. Immédiatement après l'infection, puis 6, 12, 24 et 32 h post-infection (PI), les souris ont été imagées et le signal luminescent représentant la charge bactérienne active a été quantifié. Dix souris par souche ont été infectées lors de deux expériences indépendantes. **A**. Superposition des photographies et des acquisitions de luminescence de cinq souris représentatives, infectées aux temps 6 h et 24 h PI. L'échelle de signal lumineux est identique dans toutes les images, permettant de les comparer visuellement en plus de procéder aux quantifications. **B**. Quantifications normalisées de la luminescence totale associée à chaque souris à chaque temps d'acquisition. Une des dix souris infectées avec la souche KO a été exclue de l'analyse puisque l'infection n'a pas été productive et aucun signal luminescent n'était détecté à 6 h PI. Les moyennes sont présentées, et les souches ont été comparées statistiquement à chaque temps à l'aide de tests de Student pairés. ns non significatif, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. **C**. Courbes de survie des souris infectées. Pour des raisons éthiques, l'expérience a été interrompue à 32 h PI car plusieurs des souris infectées ont atteint un point critique.

De toute évidence, SsnA est important au maintien de la charge bactérienne en contexte infectieux, ce qui pourrait impliquer une meilleure évasion immunitaire ou alors un potentiel d'adaptation plus grand. Dans le but de mieux comprendre les implications de cette nucléase dans la virulence de *N. meningitidis*, nous développons actuellement un modèle de portage asymptomatique de cette bactérie chez la souris (données non montrées). Ce modèle permettra d'étudier l'impact de SsnA sur de plus longues périodes allant d'une semaine à plus d'un mois, donnant ainsi l'occasion d'évaluer son rôle dans l'adaptation et l'évolution chez l'hôte.

3.12 Autres phénotypes testés

La découverte des fonctions biologiques associées à SsnA présentées dans ce chapitre a été précédée de nombreux tests phénotypiques non concluants ou ne révélant aucune différence significative entre les souches exprimant et n'exprimant pas la nucléase. Afin d'éviter des longueurs superflues non essentielles à cet ouvrage, ces tests sont compilés dans le tableau suivant (**Tableau 12**). Brièvement, ces tests visaient à étudier les fonctions associées aux gènes flanqués de nombreuses copies de la NTS dans le génome de *N. meningitidis*. En effet, on retrouve parmi les protéines encodées par ces gènes plusieurs transporteurs de métaux, des pilines, des enzymes de résistance au stress oxydatif ainsi qu'une hélicase (voir **Figure 10**). D'après nos tests, SsnA n'est pas impliquée dans ces fonctions.

Tableau 12 Tests phénotypiques sans effets observés liés à SsnA

Phénotype testé	Détails
Morphologie et piliation	Microscopie électronique à transmission après croissance en milieu liquide.
Résistance aux dommages à l'ADN (mutagenèse physique, chimique, et dommages induits par antibiotique)	Dénombrement de colonies (CFUs) après exposition (1) aux UV, (2) au méthanesulfonate d'éthyle (EMS), et (3) à l'acide nalidixique (inhibiteur de la réplication d'ADN menant à des cassures double-brin).
Sensibilité au pH	Tests de croissance en milieu solide à pH 5-10
Résistance au stress oxydatif	Mesure du diamètre d'inhibition de disques de papier imbibés de peroxyde (H ₂ O ₂), de butyl hydroperoxyde (tBOOH) et de paraquat.
Sensibilité aux métaux	Croissance en milieu solide en liquide en présence de concentrations toxiques de magnésium, manganèse, fer, zinc, cobalt et nickel, aluminium, cadmium et cuivre.

CONCLUSIONS

Pour résumer et conclure ce chapitre, nous avons généré des mutants d'expression de SsnA dans une panoplie de souches de *N. meningitidis* pour étudier les fonctions biologiques de cette nucléase. La délétion ou surexpression de SsnA a peu d'impact sur le transcriptome de cette espèce. Cependant, nous avons confirmé que l'interaction entre SsnA et la séquence répétée NTS, caractérisée dans le chapitre précédent par des tests biochimiques, a bien lieu *in vivo* chez la bactérie *N. meningitidis*. En effet, l'activité ssDNase spécifique de SsnA module de façon importante l'intégration chromosomique d'ADN transformant contenant des NTS via un mécanisme probable de restriction d'ADN simple-brin. Lorsque les NTS sont absentes ou mutées, le phénotype est perdu, voire même inversé.

Pour rappel, lors de la transformation par compétence naturelle, l'ADN se retrouve sous forme simple-brin une fois au cytoplasme, avant d'être apparié au génome de l'hôte par la recombinaise RecA. Lors de l'import de gènes flanqués par des NTS, SsnA clive à proximité de la séquence répétée, ce qui a pour effet de réduire la longueur d'homologie avec le génome hôte et donc de réduire l'intégration de ces régions. Cette activité n'est pas limitée aux transformations en présence de quantités saturantes d'ADN exogène, puisque les mêmes phénotypes sont observés lors de simples cocultures de *Neisseria*, y compris lorsque le donneur est une espèce commensale. Vraisemblablement, SsnA joue un rôle important dans les HGT des *Neisseria* et tout particulièrement chez *N. meningitidis* qui possède une forte abondance de NTS dans son génome, s'accumulant autour de gènes codant pour des protéines de surface dont plusieurs facteurs de virulence. SsnA agit avant l'étape de recombinaison homologue et indépendamment de celle-ci, puisque cette nucléase n'influence pas la variation antigénique des pili, malgré la présence de nombreuses NTS autour des gènes responsables de ce mécanisme.

Bien que le mécanisme sous-jacent soit encore inexpliqué, nous révélons que SsnA est un déterminant majeur de la virulence de *N. meningitidis*. N'agissant ni sur l'adhésion aux cellules cibles ni sur leur invasion, cette nucléase atypique semble promouvoir le maintien de l'infection chez la souris en favorisant potentiellement l'évasion immunitaire. Cette hypothèse est explorée davantage dans la section discussion de cet ouvrage.

4 CHAPITRE IV – FAMILLE DES SSN

Initialement, ce projet avait pour but principal de caractériser la protéine SsnA et son interaction avec la NTS chez les *Neisseria*. Après des études bio-informatiques, biochimiques et phénotypiques approfondies révélant une enzyme unique aux fonctions biologiques particulièrement intéressantes, nous avons pris un peu de recul dans le but de préparer la suite de ce projet de recherche. En effet, l'impact important de ces découvertes ouvre la porte à de nombreux autres projets qui pourraient être axés autant sur l'aspect mécanistique et structural de SsnA que sur son rôle dans la virulence, son potentiel biotechnologique ou encore sur l'identification et la caractérisation d'homologues chez d'autres espèces. Cet ultime chapitre explore justement ce dernier point. Appuyé par des analyses bio-informatiques et quelques tests enzymatiques préliminaires, ce court chapitre a pour but d'introduire SsnA comme membre d'une vaste famille de nucléases non caractérisées aux diverses spécificités de clivage, dont le potentiel biologique et biotechnologique reste encore à définir.

MATÉRIEL ET MÉTHODES

4.1 Identification et classification des homologues de SsnA

Afin d'étendre au maximum notre recherche d'homologues, les 23 092 protéines appartenant à la famille cd10448 (GIY-YIG_unchar_3) ont été extraites de la base de données des domaines conservés du NCBI (Lu *et al.*, 2020). Cette famille non caractérisée abrite la nucléase SsnA de *N. meningitidis*. Afin de conserver uniquement les séquences non redondantes, un filtre ne gardant qu'une séquence parmi celles identiques à 90% et plus a été appliqué avec CD-HIT (Li & Godzik, 2006). Pour s'assurer de n'avoir que des homologues avec un seul domaine fonctionnel (GIY-YIG), les protéines de plus de 150 acides aminés ont été exclues à l'aide de SeqKit (Shen *et al.*, 2016). Un total de 7051 protéines homologues uniques a ainsi été obtenu afin de procéder à l'analyse de clusters et à la génération de la phylogénie de toutes les Ssn.

L'outil CLANS a été utilisé sur ces 7051 protéines pour identifier et illustrer les sous-familles SsnA, SsnB et SsnC (Frickey & Lupas, 2004). Les paramètres par défaut ont été utilisés en faisant varier le seuil de *p-value*. Ce seuil a été établi empiriquement à 0.01 puisqu'il offrait la meilleure séparation entre les trois clusters observés.

Pour la phylogénie des Ssn, les mêmes 7051 protéines homologues ont été alignées par MAFFT v7.407 (Katoh & Standley, 2013). L'arbre phylogénique a été inféré à partir de 1000 itérations par maximum de vraisemblance avec le programme IQ-TREE v2.1.2 (Hoang *et al.*, 2018; Nguyen *et al.*, 2015). Les informations taxonomiques ont été ajoutées pour chaque séquence à l'aide d'un script maison sur Python. L'arbre est illustré à l'aide du serveur iTOL v6.7.3 (Letunic & Bork, 2021).

4.2 Recherche d'éléments répétés associés aux homologues de SsnA

Pour trouver les séquences répétées associées aux homologues de SsnA, les séquences intergéniques immédiatement en amont et en aval des gènes codant pour ces protéines ont été extraites de NCBI. Ensuite, deux approches ont été utilisées : une ciblée visant à identifier des répétitions semblables aux NTS, et une seconde non ciblée visant à identifier de nouveaux éléments répétés non reliés aux NTS.

Pour la première approche, la séquence consensus des NTS de *N. meningitidis* a été cherchée par BLAST ou alignement à proximité des homologues *ssnA*. Pour la seconde approche, les régions intergéniques flanquant *ssnA* ont été soumises à l'outil LALIGN qui cherche des duplications internes par alignements locaux (Madeira *et al.*, 2022; Pearson, 2001). Les répétitions trouvées ont par la suite été validées par alignement multiple utilisant l'algorithme MAFFT, puis illustrées avec JalView (Katoh & Standley, 2013; Waterhouse *et al.*, 2009). La structure secondaire consensus des séquences alignées a été prédite avec l'outil RNAalifold en substituant les uraciles par des thymines (Bernhart *et al.*, 2008). Les structures secondaires individuelles ont quant à elles été prédites à l'aide de RNAfold avec les paramètres par défaut pour l'ADN (Gruber *et al.*, 2008).

4.3 Purifications de protéines

Des homologues de SsnA provenant de quatre espèces ont été exprimés et purifiés pour tester leur activité enzymatique. Il s'agit de l'espèce commensale *Neisseria elongata* subsp. *glycolytica* ATCC 29315, de *Neisseria wadsworthii* DSM 22245, qui possède uniquement des variantes de la NTS dans son génome (NTSvar), de *Dehalogenimonas lykanthroporepellens* BL-DC-9 du phylum très éloigné Chloroflexota, ainsi que de *Vibrio vulnificus* MO6-24/O, une gammaprotéobactérie pathogène. Le gène *ssnA* de *N. wadsworthii* a été amplifié par PCR puis cloné dans le vecteur d'expression pET-15b(+), tandis que les deux autres ont été directement

synthétisés dans ce même vecteur, générant des protéines recombinantes avec un tag polyhistidine en N-terminal (GenScript) (voir **Tableau 5**, Chapitre II).

La protéine de *N. elongata* a été purifiée de la même façon que celle de *N. meningitidis* (voir Chapitre II). Afin d'optimiser la solubilité et le rendement des purifications des autres homologues, un protocole d'expression et de purification complètement différent a été utilisé. Un milieu M9 modifié a été utilisé pour les inductions (sels M9, 0.1 M thiamine, 2 mM MgCl₂, 0.1 mM CaCl₂, 0.2% glucose, 1% casaminoacides). Deux L de culture ont été induits avec 0.1 mM d'IPTG, ou 0.2 mM pour *N. wadsworthii*, puis incubés 18 h à 14°C. La lyse a été faite à l'aide d'une *french press* dans un tampon à base de 50 mM Tris pH 7.5 et 100 mM NaCl, supplémenté avec 0.1 mM PMSF. La fraction soluble a été incubée avec la résine Ni-NTA (Thermo), puis lavée et éluée à l'aide d'un gradient d'imidazole. Les protéines ont été quantifiées par Qubit 4 (Invitrogen), visualisées par SDS-PAGE, puis conservées à -80°C dans 50% de glycérol.

4.4 Tests de nucléase

Les tests de nucléase réalisés pour ce chapitre sont très similaires à ceux présentés dans le Chapitre II. Brièvement, 0.5 µM de protéines recombinantes purifiées ont été incubées avec 0.8 µM d'oligonucléotides marqués en 5' à la fluorescéine (Sigma) dans du tampon NEBuffer 2.1 (NEB). Pour les homologues de *N. wadsworthii* et de *D. lykanthroporeppelens*, des concentrations de 0.05 µM et 2 µM respectivement ont été utilisées, car ces dernières n'ont pas la même efficacité de clivage. Les réactions ont été incubées 1h30 à 37°C, stoppées via l'ajout d'un tampon de chargement contenant de l'EDTA et inactivées à la chaleur, puis résolues par SDS-PAGE sur des gels dénaturants à 17.5% contenant 8 M d'urée, dans du tampon Tris-Borate-EDTA (TBE). Les gels ont été imagés sur un scanner Typhoon FLA 9500 (GE).

L'identité et la séquence des oligonucléotides fluorescents utilisés pour ce chapitre sont détaillées dans le (**Tableau 13**).

Tableau 13 Liste des oligonucléotides fluorescents

Nom	Séquence (5'-3')
RSE75_F	[6-fam]TTTTTATCAAATTCAAAAATATTC <u>CGTCATTCCC</u> CGCAAAGCGGGAATCTAGAACGT-AAAATCTAAAGAAACCGT
RSE75_R	[6-fam]ACGGTTTCTTTAGATTTTACGTTCTAG <u>ATTCCC</u> GCTTTTCGCGGGAATGACGGAATAT-TTTTGAATTTGATAAAAA
NTS_Nelon_F	[6-fam]TTCACCCTGGTACCCCTTGAATTT <u>CGTCATTCCC</u> CGCGCAGGCGGGAATCCAGACTC-TTAGGTTTCAGGAATCTT
NTS_Nelon_R	[6-fam]AAGATTCCTGAAACCTAAGAGTCTGG <u>ATTCCC</u> GCCTGCGCGGGAATGACGAAATTC-AAGGGGTACCAGGGTGAA
NTSvar_F (<i>Nwad</i>)	[6-fam]TGATTGATTTCGCCAAAGGCATTACCTTTTTCAGACAAGCTTGTTCTGTAC <u>CGTCATAC</u> -TCGGGCTTGACCCGAGTATCTCCCAAATATCAGGGGGGAGGT
NTSvar_R (<i>Nwad</i>)	[6-fam]AACCTCCCCCTGATATTTGGGAGATACTCGGGTCAAGCCCGAGTATGACGGTAC-AGAACAAGCTTGTCTGAAAAGGTAATGCCTTTGGCGAATCAATCA
NTS_Dlyk_F	[6-fam]ATTGTCATTACCCGCTCTCCCTTTG <u>TCATTACCC</u> CGCAAAGCGGGTAATCCACG-GTCAAATAAGCAACAGAAGT
NTS_Dlyk_R	[6-fam]ACTTCTGTTGCTTATTTGACCGTGG <u>ATTACCC</u> GCTTTTCGCGGGTAATGACAAAGG-GAAGAGCGGGTAATGACAAT
NTS_Vvulni_F	[6-fam]TCATTCAACCGTATCCGCCGCCAGT <u>CGTCATTCCC</u> GTGCACACGGGAAACCATTTA-TCCACATGCGCCGGATGCC
NTS_Vvulni_R	[6-fam]GGCATCCGGCGCATGTGGATAAATGGTTTCCCGTGTGCACGGGAATGACGACTGG-CGGCGGATACGGTTGAATGA

La NTS ou son équivalent est soulignée pour chaque oligonucléotide. RSE75 correspond à la séquence de *N. meningitidis*, tandis que NTSvar correspond à la séquence de *N. wadsworthii*. [6-FAM] : 6-carboxyfluorescéine.

RÉSULTATS ET DISCUSSION

4.5 Protéines Ssn dans le domaine bactérien

Une recherche de la séquence protéique de SsnA provenant de *N. meningitidis* dans différentes bases de données révèle que cette nucléase se classe dans la famille KEGG K07461 (endonucléase hypothétique) et cd10448 (GIY-YIG_unchar_3). Ces dernières regroupent plusieurs milliers de protéines non caractérisées provenant de l'ensemble du domaine bactérien. Comprenant une centaine d'acides aminés, ces protéines ont pour unique domaine fonctionnel le domaine GIY-YIG, mais sont toutefois représentées comme une famille à part entière de la superfamille GIY-YIG en raison de leurs différences significatives avec les familles déjà caractérisées. La majorité des membres de cette famille possède une homologie importante avec SsnA, dépassant les 50% d'identité. Dans le but de mieux définir ce groupe de nucléases homologues, nous avons récupéré l'ensemble des séquences classées dans la famille cd10448, que nous avons triées, alignées et comparées à l'aide de différents outils bio-informatiques.

En premier lieu, nous avons réalisé une analyse CLANS (Cluster ANalysis of Sequences), qui permet d'illustrer en deux ou trois dimensions les liens entre de nombreuses séquences protéiques selon des alignements pairés (**Figure 47**). Cette analyse a révélé l'existence de trois clusters distincts dont deux sont presque indissociables et regroupent la grande majorité des séquences homologues. Le troisième cluster ne comprend que quelques dizaines de protéines, dont les séquences diffèrent considérablement des deux précédents groupes mais sont très rapprochées entre elles. De façon intéressante, le cluster abritant le plus grand nombre de protéines est également celui dans lequel on retrouve le SsnA de *N. meningitidis*. Pour ces raisons, nous redéfinissons ces clusters comme les familles de nucléases hypothétiques SsnA, SsnB et SsnC.

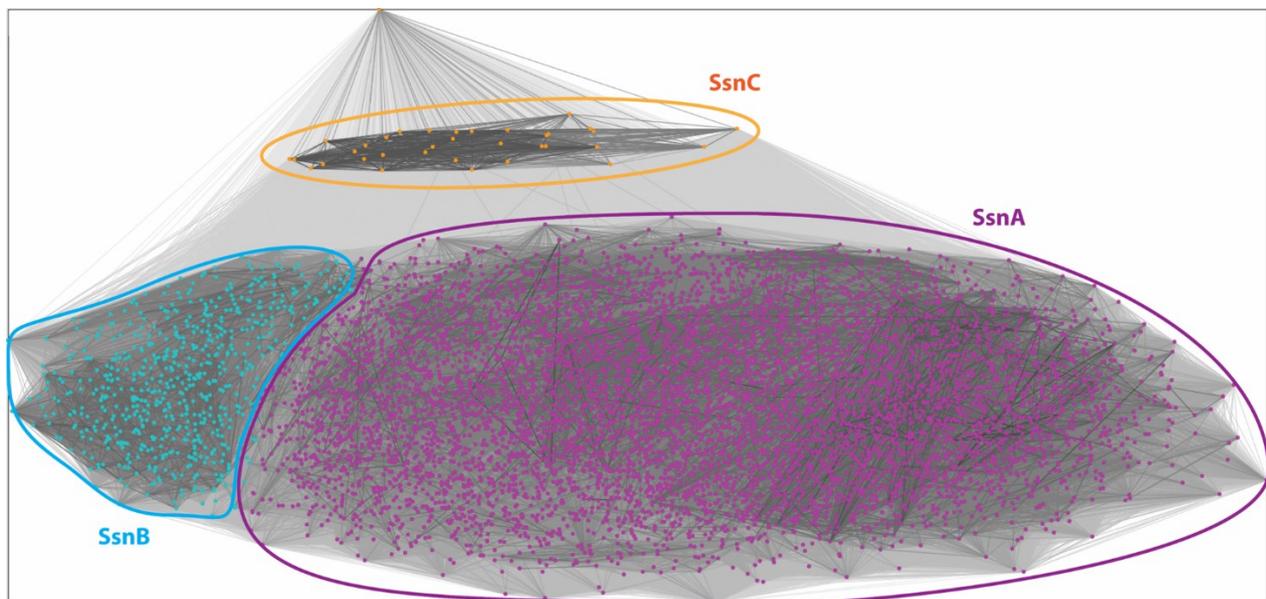


Figure 47 Analyse CLANS de la famille cd10448

Un ensemble de 7051 séquences protéiques de moins de 150 aa appartenant à la famille cd10448 a été récupéré de la base de données des domaines conservés de NCBI (CDD). L'outil CLANS a été utilisé pour générer la matrice en utilisant les paramètres par défaut et un seuil de p-value=0.01. Chaque point représente une séquence protéique, reliée aux autres par des traits donc l'intensité de couleur grise dénote le score de similarité par BLAST/PSIBLAST. Les clusters distincts sont encerclés et colorés.

Dans un second temps, nous avons voulu déterminer par des analyses plus laborieuses où se situent toutes ces protéines homologues dans un contexte évolutif. Pour ce faire, la même liste de séquences protéiques a été soumise à des alignements multiples puis utilisée pour générer une phylogénie, à laquelle chaque séquence est associée à sa distribution taxonomique (**Figure 48**). Cette analyse révèle l'incroyable diversité d'espèces bactériennes possédant un homologue de SsnA. En effet, ces protéines sont distribuées dans l'ensemble du domaine bactérien, mais en grande majorité chez des espèces à Gram négatif. Malgré les rôles de SsnA dans la virulence et la compétence naturelle de *N. meningitidis*, on constate que des homologues sont identifiés dans une panoplie d'espèces pathogènes, commensales et environnementales, dont beaucoup ne sont pas naturellement compétentes. Cette distribution suggère que ces homologues accomplissent peut-être d'autres fonctions importantes chez d'autres espèces, qui ne sont pas encore définies. Enfin, cette analyse phylogénétique permet d'observer une séparation assez claire des homologues de SsnA des différents clades bactériens, notamment entre les protéobactéries et le superphylum FCB. Cette séparation, en ligne avec la phylogénie globale du domaine bactérien, suggère une transmission principalement verticale de ces nucléases hypothétiques et supporte donc l'hypothèse que d'autres fonctions importantes y soient associées qui justifieraient leur conservation au travers de l'évolution. Les clusters identifiés comme SsnB

et SsnC par l'analyse CLANS affichent clairement une séparation phylogénétique du reste des protéines appartenant aux SsnA. En effet, les SsnB semblent avoir considérablement dérivé au sein du superphylum FCB et s'assemblent dans un groupe monophylétique supportant leur classification comme une famille à part. Ces derniers possèdent potentiellement des fonctions biologiques distinctes des SsnA. La famille SsnC quant à elle, semble avoir émergé très récemment dans le superphylum FCB à partir d'un ancêtre commun de SsnA. Ces homologues, peu nombreux et très rapprochés, appartiennent principalement au genre *Pedobacter* et sont possiblement en pleine expansion. Il est attendu que les homologues de SsnA ainsi que les SsnB et SsnC soient toutes des endonucléases d'ADN simple-brin en raison de leur homologie avec la protéine caractérisée chez *N. meningitidis*. Ces dernières ont potentiellement un rôle dans les mécanismes de réparation d'ADN ou encore dans les transferts horizontaux de gènes par conjugaison et transduction, puisque tous ces processus font intervenir de l'ADN simple-brin.

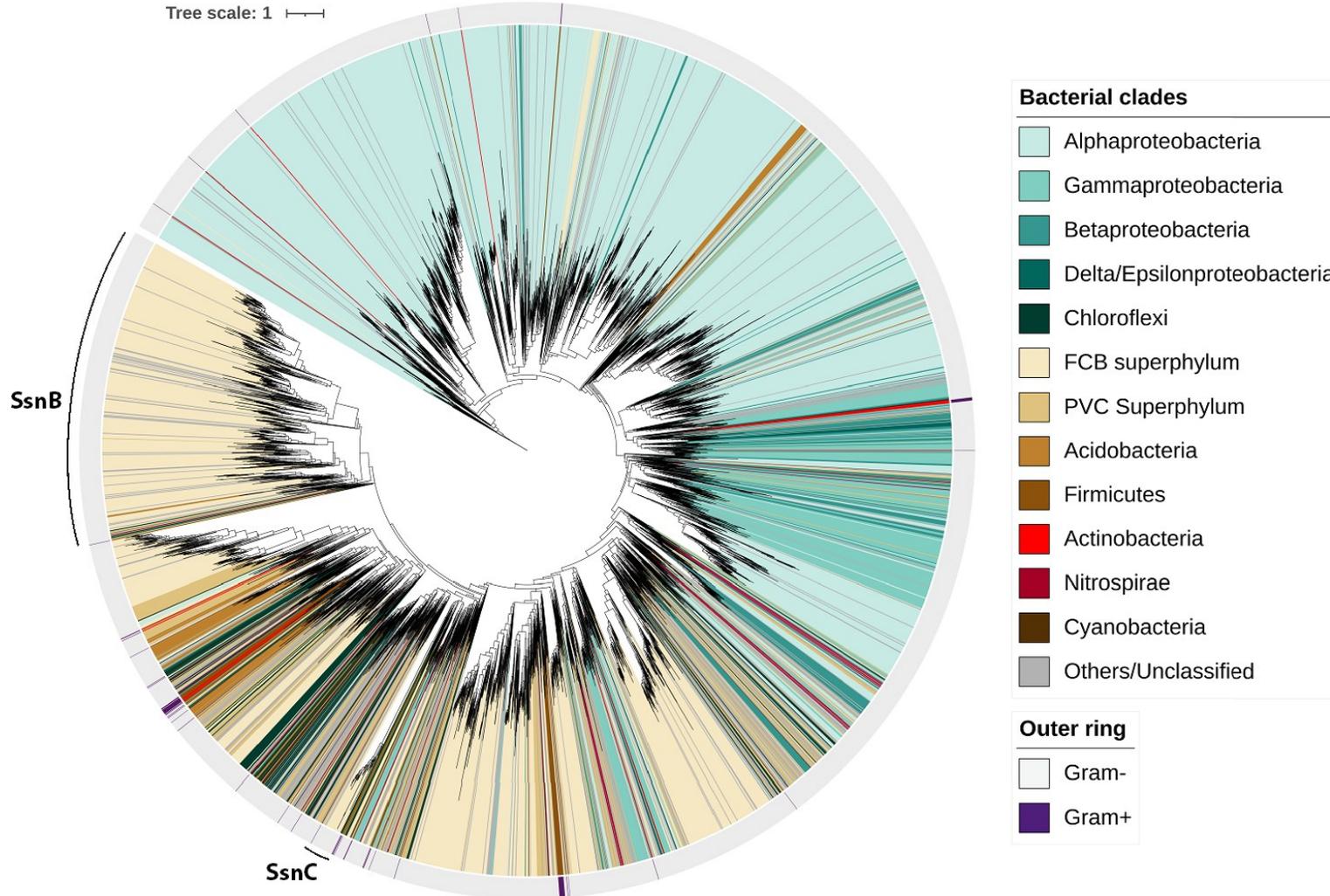


Figure 48 Phylogénie de la famille cd10448

Les 23 092 séquences protéiques de la famille cd10448 ont été extraites de la base de données CDD, puis filtrées pour ne garder que les séquences de moins de 150 acides aminés possédant au moins 50% d'identité avec la séquence du SsnA de *N. meningitidis*. À la suite d'un alignement multiple par MAFFT des 7051 séquences restantes, une phylogénie a été inférée par maximum de vraisemblance avec l'outil IQ-TREE, puis visualisée via l'outil iTOL. Chaque couleur désigne un phylum bactérien. L'anneau externe différencie les espèces à Gram positif (violet) des espèces à Gram négatif (gris). Les familles orthologues SsnB et SsnC identifiées lors de l'analyse CLANS sont indiquées.

4.6 Association des Ssn avec des séquences similaires aux NTS

Dans le tout premier chapitre de cet ouvrage, nous avons fait état de la forte association par proximité génomique entre le gène *ssnA* et la NTS chez les *Neisseria*, en plus d'observer que la NTS est également retrouvée à proximité des homologues de *ssnA* chez des espèces distantes (voir **Tableau 4**, Chapitre I). Après avoir déterminé que les homologues de SsnA forment une grande famille monophylétique de nucléases GIY-YIG, nous avons tenté d'explorer davantage l'association entre ces protéines et des éléments répétés dans tout le domaine bactérien. Pour ce faire, nous avons d'abord procédé à une recherche ciblée de NTS à proximité des homologues de *ssnA* par BLAST sur les régions intergéniques flanquant ces gènes (**Figure 49**). Cette analyse a détecté des NTS à côté des *ssnA* provenant de différents phyla, dont des protéobactéries, deinococota, thermodesulfobacteria et chlorobiota, confirmant que cette association n'est pas du tout limitée aux *Neisseria*. De façon intéressante, ces différentes variantes espèce-spécifiques des NTS affichent presque toutes un noyau palindromique riche en GC ainsi que la séquence d'amorce GTCA et le dinucléotide terminal CC. (**Figure 49A**). Ces dernières leur confèrent une structure prédite en tige-boucle avec une orientation définie par les extrémités non palindromiques (**Figure 49B**), identique à celle retrouvée chez les *Neisseria*. Cette structure est hautement conservée parmi toutes les espèces analysées malgré les multiples variations de séquences observées, indiquant que ces substitutions correspondent à des covariations. Ces observations supportent davantage une probable coévolution entre les nucléases et leurs NTS respectives et suggèrent fortement que ces enzymes clivent également à proximité des éléments répétés qui leur sont associés.

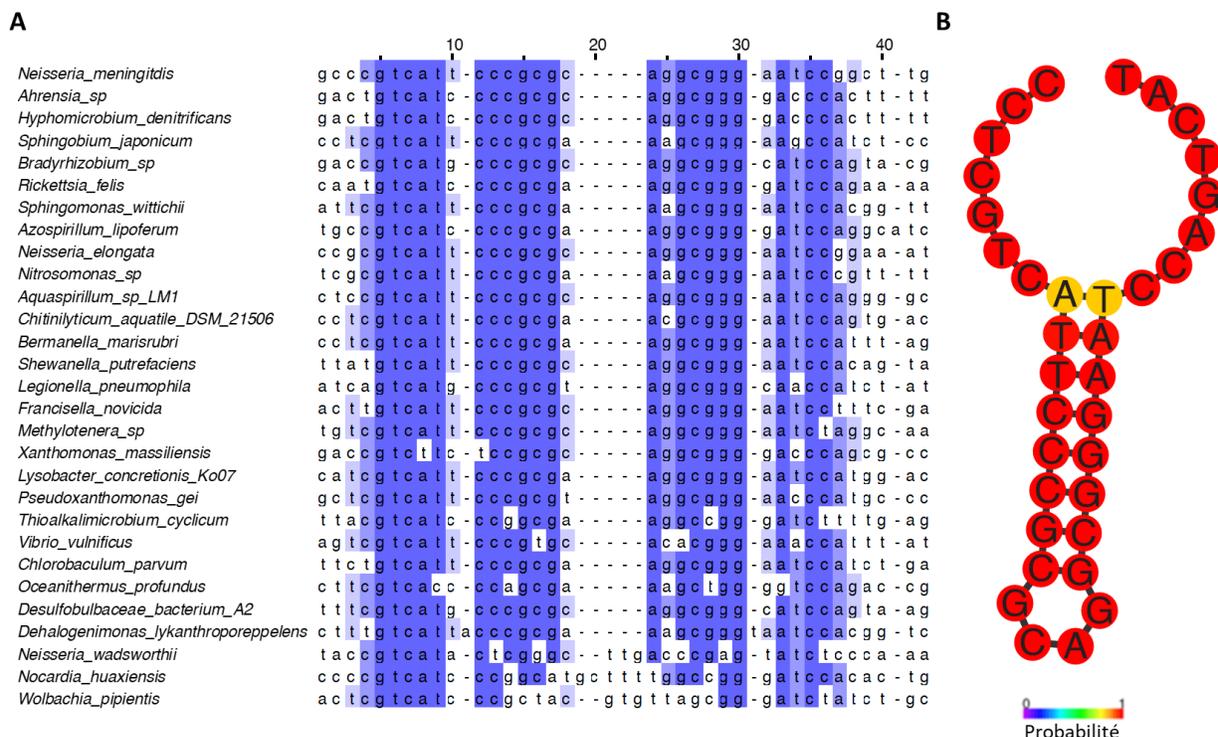


Figure 49 Homologues de SsnA associés à des éléments répétés similaires aux NTS

L'élément répété NTS a été recherché par BLAST dans les régions génomiques immédiatement en amont et en aval des gènes codant pour des homologues de SsnA répartis dans l'ensemble de la phylogénie de cette famille. **A.** Alignement multiple par MAFFT des séquences fortement similaires aux NTS trouvées à proximité des homologues SsnA. L'intensité du bleu indique la conservation de chaque nucléotide. **B.** Prédiction de la structure secondaire consensus à partir des séquences en A avec l'outil RNAalifold. Les couleurs indiquent la probabilité d'appariement. Les uraciles (U) ont été manuellement substituées en thymines (T).

La présence de plusieurs substitutions au sein des NTS associées aux *ssnA*, ainsi que la découverte d'une variante de la NTS conservée chez certaines *Neisseria* laissent croire à l'existence d'autres variantes espèce-spécifiques au travers de la phylogénie des SsnA (voir **Figure 18**, Chapitre I). Pour identifier des répétitions associées à différents homologues sans chercher spécifiquement les NTS des *Neisseria*, nous avons procédé à une recherche d'éléments dupliqués dans les régions intergéniques immédiatement devant et derrière les gènes *ssnA*. Cette méthodologie a révélé la présence de plusieurs séquences répétées sensiblement différentes mesurant 20 à 40 nt. Lorsqu'illustrées en relation avec leur gène *ssnA* respectif, on constate que toutes ces répétitions se trouvent en direction opposée l'une de l'autre immédiatement devant le promoteur et derrière le codon STOP de la nucléase, parfois en plusieurs copies (**Figure 50A**). Ces dernières partagent avec la NTS un noyau semi-palindromique riche en GC, ce qui leur donne également une structure prédite en épingle à cheveux (**Figure 50B**). Ces structures

affichent une diversité de longueurs de tige et de boucle, et possèdent parfois des mésappariements dans leur région palindromique causant un renflement dans la tige.

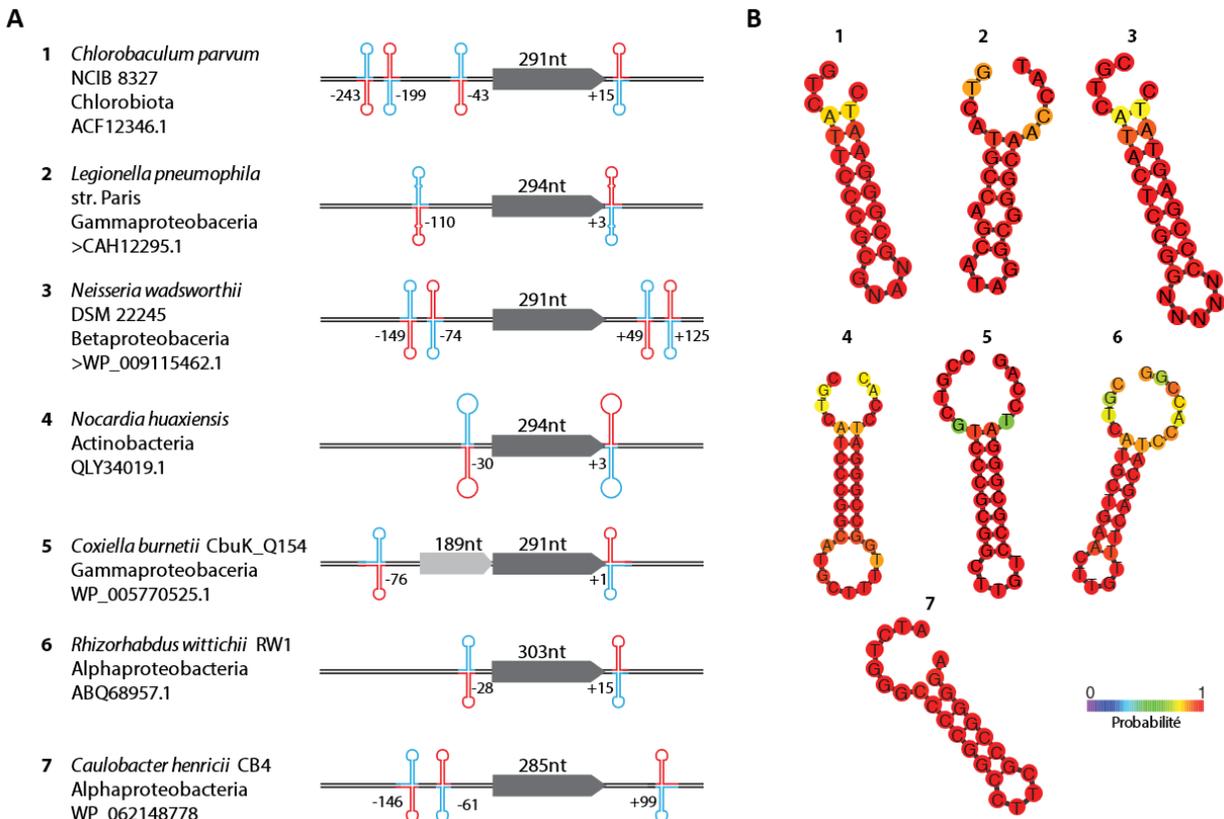


Figure 50 Homologues de *ssnA* associés à d'autres répétitions palindromiques

Les séquences intergéniques immédiatement en amont et en aval de plusieurs homologues de *ssnA* ont été soumises à une recherche d'éléments dupliqués par l'outil LALIGN. **A.** Représentation graphique des régions génomiques abritant les homologues *ssnA* (flèches grises). L'espèce, la souche, la classification taxonomique et le numéro de référence des gènes correspondant à *ssnA* sont indiqués, de même que l'orientation de leurs séquences répétées respectives et leur position par rapport à *ssnA*. La couleur des répétitions indique leur orientation. **B.** Prédications de la structure secondaire des répétitions numérotées en A par l'outil RNAfold. Les couleurs indiquent la probabilité d'appariement de chaque nucléotide.

La découverte de différentes séquences répétées autour des *ssnA* d'espèces lointaines, de même que leur distribution particulière, soulève plusieurs interrogations et hypothèses qui sont discutées plus en détail dans une prochaine section. D'abord, il est attendu que chaque homologue de *ssnA* encode pour une nucléase capable de cliver spécifiquement à proximité des séquences répétées qui l'entourent, tel que démontré chez *N. meningitidis*. Ensuite, ces répétitions se classent dans la famille des *Repetitive Extragenic Palindromes* (REP), qui lorsque retrouvés en tandem forment des *Bacterial Interspaced Mosaic Elements* (BIME). Ces éléments répétés sont fréquemment associés à des séquences d'insertion (IS) et sont la cible de transposases (Tobes & Pareja, 2006). Les gènes *ssnA* associés à ces REP ressemblent justement de façon frappante à des IS, soulevant la possibilité que certaines de ces nucléases hypothétiques soient en fait des transposases.

4.7 Caractérisation enzymatique d'homologues distants de SsnA

4.7.1 Expression et purifications de protéines

Dans le but de confirmer expérimentalement les précédentes observations *in silico*, nous avons procédé à l'expression, la purification et la caractérisation enzymatique de plusieurs homologues de SsnA provenant de l'ensemble du domaine bactérien. Après plusieurs échecs dus à des problèmes de solubilité, quatre protéines ont été obtenues avec une pureté, une concentration et une stabilité satisfaisantes (**Figure 51**). La première provient de *Neisseria elongata*, une espèce commensale dont les NTS sont identiques à celles de *N. meningitidis* et dont la protéine SsnA est très similaire. La seconde provient de *Neisseria wadsworthii*, une des quelques *Neisseria* qui possède de nombreuses copies d'une variante de la NTS (NTSvar) dans son génome, notamment de part et d'autre de son gène *ssnA*. La troisième protéine provient de *Dehalogenimonas lykanthroporepellens*, une bactérie environnementale anaérobie stricte du phylum Chloroflexota. Son génome possède plusieurs dizaines de séquences quasi identiques aux NTS, mais qui affichent toutes une insertion parfaitement conservée de deux nucléotides palindromiques (GTCATTACCCGC[N₄]GCGGGIAATC). La dernière protéine provient de *Vibrio vulnificus*, une gammaprotéobactérie pathogène. Cette dernière ne possède dans son génome que cinq copies de la NTS, dont une immédiatement en amont de son *ssnA* mais pas en aval. Il est possible qu'une autre variante de la NTS ou une autre séquence répétée de la famille des REP y soit plus abondante, mais nous n'avons pas été en mesure de trouver un tel élément. Ces quatre homologues possèdent respectivement 82%, 66%, 49% et 57% d'identité en acides aminés par rapport à la protéine de *N. meningitidis*, et pas plus de 58% d'identité entre eux à l'exception des

deux *Neisseria*. Ils forment donc un échantillon représentatif de la diversité de cette vaste famille d'enzymes. Des tentatives de purification et de caractérisation de plusieurs autres protéines ont été faites, notamment les homologues provenant de *Legionella pneumophila*, *Coxiella burnetii*, *Rickettsia felis* et *Nocardia huaxiensis*. Ces dernières n'ont cependant pas été fructueuses.

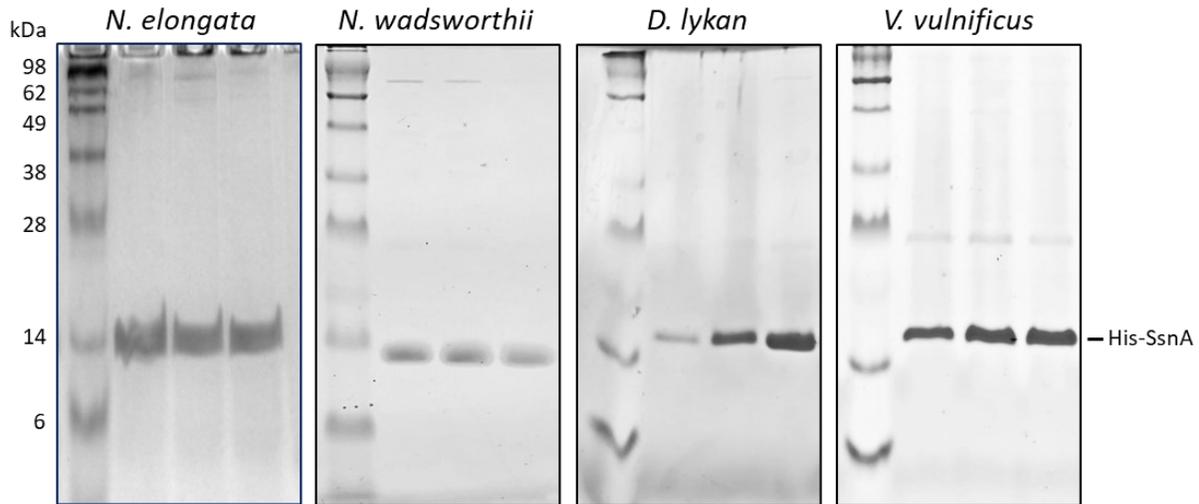


Figure 51 Purifications de protéines homologues à SsnA

SDS-PAGE des fractions d'élution des purifications d'homologues de SsnA provenant de *Neisseria elongata*, *Neisseria wadsworthii*, *Dehalogenimonas lykanthroporepellens*, et *Vibrio vulnificus*.

4.7.2 Tests d'activité nucléase

Puisque la protéine de *N. meningitidis* clive de façon spécifique de l'ADN simple-brin contenant la NTS, nous avons d'abord vérifié si les homologues de SsnA ont une activité enzymatique similaire. Pour ce faire, chaque protéine a été incubée en présence d'oligonucléotides fluorescents contenant la séquence répétée analogue aux NTS propre à chaque génome. Afin de démontrer leur spécificité, les mêmes tests ont également été réalisés sur de l'ADN simple-brin complémentaire inverse ainsi que sur de l'ADN double-brin de même séquence (**Figure 52**). Tout comme la protéine de *N. meningitidis* (**A**), celles de *N. elongata* (**B**), *N. wadsworthii* (**C**) et *D. lykanthroporeppelens* (**D**) clivent leurs séquences respectives sous forme simple-brin avec une orientation définie. Cependant, les homologues de *N. wadsworthii* et de *D. lykanthroporeppelens* semblent cliver leur substrat à différentes positions par rapport aux enzymes de *N. meningitidis* et *N. elongata*. Il est à noter ici que la NTS retrouvée dans le substrat de *N. wadsworthii* est décalée d'une vingtaine de nucléotides par rapport aux NTS des autres substrats, signifiant qu'un produit de clivage de même taille que les autres substrats correspond à un site de clivage différent.

L'homologue de *V. vulnificus* affiche quant à lui une activité ssDNase minime, difficile à reproduire car trop proche du seuil de détection de la technique, mais qui semble néanmoins spécifique à son substrat. Ces premiers tests confirment que la plupart de ces enzymes sont bien des endonucléases spécifiques d'ADN simple-brin (ssDNase) capables de cliver les séquences répétées similaires aux NTS retrouvées dans leur génome à proximité du gène *ssnA* ou ailleurs.

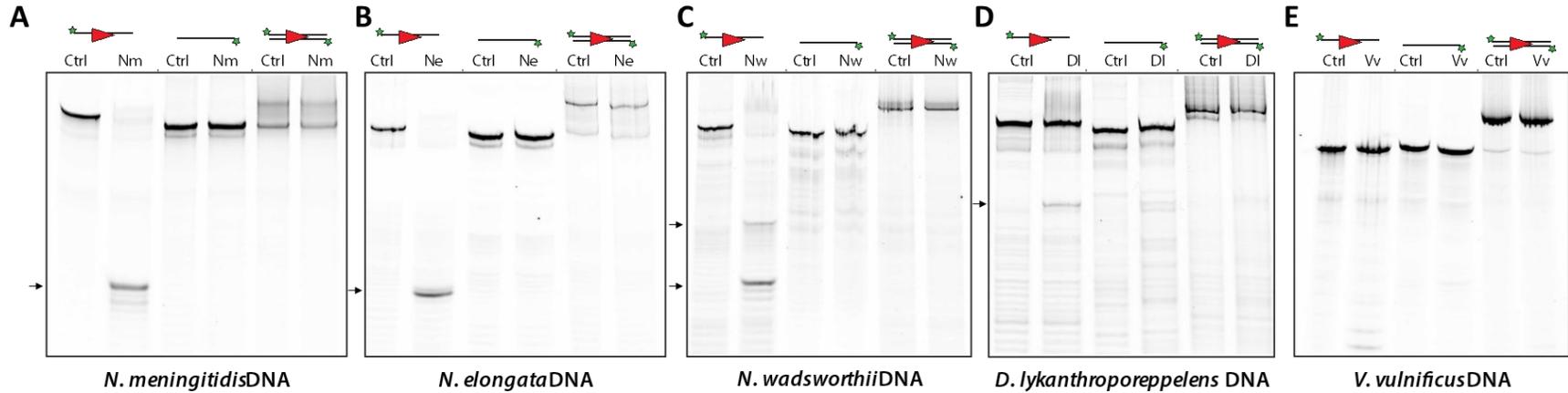


Figure 52 Activité nucléase des homologues SsnA sur leurs REP respectives

Test de nucléase où chaque homologue a été testé avec la séquence répétée qui lui est associée sous forme d'ADN simple-brin, simple-brin complémentaire inverse, et double-brin. **A.** Protéine de *N. meningitidis* (*Nm*) **B.** de *N. elongata* (*Ne*) **C.** de *N. wadsworthii* (*Nw*) **D.** de *D. lykanthroporeppelens* (*DI*) **E.** de *V. vulnificus* (*Vv*). Les produits de clivage sont pointés.

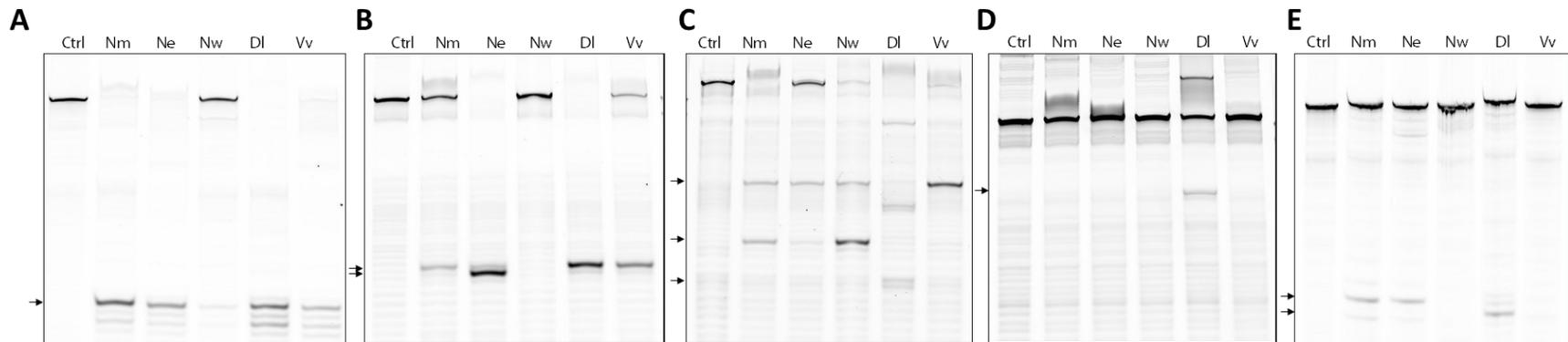


Figure 53 Spécificité de clivage des homologues de SsnA

Test de nucléase où les séquences associées à chaque SsnA ont été testées avec toutes les protéines. **A.** Séquence contenant une NTS provenant du génome de *N. meningitidis* **B.** Séquence avec NTS de *N. elongata* **C.** Séquence avec NTS_{var} de *N. wadsworthii* **D.** Séquence avec la NTS de *D. lykanthroporeppelens* **E.** Séquence avec une NTS de *V. vulnificus*. Les produits de clivage sont pointés.

Bien que les oligonucléotides testés précédemment correspondent aux répétitions propres à chaque génome d'où provient les homologues SsnA, certaines de ces séquences sont similaires. Celle de *N. elongata* par exemple, possède une NTS identique à la séquence testée de *N. meningitidis*, et ne diffère de celle-ci que par les extensions de part et d'autre de la NTS. Nous avons donc dans un second temps testé la capacité de chaque SsnA de cliver les séquences avec NTS provenant des autres espèces (**Figure 53**). Toutes les protéines clivent la séquence de *N. meningitidis* à la même position, bien que *N. wadsworthii* montre une activité nettement plus faible pour ce substrat (**A**). De façon surprenante, des résultats similaires sont observés pour la séquence de *N. elongata*, sauf que le SsnA de *N. elongata* coupe cette séquence avec un décalage d'un nucléotide par rapport aux autres enzymes (**B**). Pour la séquence de *N. wadsworthii*, plusieurs différents produits de clivage sont obtenus selon les protéines utilisées, mais la meilleure activité nucléase est associée aux SsnA de *N. wadsworthii* et de *V. vulnificus*, qui pourtant ne clivent pas au même endroit (**C**). La séquence de *D. lykanthroporeppelens*, à l'inverse des précédente, est uniquement clivée par sa propre protéine, mais seulement à de très faibles niveaux (**D**). Enfin, la séquence de *V. vulnificus* est faiblement clivée par les protéines de *N. meningitidis*, *N. elongata* et *D. lykanthroporeppelens*, mais pas par sa propre protéine dans les conditions testées (**E**).

Ces résultats, bien qu'exploratoires, suggèrent que les différents homologues de SsnA ont des spécificités complexes et distinctes. Même si plusieurs enzymes différentes peuvent cliver une même séquence, les sites de clivage semblent varier d'une protéine à une autre, de même que l'efficacité de clivage. À l'exception du SsnA de *V. vulnificus*, toutes les protéines testées peuvent cliver la séquence répétée analogue aux NTS retrouvée dans leur génome respectif, suggérant que ces protéines accomplissent également chez ces espèces diverses fonctions liées à l'intégrité génomique ou à la diversité génétique, tout comme chez *N. meningitidis*. Il est probable que l'homologue de *V. vulnificus* clive une autre séquence analogue aux NTS que nous n'avons pas été en mesure d'identifier, puisque cette enzyme clive efficacement les séquences provenant des différentes *Neisseria*. Évidemment, il est attendu que les conditions et le substrat optimal de clivage divergent d'une protéine à une autre vu leur grande diversité et disparité phylogénétique. Néanmoins, ces premières tentatives de caractérisation de la grande famille des NTS indique que la majorité de ces nucléases sont bien spécifiques à l'ADN simple-brin, confirmant leur classification dans une famille de protéines GIY-YIG distinctes de celles déjà décrites.

CONCLUSIONS

Pour résumer et conclure cet ultime chapitre, nous avons cherché à étendre les résultats de cette étude au-delà de *N. meningitidis* en explorant différents homologues de cette enzyme dans l'ensemble du domaine bactérien. Plusieurs milliers d'homologues ont été identifiés, tous jusqu'alors non caractérisés. Ces derniers peuvent être regroupés en trois sous-familles distinctes que nous avons appelées SsnA, SsnB et SsnC. Une analyse phylogénétique exhaustive de ces homologues révèle l'énorme diversité de la grande famille des Ssn, dont des membres sont retrouvés dans de nombreux phyla bactériens, avec une forte prépondérance chez les espèces à Gram négatif. Loin d'être particulièrement récentes ou ancestrales, les SsnA des *Neisseria* se trouvent au beau milieu de l'arbre phylogénétique, suggérant que d'autres homologues appartenant à des espèces non pathogènes ou non douées de compétence naturelle ont également une activité catalytique et des fonctions qui leur sont propres.

En explorant la diversité des homologues de SsnA, nous avons révélé que nombre d'entre eux sont associés à des séquences répétées de la famille des REP, fréquemment retrouvées immédiatement en amont et en aval de ces gènes. Bien que la plupart des REP associés à des homologues de SsnA s'apparentent fortement aux NTS des *Neisseria* pathogènes, plusieurs autres affichent une séquence qui en diverge davantage tout en conservant la même structure secondaire prédite. C'est le cas notamment de *Neisseria wadsworthii*, dont le génome abrite de nombreuses copies d'une variante conservée de la NTS (NTS_{var}, voir Chapitre I).

Quatre homologues de SsnA ont été purifiés et soumis à des tests préliminaires d'activité nucléase pour explorer la diversité enzymatique de cette grande famille; deux provenant d'espèces commensales de *Neisseria*, et deux provenant d'espèces éloignées. Ces tests indiquent que tous les homologues étudiés sont également des endonucléases d'ADN simple-brin (ssDNase) spécifiques. Bien que des résultats encourageants semblent supporter l'hypothèse que chaque homologue clive la séquence répétée qui lui est associée, il est encore difficile de se prononcer sur leur spécificité exacte. Néanmoins, cette exploration en surface de la famille des Ssn semble prometteuse et ouvre la voie à de nombreuses études, tant sur le potentiel biotechnologique de ces enzymes que sur leurs fonctions biologiques.

5 DISCUSSION GÉNÉRALE

5.1 Importance de l'interaction SsnA-NTS chez les *Neisseria*

5.1.1 NTS : Nouvelle séquence répétée des *Neisseria*

Les séquences répétées dRS3, bien plus fréquentes dans le génome des *Neisseria* pathogènes que des espèces commensales, ont soulevé de nombreuses hypothèses depuis leur découverte il y a plus de vingt ans. Des études *in silico* ont associé ces éléments répétés à des zones chaudes de recombinaison et à des gènes codant pour des protéines de surface (Bentley *et al.*, 2007; Budroni *et al.*, 2011; Parkhill *et al.*, 2000). Toutefois, aucune étude expérimentale ne s'est intéressée à l'importance biologique de ces séquences répétées, à l'exception de deux équipes qui ont démontré que certaines dRS3 peuvent être ciblées par une intégrase de phage (Bille *et al.*, 2005; Kawai *et al.*, 2005).

En réexaminant ces répétitions dans un large ensemble de génomes de *Neisseria* commensales et pathogènes, nous avons identifié les NTS, plus longues et complexes que la séquence dRS3 décrite dans la littérature. En plus du palindrome permettant la formation de structures en épingle à cheveux, les NTS possèdent quelques nucléotides supplémentaires hautement conservés en amont et en aval, qui eux ne sont pas parfaitement palindromiques. Ces extensions, cruciales à l'interaction avec SsnA, donnent à ces éléments une orientation définie, chose qui n'avait jamais été décrite pour les dRS3. En raison de ces différences majeures et pour mieux refléter leur impact biologique, nous avons nommé ces séquences répétées étendues NTS pour *Neisseria transformation sequences*.

La découverte d'un variant de la NTS exclusivement retrouvé chez certaines espèces phylogénétiquement proches dont *N. wadsworthii* et *N. canis* indique sans ambiguïté que ces éléments répétés coévoluent avec leur hôte. Un parallèle peut être tracé entre cette découverte et les DUS, qui eux aussi se déclinent en différents variants spécifiques à certaines espèces (Frye *et al.*, 2013). Ces observations supportent l'utilisation des éléments répétés des *Neisseria* dans leur classification phylogénétique et l'étude de leur évolution.

En raison de leur présence dans presque toutes les *Neisseria*, les NTS proviennent probablement d'un ancêtre commun. Une expansion nette de ces éléments est observée dans le clade prépathogène qui regroupe les deux espèces pathogènes ainsi que *N. lactamica*, *N. polysacchareae*, *N. bergeri* et plusieurs autres. Au contraire, *N. cinerea*, qui est à la base de ce clade, ne possède aucune copie de la NTS. Cette disparité suggère que l'expansion des NTS

dans un ancêtre commun a débuté après la spéciation de *N. cinerea*. Il est intéressant de noter que cette espèce est la seule parmi celles mentionnées précédemment qui n'encode pas *ssnA*, alors que toutes les autres partagent un gène quasi identique. Les espèces commensales plus éloignées, qui ont également peu ou pas de NTS dans leur génome, n'encodent généralement pas non plus de *ssnA*. Cette observation suggère fortement une coévolution des NTS et de SsnA au sein des *Neisseria*. Puisque SsnA clive spécifiquement à proximité de ces séquences répétées, tel que démontré dans cette étude, cette nucléase est probablement impliquée dans la propagation des NTS via la transformation par compétence naturelle. Il existe néanmoins quelques souches et espèces qui possèdent très peu de NTS mais qui encodent pour *ssnA* et vice-versa. Étant donné la forte fréquence de HGT chez ces espèces dont la majorité cohabite dans les voies aériennes supérieures, il est probable que certaines souches aient récemment acquis *ssnA* sans avoir eu le temps d'accumuler des NTS. À l'inverse, on peut supposer que certaines souches abritant de nombreuses copies de NTS aient récemment perdu *ssnA*. À l'heure actuelle, nous ne savons pas si l'activité ssDNase de SsnA est également impliquée dans les remaniements chromosomiques. Cette hypothèse est cependant tout à fait plausible et mériterait d'être explorée, puisque de l'ADN simple-brin est formé de façon transitoire à plusieurs étapes de la réplication et de la réparation du génome. Lors de l'écriture de cette thèse, un article a été publié révélant que la recombinaison homologue lors de la transformation par compétence naturelle est justement initiée aux fourches de réplication chez *S. pneumoniae* (Johnston *et al.*, 2023).

5.1.2 Médiation des transferts horizontaux de gènes par SsnA

Malgré sa petite taille et son unique domaine fonctionnel, SsnA s'est révélée être un important médiateur des transferts horizontaux de gènes (HGT) chez *N. meningitidis* via la transformation par compétence naturelle. Cette endonucléase semble agir à l'étape du traitement de l'ADN transformant simple-brin (ADNsb) dans le cytoplasme, avant sa recombinaison au génome de l'hôte. En effet, notre étude démontre que SsnA clive de façon spécifique l'ADNsb à proximité des séquences répétées NTS. Cette activité ssDNase a pour effet de moduler le taux d'intégration chromosomique de l'ADN transformant selon la présence et la distribution de NTS au sein de celui-ci. Un modèle simplifié du rôle de l'interaction SsnA-NTS est illustré en **Figure 54**.

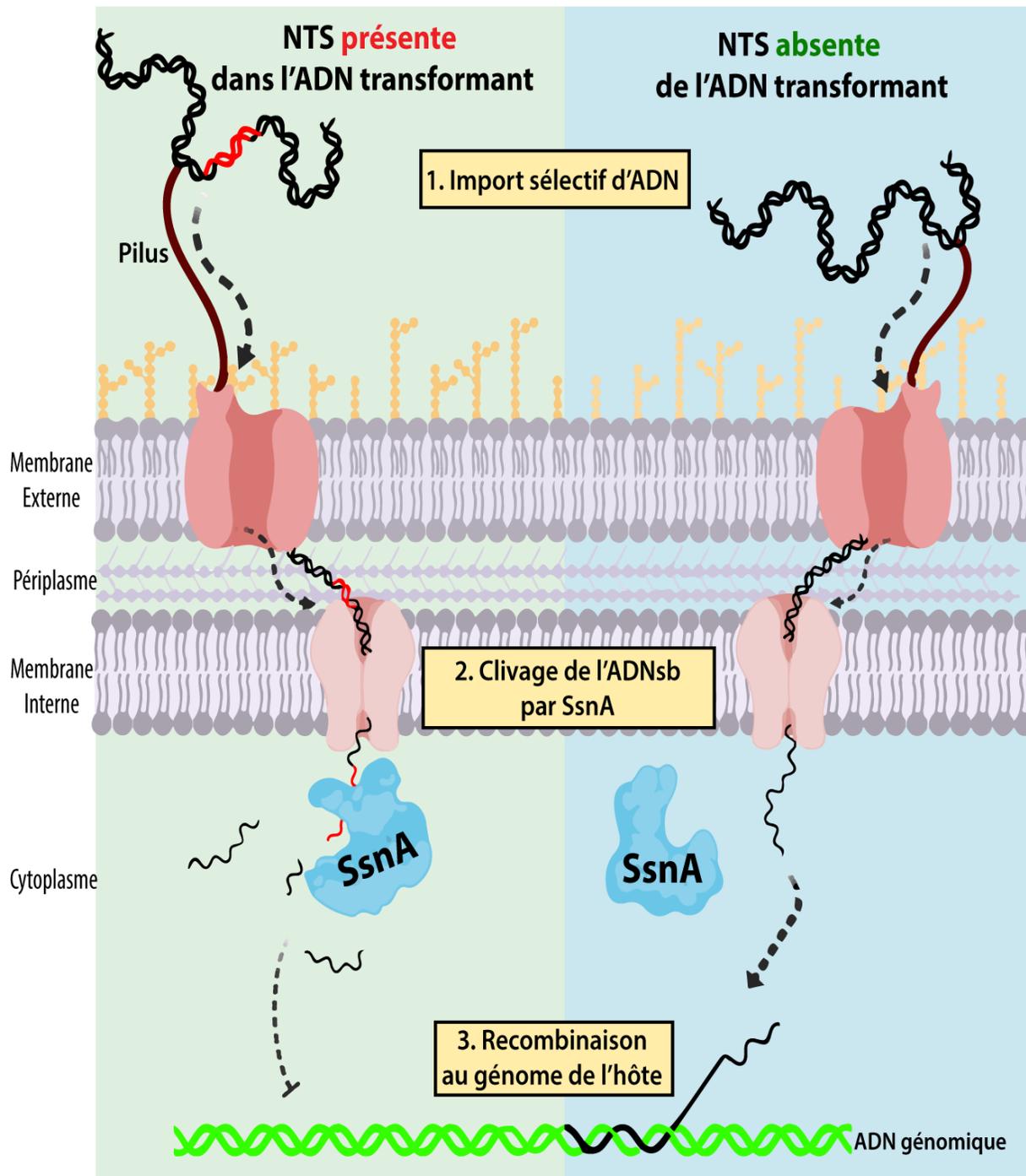


Figure 54 Modèle de transformation par compétence naturelle

La machinerie de compétence naturelle des *Neisseria* importe activement dans son périplasma de l'ADN double-brin provenant majoritairement d'autres *Neisseria*. Seul un brin est transporté au cytoplasme. Si cet ADN contient la séquence répétée NTS, SsnA coupe à proximité de celle-ci, ce qui affecte ses chances d'être intégré au chromosome de l'hôte par recombinaison homologue. En absence de NTS, SsnA n'interagit pas avec l'ADN transformant, ce qui lui permet d'être recombiné efficacement au génome.

Image créée avec BioRender.com

Nous avons démontré que des gènes flanqués de part et d'autre par des NTS sont transformés plus efficacement dans une souche n'exprimant pas *ssnA* que dans une souche exprimant la nucléase. Cet effet est indépendant de l'origine et de la méthylation de l'ADN transformant puisque le même phénotype est obtenu avec de l'ADN génomique et des plasmides. Ces résultats indiquent que SsnA diminue le taux de recombinaison homologe de ces régions au génome de l'hôte via un mécanisme de restriction d'ADN simple-brin. L'ADNsb transformant est habituellement protégé des nucléases cytoplasmiques de l'hôte par la liaison rapide des protéines SsbA, SsbB et DprA, qui bloquent physiquement l'accès aux exonucléases et DNAses (Dwivedi *et al.*, 2013). Nous avons cependant démontré que SsnA possède une affinité particulièrement forte pour de l'ADNsb contenant une NTS. Deux scénarios sont donc possibles; (1) SsnA compétitionne avec les Ssb et DprA pour cliver l'ADNsb avant qu'il soit protégé, ou bien (2) SsnA est capable de simplement déloger ces protéines pour accomplir son activité nucléase. Cette deuxième option est tout à fait plausible puisque le complexe DprA-RecA, est capable de prendre la place de SsbA sur l'ADNsb malgré une affinité nettement plus faible (Yadav *et al.*, 2014). La **Figure 55** illustre le mécanisme proposé de restriction d'ADNsb par SsnA. Concrètement, la fragmentation de l'ADN transformant médiée par SsnA diminue sa taille et donc la longueur de l'homologie avec le génome. Plusieurs études ont effectivement démontré une corrélation linéaire entre la longueur de l'homologie et la fréquence de recombinaison, et ce dans plusieurs espèces naturellement compétentes (Dalia *et al.*, 2014; de Vries & Wackernagel, 2002; Kung *et al.*, 2013).

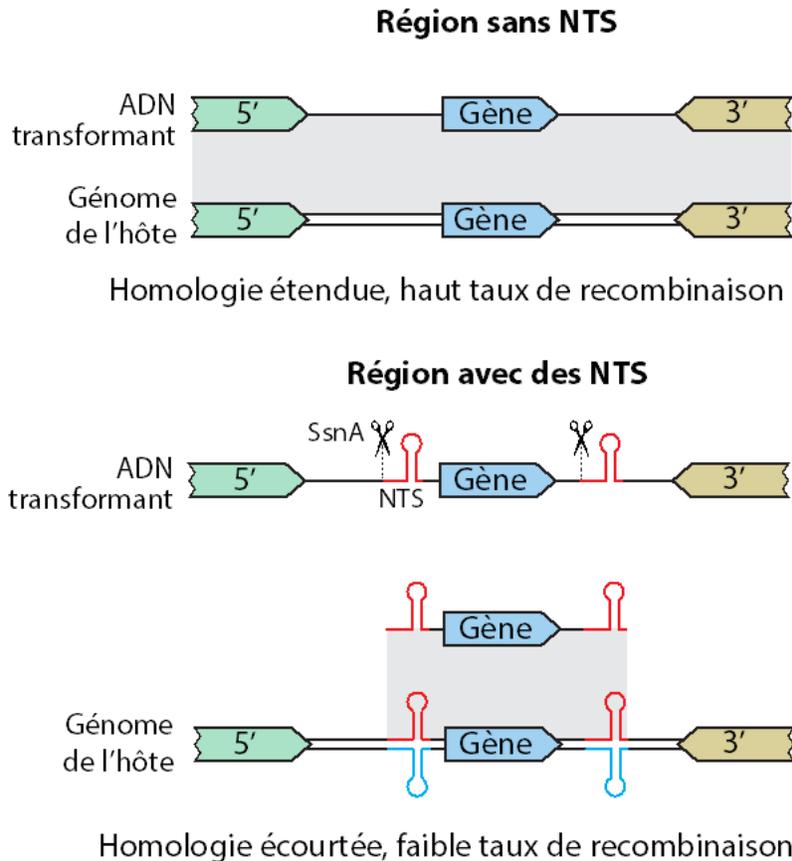


Figure 55 Mécanisme proposé de restriction d'ADN transformant par SsnA

En absence de NTS dans l'ADN transformant, RecA catalyse sa recombinaison au génome en alignant les régions homologues. Lorsque l'ADN transformant contient un gène flanqué par des NTS, SsnA clive à proximité de ces séquences, générant des fragments plus courts. L'homologie entre ces fragments et le génome est alors moins étendue que si l'ADN était resté intact, ce qui réduit le taux de recombinaison.

Les différences de taux de transformation entre des mutants de *ssnA* sont facilement justifiables lorsque l'ADN transformant contient des NTS. Cependant, notre étude révèle également qu'un léger phénotype inverse est parfois observé quand l'ADN ne contient pas ces séquences répétées, tout particulièrement lorsque *ssnA* est surexprimée. En effet, certains plasmides sans NTS sont transformés plus efficacement dans une souche de *N. meningitidis* surexprimant *ssnA* que dans une souche n'exprimant pas la nucléase. Toutefois, il est important de noter ici que ce phénotype, bien que statistiquement significatif pour certaines constructions testées, est beaucoup moins prononcé que le phénotype inverse obtenu lorsque des NTS se retrouvent dans l'ADN transformant. De plus, ces différences ne sont pas reproduites systématiquement avec d'autres plasmides s'intégrant ailleurs dans le génome. Ces observations ne sont pas expliquées par nos données *in vitro*, puisque nous avons démontré que SsnA interagit avec l'ADN simple-brin uniquement lorsqu'il contient la NTS. Cependant, nos tests enzymatiques ont été réalisés

avec de la protéine isolée pure et ignorent toute potentielle interaction avec d'autres effecteurs de la compétence naturelle et de la recombinaison homologue. Il est donc possible que SsnA interagisse avec d'autres protéines impliquées dans le mécanisme de recombinaison et module leur activité, indépendamment de la présence de NTS. Ceci nécessiterait une étude plus approfondie de l'interactome de cette nucléase.

Il est important de noter ici que bien que cette étude ait été réalisée chez *N. meningitidis*, nous nous attendons à ce que ces conclusions soient transposables à *N. gonorrhoeae* ainsi qu'à plusieurs espèces commensales. En effet, le gonocoque possède de nombreuses NTS dans son génome ainsi qu'un gène identique au *ssnA* de *N. meningitidis*. De leur côté, les espèces commensales possèdent relativement peu de NTS, mais plusieurs encodent une protéine identique à près de 80% avec un site catalytique conservé. Malgré ces légères différences, nos tests enzymatiques avec la protéine purifiée de *N. elongata* indiquent que cette nucléase a une activité et spécificité *in vitro* très similaire à celle du méningocoque (voir Chapitre IV). Lorsque nous avons tenté de reproduire les phénotypes de transformation dans *N. gonorrhoeae*, nous avons été confrontés à des taux de transformation trop faibles pour être quantifiés avec certaines constructions génétiques (données non montrées). Étant donné la grande variabilité interespèce des régions riches en NTS qui nous intéressaient, il faudra recréer des constructions génétiques spécifiques à *N. gonorrhoeae* ou aux espèces commensales pour améliorer ces taux. Toutefois, nous avons démontré que la transformation de *N. meningitidis* par coculture avec une autre espèce comme *N. muscili* produit les mêmes résultats que lorsque de l'ADN pur est utilisé. Ces observations suggèrent fortement que la modulation de l'intégration d'ADN exogène via SsnA est un mécanisme partagé avec le gonocoque et les autres espèces commensales possédant la nucléase.

5.1.3 Nouveau mécanisme d'interférence de la compétence naturelle

Bien que la compétence naturelle contribue aux mécanismes de réparation du génome et à la plasticité génétique des *Neisseria*, l'import d'ADN exogène n'est pas toujours bénéfique aux bactéries. En effet, l'ADN viral ainsi que certains éléments génétiques mobiles peuvent être toxiques pour leur hôte, notamment quand ces derniers s'insèrent dans des gènes essentiels. Un article récent a par ailleurs démontré que l'ADN de *Neisseria* commensales était toxique pour les *Neisseria* pathogènes en raison de sa faible méthylation, soulignant les effets potentiellement néfastes de la compétence naturelle (Kim *et al.*, 2019). Pour minimiser ces inconvénients, les espèces naturellement compétentes ont développé plusieurs systèmes de protection contre les

transferts horizontaux de gènes (HGT) pouvant interférer avec la compétence naturelle (Liu *et al.*, 2020).

Les systèmes de restriction-modification (R-M) sont particulièrement abondants chez les espèces naturellement compétentes comme les *Neisseria* pathogènes, *Helicobacter pylori*, *Haemophilus influenzae* et *Streptococcus pneumoniae* (Vasu & Nagaraja, 2013). Ces systèmes modulent les HGT en dégradant l'ADN transformant possédant un profil de méthylation différent de l'hôte (Ambur *et al.*, 2012; Beauchamp *et al.*, 2017; Donahue *et al.*, 2000). Les systèmes R-M sont particulièrement importants à l'évolution de *N. meningitidis*, puisque leur distribution corrèle fortement avec la phylogénie de cette espèce (Budroni *et al.*, 2011). Les systèmes CRISPR-Cas, qualifiés comme « immunité adaptative » des bactéries, sont principalement utilisés par leurs hôtes pour dégrader l'ADN viral. Cependant, ces derniers peuvent également bloquer la transformation par compétence naturelle chez *S. pneumoniae* et *N. meningitidis* (Bikard *et al.*, 2012; Zhang *et al.*, 2013). De façon similaire, les protéines Argonautes agissent comme barrière de protection contre l'import d'ADN exogène chez la bactérie compétente *T. thermophilus*, mais ce système n'a encore été identifié chez les *Neisseria* (Swarts *et al.*, 2014).

Notre étude révèle un tout nouveau mécanisme d'interférence de la compétence naturelle. Combiné aux autres systèmes de protection contre les HGT, le système SsnA-NTS permettrait à ces espèces de contrecarrer les désavantages de leur compétence constitutive en promouvant l'intégrité et la stabilité génétique de certaines régions.

5.1.4 Antagonisme des séquences répétées chez les *Neisseria*

Nous avons démontré dans cette étude que SsnA limite l'intégration à son génome de gènes flanqués par des NTS. De manière frappante, la présence de ces éléments répétés à un effet diamétralement opposé sur les taux de transformation par rapport à la présence des *DNA uptake sequences* (DUS), qui favorisent l'import d'ADN (Berry *et al.*, 2013; Cehovin *et al.*, 2013). De plus, alors que les DUS sont retrouvés plus souvent près de gènes de maintenance très conservés, les NTS sont fortement associées à des gènes hautement variables qui encodent pour des protéines de surface (Davidsen *et al.*, 2004). En raison de cette dichotomie frappante entre la distribution de ces deux éléments répétés spécifiques aux *Neisseria*, ainsi que leurs rôles opposés dans la transformation, nous voyons ces séquences comme les déterminants génétiques d'un complexe mécanisme antagoniste régulant les HGT par compétence naturelle. Contrairement aux autres espèces compétentes comme *S. pneumoniae*, *V. cholerae* ou *B. subtilis*, les *Neisseria* expriment de façon constitutive leurs gènes de compétence, soulevant

plusieurs interrogations sur la façon dont elles régulent cette caractéristique (Seitz & Blokesch, 2013). Le mécanisme antagoniste des séquences répétées DUS et NTS pourrait bien être le chaînon manquant de la compétence constitutive des *Neisseria*, qui permettrait à la fois de promouvoir la plasticité génétique de certaines régions du génome tout en maintenant l'intégrité d'autres régions.

5.1.5 Paradigme des dRS3/NTS

L'accumulation des dRS3 autour de gènes codant pour des protéines de surface a soulevé plusieurs hypothèses par le passé quant à leurs fonctions biologiques (Parkhill *et al.*, 2000). Une étude de 2007 s'est attardée sur la distribution de ces séquences répétées chez différentes souches de méningocoque (Bentley *et al.*, 2007). Cette étude montre que les clusters de dRS3 ainsi que les gènes qui y sont retrouvés sont particulièrement variables par rapport au reste du génome (**Figure 56**). Pour expliquer cette particularité, les auteurs de cette étude ont posé l'hypothèse que les dRS3 seraient la cible d'une enzyme qui favoriserait la recombinaison avec de l'ADN exogène acquis par compétence naturelle. Dans la même optique, plusieurs autres études ont associé *in silico* les dRS3 à des zones précises de réarrangements chromosomiques caractérisées par des inversions, des insertions et des délétions de gènes (Budroni *et al.*, 2011; Schoen *et al.*, 2008). Globalement, toutes ces études bio-informatiques suggèrent que les dRS3 agissent comme médiateurs importants de la diversité génétique des *Neisseria*.

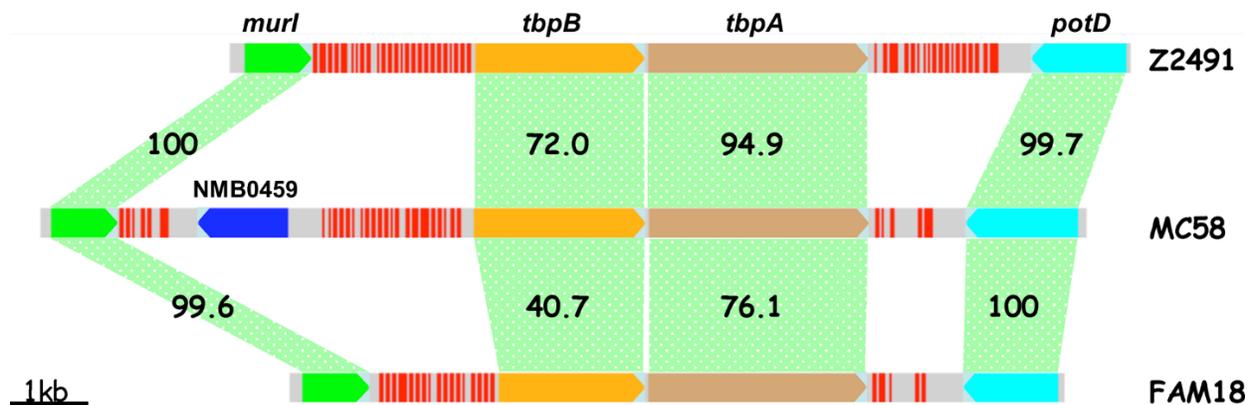


Figure 56 Hypervariabilité des régions contenant des dRS3 chez *N. meningitidis*

La dRS3 est représentée par des lignes rouges. Les zones vertes indiquent le pourcentage d'identité entre les séquences en acides aminés des orthologues. Adaptée de (Bentley *et al.*, 2007).

Notre étude amène enfin des preuves expérimentales de l'influence d'une sous-famille de dRS3 sur l'évolution et la plasticité génétique de *N. meningitidis*. Cependant, nos résultats divergent de toutes les précédentes hypothèses formulées à ce sujet. En effet, nous avons bel et bien identifié une enzyme qui interagit spécifiquement avec les NTS, des séquences répétées homologues aux dRS3. Cependant, plutôt que de promouvoir la recombinaison des régions contenant ces éléments répétés, la nucléase SsnA réduit considérablement leur taux d'intégration chromosomique via un mécanisme de restriction d'ADN simple-brin. L'interaction SsnA-NTS serait donc importante au maintien de la stabilité génétique de ces locus plutôt qu'à leur diversification.

Une question se pose alors : Comment et pourquoi les NTS s'accumulent dans des régions hypervariables ? Tout d'abord, il est important de noter que la majorité des gènes flanqués de nombreuses NTS possèdent des mécanismes de variabilité génétique qui ne dépendent pas de l'import d'ADN exogène par compétence naturelle. En effet, des gènes tels que *fetA*, *tdfH*, *pilC1-2*, *porA*, *lbpA-B*, *opaA-D*, *hpuA*, *vapA*, *nalP* et *opc* sont tous entourés de NTS et capables de variation de phase, tandis que *pilE*, *pilS1-5* et certains *opa* sont capables de variation antigénique (Rotman & Seifert, 2014). La variation de phase intervient lors de la réplication de l'ADN et est dépendante de très courtes répétitions retrouvées directement dans le gène ou son promoteur. Elle permet de moduler de façon réversible l'expression, le cadre de lecture ou la longueur des transcrits par un mécanisme appelé glissement de brin (Rotman & Seifert, 2014). Pour ces gènes, il pourrait être désavantageux de recombiner avec des séquences exogènes qui pourraient impacter leur capacité de variation de phase. L'accumulation de NTS à ces endroits permettrait donc peut-être de préserver cette caractéristique, qui est particulièrement importante dans l'évasion immunitaire de l'hôte et donc dans la virulence des *Neisseria* pathogènes. Contrairement à la variation de phase, la variation antigénique de *pilE* et des *opa* modifie directement leur séquence nucléotidique par recombinaison homologue avec d'autres locus similaires présents dans le génome. Ce mécanisme est dépendant d'une séquence particulière appelée G4 retrouvée en amont de *pilE* et ne nécessite pas d'import d'ADN par compétence naturelle (Palmer *et al.*, 2016). Encore là, il est possible que la présence de NTS autour de ce gène serve à prévenir les modifications par recombinaison avec de l'ADN exogène, afin de laisser place aux mécanismes de variation antigénique déjà présents. Comparées aux espèces pathogènes, les *Neisseria* commensales possèdent beaucoup moins de gènes variables (Palmer *et al.*, 2016). Parallèlement, le génome de ces espèces arbore également nettement moins de NTS, ce qui supporte l'hypothèse que ces séquences répétées puissent contribuer à la

« protection » des gènes variables en minimisant leur fréquence de recombinaison homologue avec de l'ADN exogène.

Nous avons tenté d'évaluer l'influence de SsnA sur la variation antigénique de *pilE* en analysant le nombre de variants chez des mutants de *N. meningitidis* en présence et en absence d'ADN transformant. Ces tests n'ont pas permis de détecter des différences significatives, ce qui n'exclut pas que SsnA puisse être impliquée dans la plasticité d'autres gènes, notamment par variabilité de phase. Cette hypothèse devra être explorée expérimentalement.

5.2 Importance de SsnA dans la virulence de *N. meningitidis*

Nous avons démontré dans cette étude que la délétion de *ssnA* chez des souches hyperinvasives de *N. meningitidis* diminue considérablement leur potentiel infectieux chez la souris. En effet, une infection systémique au méningocoque est éliminée plus rapidement et affiche un taux de mortalité nettement plus bas lorsque la souche utilisée n'exprime pas la nucléase. Ces résultats sont cohérents avec l'observation que les *Neisseria* pathogènes possèdent un nombre beaucoup plus élevé de NTS dans leur génome que les espèces commensales. L'interaction SsnA-NTS joue donc un rôle majeur dans la virulence, bien que le mécanisme sous-jacent ne soit pas encore connu.

La compétence naturelle est généralement considérée comme bénéfique dans un contexte infectieux. Ce mécanisme a justement été découvert il y a près d'un siècle en observant une souche inoffensive de *S. pneumoniae* devenir mortelle simplement en l'incubant avec un lysat de souche virulente. Chez cette espèce, la virulence est étroitement reliée à la compétence puisque de nombreux gènes spécifiquement induits lors du stade compétent sont impliqués dans la survie et la prolifération lors d'infection (Zhu *et al.*, 2015). De plus, chez *Acinetobacter baumannii*, la compétence naturelle est fortement augmentée en présence d'albumine, une composante majeure du sérum (Traglia *et al.*, 2016). Une étude très récente a même démontré que l'induction de la compétence chez *Staphylococcus aureus* est essentielle à l'établissement et au maintien d'une infection (Cordero *et al.*, 2022). Quant aux *Neisseria*, la délétion d'une protéine essentielle à leur compétence, *dprA*, est associée à une augmentation du taux de variation antigénique des pili, et donc indirectement à leur potentiel infectieux (Duffin & Barber, 2016). Cette étude ainsi que plusieurs autres soulèvent l'importance des gènes de compétence dans d'autres fonctions que la transformation, dont le métabolisme et la variabilité génétique (Cordero *et al.*, 2022).

Nous nous sommes penchés sur la façon dont l'interaction SsnA-NTS pouvait influencer la virulence de *N. meningitidis*. Pendant les premières heures suivant l'infection dans nos modèles murins, aucune différence de charge bactérienne entre des souches exprimant et n'exprimant pas *ssnA* n'est observée. Ceci suggère que SsnA n'est pas impliquée dans l'établissement de l'infection, mais plutôt dans sa survie à plus long terme chez l'hôte. Cette nucléase semble donc fournir à *N. meningitidis* un meilleur potentiel d'adaptation et d'évasion immunitaire, sans affecter son potentiel d'adhésion, d'invasion ou de prolifération. Lorsque nous avons infecté *in vitro* des lignées cellulaires, la présence ou l'absence de SsnA n'a pas eu d'impact significatif sur l'adhésion et l'invasion, ce qui supporte l'hypothèse d'un rôle plus indirect dans la virulence nécessitant une interaction avec le système immunitaire. Ces résultats sont également cohérents avec le fait que SsnA soit cytoplasmique, sans aucune indication d'expression à la surface ou de sécrétion. Dans les conditions testées, SsnA n'est pas impliquée dans la variation antigénique des pili, mais nous ne connaissons toujours pas son rôle potentiel dans la variation de phase d'autres protéines de surface. Plusieurs expériences sont proposées dans une prochaine section de cet ouvrage dans le but d'élucider le rôle de SsnA et des NTS dans la virulence du méningocoque.

5.3 Nouvelle famille de protéines GIY-YIG

Nous avons démontré au travers de cette étude que la protéine SsnA de *N. meningitidis* possède plusieurs milliers d'homologues répartis dans l'entièreté du domaine bactérien. Ces protéines sont identiques à plus de 50% et sont composées d'une centaine d'acides aminés formant un unique domaine fonctionnel GIY-YIG. À notre connaissance, aucune de ces protéines n'a été caractérisée auparavant. Elles sont majoritairement annotées comme nucléases hypothétiques en raison de leur petite taille et de leur homologie partielle avec d'autres protéines GIY-YIG multidomaines. Puisque nous avons été en mesure de démontrer l'activité catalytique unique et les fonctions biologiques importantes de certaines de ces protéines chez les *Neisseria*, de même que l'activité ssDNase d'autres homologues plus éloignés, nous proposons de réunir ces enzymes sous le nom Ssn, pour *Specific Single-strand nucleases*.

La première observation que l'on peut faire quand on regarde la distribution phylogénétique des Ssn est sa présence quasi exclusive chez les bactéries à Gram négatif, à l'exception de quelques actinobactéries et firmicutes. Étant donnée sa localisation prédite comme étant cytoplasmique, nous ne sommes pas en mesure d'expliquer ce phénomène. Il est possible que ces protéines

interagissent avec d'autres facteurs prédominants chez les bactéries à Gram négatif comme les bêta-lactamases, les récepteurs TonB-dépendants ou encore les systèmes de sécrétion de type III. Plusieurs espèces naturellement compétentes possèdent un homologue de Ssn, comme *Legionella pneumophila*, *Vibrio vulnificus*, *Agrobacterium tumefaciens*, et *Bradyrhizobium japonicum*. Cependant, la grande majorité des homologues sont retrouvés dans des espèces non compétentes. De façon similaire, ces homologues sont présents dans plusieurs pathogènes humains d'importance tels que *Coxiella burnetti* et *Rickettsia felis*, mais aussi dans une panoplie d'espèces commensales ou environnementales. Cette diversité suggère que les protéines Ssn ont d'autres fonctions biologiques que celles révélées dans cette étude.

La famille des Ssn se distingue de toutes les familles de protéines GIY-YIG précédemment définies (**Figure 57**). En effet, les Ssn sont particulièrement petites et ne possèdent que le domaine GIY-YIG qui couvre la quasi-totalité de leur séquence. De plus, ces protéines semblent agir sous forme de monomères. En effet, beaucoup de Ssn ne peuvent pas former de ponts disulfures dû à une absence de cystéines, et aucun oligomère n'a été détecté par MALDI-TOF ni lors des EMSA pour les protéines de *N. meningitidis* et de *N. elongata*. À l'inverse, les autres familles GIY-YIG sont composées de plusieurs domaines fonctionnels, ou alors affichent de larges extensions autour du domaine GIY-YIG. La majorité de ces dernières est active sous forme multimérique (Dunin-Horkawicz *et al.*, 2006). D'un point de vue fonctionnel, l'activité nucléase particulière des Ssn ainsi que leurs fonctions biologiques chez les *Neisseria* pathogènes les classent également à part. Aucune autre famille de GIY-YIG n'est spécifique à l'ADN simple-brin. Ces dernières dépendent également de domaines de liaisons à l'ADN distincts pour reconnaître leur substrat.

D'un point de vue phylogénétique, les Ssn forment un groupe monophylétique lorsqu'alignées avec des membres des autres familles GIY-YIG. Pour toutes ces raisons, il semble clair que les Ssn peuvent être considérées comme une famille distincte de protéines GIY-YIG, ce qui ouvre la porte à de nombreuses autres études. D'après nos analyses, cette famille peut ensuite être subdivisée en trois sous-familles : SsnA, SsnB et SsnC. Des études supplémentaires seront nécessaires pour déterminer si les sous-familles B et C possèdent une activité nucléase simple-brin similaire aux SsnA. Malheureusement, nos multiples tentatives de purification et de caractérisation des groupes B et C n'ont pas porté fruit, principalement en raison d'une très faible expression et d'une encore plus faible solubilité (données non montrées).

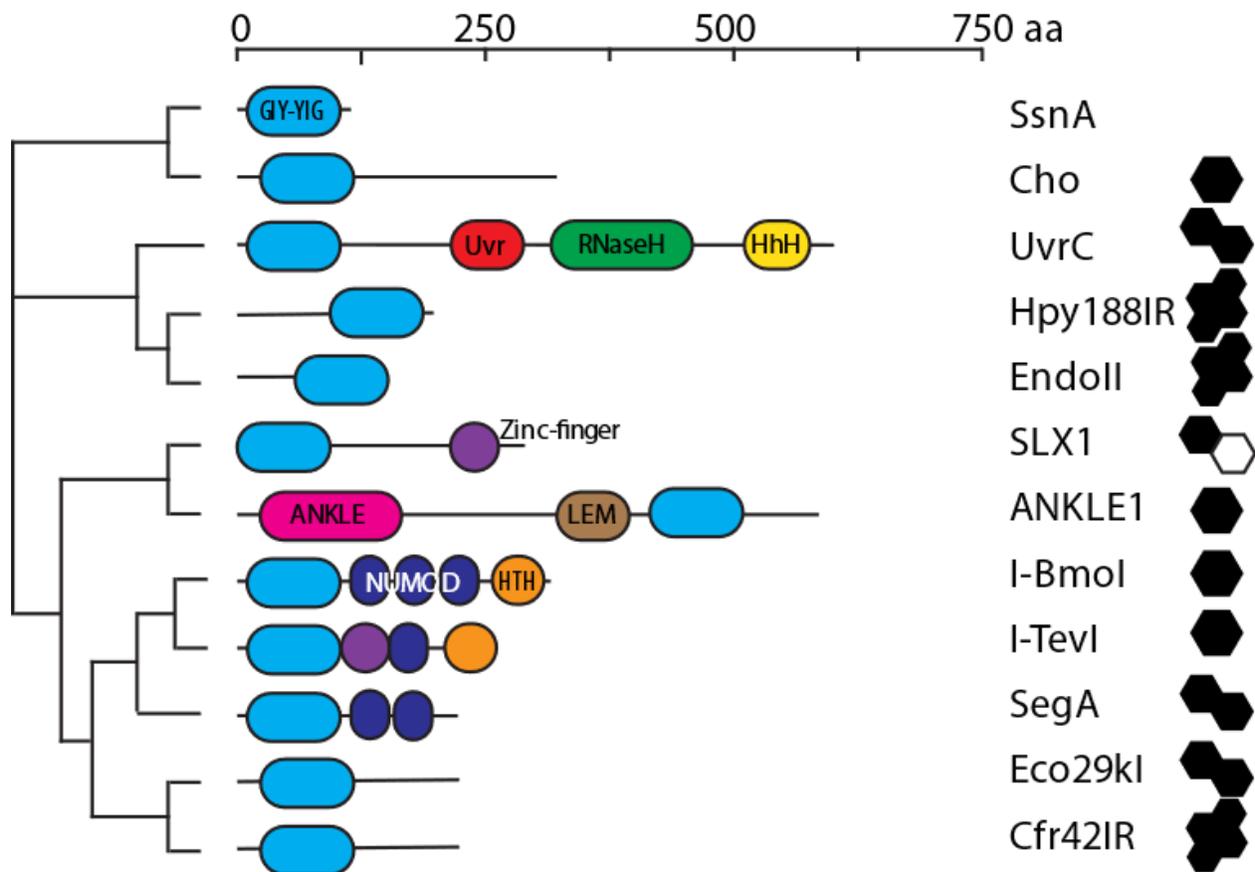


Figure 57 Nouvelle addition à la superfamille des nucléases GIY-YIG

Cladogramme, architecture fonctionnelle, taille et état oligomérique de membres représentatifs des familles caractérisées de GIY-YIG. Cladogramme obtenu à partir d'alignements multiples réalisés avec Clustal Omega.

5.3.1 Activité enzymatique unique de SsnA

Notre étude, focalisée sur le SsnA de *N. meningitidis*, a révélé que cette enzyme est une endonucléase spécifique d'ADN simple-brin (ssDNase). Cette dernière lie la structure en épingle à cheveux formée par la séquence répétée NTS et clive quelques nucléotides en amont de celle-ci. Une telle spécificité est tout à fait unique et remarquable considérant la petite taille de SsnA, son unique domaine fonctionnel, et son indépendance vis-à-vis d'autres protéines. En effet, la majorité des nucléases agissant sur l'ADN simple-brin peuvent également cliver l'ADN double-brin ou l'ARN, ou encore dégradent l'ADN simple-brin de façon non spécifique (Desai & Shankar, 2003).

À notre connaissance, la seule autre famille de nucléases capable de lier et de cliver de façon séquence spécifique de l'ADN simple-brin est la famille des HUH (Chandler *et al.*, 2013). Ces dernières regroupent des transposases et des relaxases importantes dans la mobilité d'éléments

généétiques et dans la conjugaison. Leur domaine HUH est généralement associé à d'autres domaines fonctionnels permettant la liaison à l'ADNsb ou la génération d'ADNsb à partir d'ADNdb (domaine hélicase). Parmi les nucléases HUH, on retrouve la transposase TnpA qui affiche des similarités frappantes avec SsnA. En effet, TnpA est également capable de lier spécifiquement de l'ADNsb palindromique en épingle à cheveux et de couper à proximité, en plus d'afficher une spécificité de brin (Ton-Hoang *et al.*, 2005). Cette enzyme, composée uniquement du domaine HUH, est la plus petite transposase caractérisée, mesurant à peine 150 acides aminés. Pour reconnaître son site de clivage, TnpA n'utilise pas de domaine de liaison à l'ADN, mais exploite plutôt l'appariement de quelques nucléotides complémentaires à proximité du palindrome (**Figure 58**). Afin de médier sa transposition, cette enzyme se dimérise pour cliver simultanément les deux répétitions inverses entourant son gène (Guynet *et al.*, 2009).

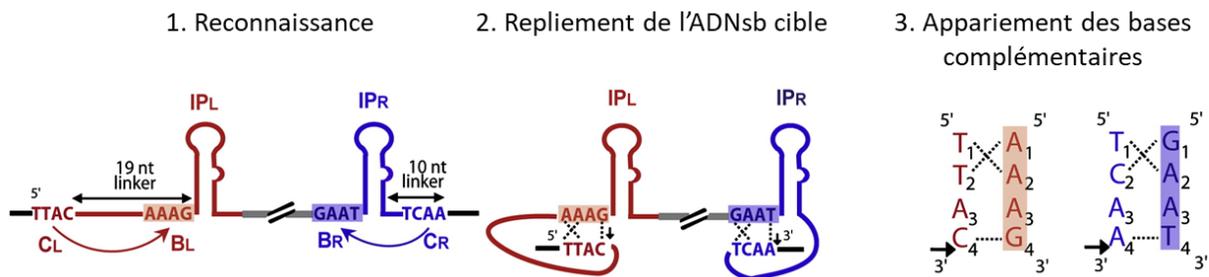


Figure 58 Mécanisme de repliement du site de clivage de TnpA

TnpA est une transposase d'ADN simple-brin clivant spécifiquement à proximité des structures en épingle à cheveux flanquant son propre gène (IP_L et IP_R). Ses sites de clivages (C_L et C_R) ne sont pas reconnus directement par la protéine, mais plutôt par des boîtes de tétranucléotides (B_L et B_R) situées à proximité. Ces boîtes s'apparient aux sites de clivage, permettant l'activité catalytique de TnpA. Figure adaptée de (Guynet *et al.*, 2009) avec autorisation de l'éditeur (Licence Elsevier #5495440575292 obtenue le 24 février 2023).

Au vu de ces ressemblances marquantes entre l'activité catalytique de TnpA et SsnA, bien que leur domaine fonctionnel et leur structure prédite soient différents, on pourrait envisager un mécanisme similaire permettant la spécificité de clivage de SsnA. De façon intrigante, nos modélisations *in silico* de l'interaction SsnA-NTS suggèrent que la région en amont de la NTS se replie sur elle-même pour prolonger la structure secondaire en tige-boucle générée par le palindrome (**Annexe I**). Un tel modèle de repliement semble compatible avec le mécanisme d'action de TnpA. Des études structurales approfondies en présence de différents substrats d'ADNsb seront nécessaires pour vérifier cette hypothèse. Ces dernières ont en effet été absolument essentielles à la compréhension de la spécificité et du mécanisme d'action de TnpA (Barabas *et al.*, 2008).

5.3.2 Spécificités attendues

Au tout début de cette étude, nous nous sommes intéressés au gène codant pour SsnA en raison de sa forte association avec la séquence répétée NTS dans plusieurs espèces de *Neisseria*. Cette séquence palindromique, qui forme une structure secondaire en épingle à cheveux, s'est révélée être le site de reconnaissance de SsnA qui clive à proximité. Nous nous sommes alors demandé si de telles séquences répétées étaient également retrouvées à proximité des milliers d'homologues de *ssnA* identifiés dans l'ensemble du domaine bactérien. Effectivement, plusieurs espèces très éloignées telles que *Chlorobaculum parvum* possèdent des séquences répétées identiques aux dRS3 ou aux NTS à proximité immédiate de leur *ssnA*. D'autres homologues encore, comme ceux de *Rhizorhabdus (sphingomonas) wittichii*, *Nocardia huaxiensis* et *Coxiella burnetti*, sont entourés de séquences répétées qui ressemblent aux NTS, mais qui affichent une ou plusieurs substitutions, insertions ou délétions très conservées. Enfin, certains homologues comme le *ssnA* de *Neisseria wadsworthii* et de *Caulobacter henricii*, sont également flanqués d'éléments répétés palindromiques formant probablement des structures en épingle à cheveux, mais dont la séquence diverge davantage des NTS des *Neisseria*.

Nous avons exploré l'hypothèse que chaque nucléase SsnA puisse être capable de cliver à proximité de la séquence répétée qui lui est associée. Cette hypothèse a d'abord été supportée par l'étude de l'homologue provenant de l'espèce commensale *Neisseria wadsworthii*, dont le génome ne contient aucune NTS mais plus d'une centaine de variants conservés surnommés NTS_{var}, dont plusieurs à proximité immédiate du gène *ssnA*. En effet, la nucléase de *N. wadsworthii* clive sous forme d'ADN simple-brin la NTS_{var}, mais pas la NTS, confirmant la coévolution entre les nucléases de cette famille et leurs éléments répétés. Supportant davantage ces résultats, l'homologue de *Dehalogenimonas lykanthroporepellens*, une espèce éloignée des *Neisseria*, est également capable de cliver sa propre séquence répétée, une autre variante conservée des NTS. De toute évidence, la grande famille des SsnA abrite une multitude de spécificités différentes dont la caractérisation précise nécessitera davantage d'efforts.

5.3.3 Nucléases, ou plutôt transposases domestiquées ?

Les régions codant pour les différents homologues de SsnA ressemblent de manière frappante à des séquences d'insertion (IS), une famille d'éléments génétiques mobiles comprenant un gène de transposase flanqué de répétitions inverses (**Figure 59**). Une famille en particulier, celle des IS200/IS605, possède une structure quasi identique aux locus contenant les gènes *ssnA*. Cette dernière se distingue des autres IS par sa petite taille et l'absence de répétitions directes aux extrémités, substituées par des séquences palindromiques inversées formant des structures en épingle à cheveux. Contrairement aux autres IS, les IS200/IS605 sont transposées par un mécanisme de « peler et coller » (*peel and paste*) uniquement sous forme d'ADN simple-brin (Barabas *et al.*, 2008). Leurs transposases sont effectivement des endonucléases spécifiques d'ADNsb qui reconnaissent la structure en épingle à cheveux de leurs propres répétitions inverses, exactement comme la nucléase SsnA avec les NTS.

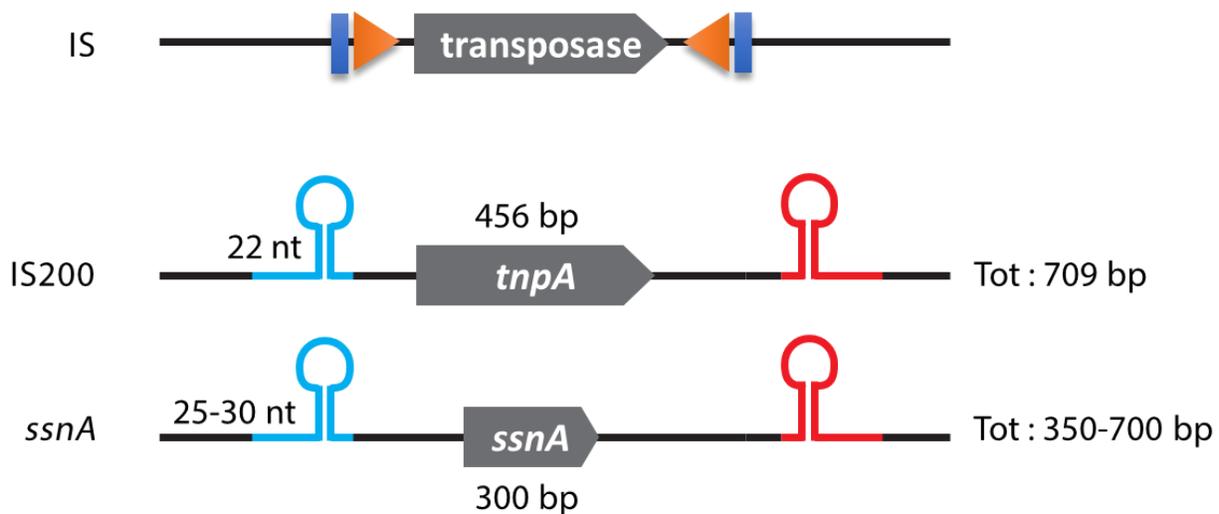


Figure 59 Structure des séquences d'insertion et des locus *ssnA*

Les flèches orange illustrent les répétitions inversées qui flanquent la transposases, tandis que les rectangles bleus représentent les courtes répétitions directes aux extrémités des IS.

Les similarités entre les IS200/IS605 et les locus encodant SsnA ne s'arrêtent pas là. Dans les dix dernières années, une famille entière de nucléases simple-brin homologues aux transposases IS200/IS605 a été associée à des séquences répétées de la famille des *Repetitive Extragenic Palindromes* (REP), qui forment des structures en épingle à cheveux (Ton-Hoang *et al.*, 2012). Ces nucléases, dénommées RAYT (*REP-associated tyrosine transposase*) ou TnpA_{REP}, sont toujours flanquées de deux REP inversés et leur présence dans un génome est fortement corrélée avec l'abondance de REP (Nunvar *et al.*, 2010). Les différentes RAYT sont associées à

des REP distincts, prouvant la coévolution de ces éléments (Corneloup *et al.*, 2021). Contrairement à la majorité des transposases, ces nucléases sont présentes en une seule copie par génome (Bertels & Rainey, 2011). Incapables de transposition, mais capables de cliver et de recombiner des REP sous forme simple-brin, les RAYT sont décrites comme le premier exemple de « domestication » de transposases chez les eucaryotes (Ton-Hoang *et al.*, 2012). Ces dernières seraient issues des transposases IS200/IS605 qui auraient perdu leur activité de transposition au profit d'autres fonctions biologiques bénéfiques à l'hôte, d'où le terme domestication (Bertels *et al.*, 2017).

Il est difficile d'ignorer les similarités entre les RAYT-REP et les SsnA-NTS. En effet, les NTS dérivent des dRS3, qui se classent dans la famille des REP. Les gènes codant pour SsnA sont presque systématiquement flanqués de copies inversées de NTS, et une forte corrélation est observée entre l'abondance de NTS dans un génome et la présence du gène *ssnA*. De plus, SsnA clive spécifiquement les NTS sous forme simple-brin. Enfin, différents homologues de SsnA sont associés à différents variants de la NTS. Aux vues de ces ressemblances, il est tout à fait envisageable que SsnA, tout comme les RAYT, provienne d'une transposase ancestrale qui aurait été domestiquée par les *Neisseria* et d'autres bactéries pour accomplir d'autres fonctions. SsnA serait donc impliquée dans la propagation et la diversité des NTS par un mécanisme similaire à celui des RAYT/REP. Un autre exemple de domestication de transposase de cette même famille d'IS200/IS605 serait l'émergence des nucléases Cas9 à partir de TnpB, aujourd'hui impliquées dans la protection du génome de l'hôte (Jangam *et al.*, 2017; Shmakov *et al.*, 2017).

Il est important de noter que les NTS ont une séquence complètement différente des répétitions associées aux RAYT, et que SsnA n'a aucune homologie avec ces protéines. Notre étude pointerait donc potentiellement vers une toute nouvelle famille de transposases ancestrales domestiquées. Depuis la démocratisation du séquençage, les transposases sont en grande majorité identifiées grâce à leur similarité aux autres IS des banques de données, et non par leur activité de transposition. Il est donc attendu que plusieurs familles non homologues à celles déjà caractérisées restent à identifier, comme peut-être la famille Ssn (Siguier *et al.*, 2014). On peut noter ici que quelques rares souches de *Neisseria* possèdent deux copies du gène *ssnA*, ce qui laisserait supposer une activité transposase. Cependant, ces dernières sont associées à la duplication ancestrale du gène *pilC* en *pilC1* et *pilC2*, qui semble être indépendant de la présence de *ssnA*.

5.4 Perspectives futures

5.4.1 Interactome de SsnA

Notre étude suggère un rôle de SsnA dans le traitement et la recombinaison d'ADN transformant au génome de l'hôte. *In vitro*, cette nucléase est suffisante pour catalyser le clivage spécifique d'ADN simple-brin à proximité des séquences répétées NTS. Chez *N. meningitidis*, la délétion ou la surexpression du gène *ssnA* suffisent à moduler le taux de transformation d'ADN contenant des NTS, mais ce gène n'est pas essentiel à l'import d'ADN ni à sa recombinaison. Les mécanismes de compétence naturelle et de recombinaison homologue sont gouvernés par l'interaction de dizaines de protéines accomplissant chacune des fonctions spécifiques. Il est donc attendu que SsnA s'imbrique dans ces mécanismes en interagissant avec d'autres effecteurs, que ce soit pour être recruté à l'ADN simple-brin, pour médier sa spécificité, ou bien pour intervenir dans la virulence de *N. meningitidis*. Nous n'avons pour l'instant pas exploré ces potentielles interactions protéiques, mais cette avenue est une suite logique dans la caractérisation de la famille SsnA.

In vitro, l'interactome de SsnA pourrait être révélé de façon non ciblée par *Tandem affinity purification-mass spectroscopy* (TAP-MS) ou par la technique plus conventionnelle de co-immunoprécipitation. Le TAP-MS consiste à ajouter des tags au gène d'intérêt directement dans le chromosome, puis à purifier le complexe natif de la protéine et à en distinguer les composantes par SDS-PAGE et spectrométrie de masse. Cette technique permet d'identifier directement dans l'hôte des faibles interactions protéiques, en plus de conserver la conformation oligomérique de la protéine d'intérêt (Völkel *et al.*, 2010). La co-immunoprécipitation de lysat de *N. meningitidis* sur une résine saturée en SsnA permettrait également de détecter des interactions protéine-protéine de façon plus conventionnelle.

In vivo, de façon plus ciblée, les potentiels interacteurs de SsnA pourraient être identifiés par *Bacterial two-hybrid assay* (B2H), dérivée de la technique du double hybride chez la levure (Y2H). SsnA servirait d'appât, et les effecteurs connus du traitement d'ADNsb et de la recombinaison homologue, comme RecA, DprA et SsbA, serviraient de proies. Ces interactions pourraient être confirmées par des techniques de microscopie comme le transfert d'énergie entre molécules fluorescentes (FRET). Puisque la détection de colocalisation de fluorescence est techniquement ardue chez les cocci, le bacille *N. elongata* qui possède un SsnA fonctionnel pourrait être utilisé. Une telle technique a été utilisée avec succès pour démontrer l'interaction de gènes de compétence chez *B. subtilis* (Kramer *et al.*, 2007).

5.4.2 Études structurales de SsnA

La cristallographie de SsnA en présence de différents substrats d'ADN simple-brin risque d'être absolument essentielle à la compréhension de la spécificité de liaison et de clivage de cette enzyme ainsi que de son mécanisme d'action. En effet, à l'heure actuelle, son unique domaine fonctionnel nucléase (GIY-YIG) n'explique pas une telle spécificité, et aucun homologue proche n'a encore été étudié. Nous avons généré plusieurs modèles de liaison et de clivage *in silico* qui supportent nos résultats expérimentaux (**Annexe I**). Ces derniers confirment la structure en épingle à cheveux des NTS, ainsi que l'attachement de SsnA au pied de cette structure et son interaction avec plusieurs nucléotides essentiels à son activité enzymatique. Cependant, une telle modélisation a ses limites et n'est pas toujours représentative de la réalité. En effet, nos modèles ont du mal à expliquer comment SsnA peut couper à une dizaine de nucléotides de son site de liaison sans une reconfiguration majeure de la protéine, peu réaliste en termes de thermodynamique. À l'avenir, il faudra donc des études structurales expérimentales comme la cristallographie à rayons X pour résoudre ce mystère. Pour les nucléases d'ADN simple-brin de la famille des IS200/IS605, qui partagent beaucoup de similarités avec SsnA, de telles études structurales ont été déterminantes dans la compréhension de leur activité catalytique (Barabas *et al.*, 2008).

5.4.3 Importance de SsnA dans la virulence du méningocoque

Bien que nous ayons révélé SsnA comme étant un déterminant important de la virulence de *N. meningitidis*, nous n'avons pas encore été en mesure d'élucider son rôle concret dans la pathogénicité. En effet, les quelques tests d'adhésion et d'invasion que nous avons tentés *in vitro* n'ont pas révélé de phénotype flagrant, suggérant que SsnA est plutôt impliquée dans l'adaptation et la survie chez l'hôte que dans sa prolifération et son potentiel d'invasion. Plusieurs approches permettraient de mieux comprendre la relation entre cette nucléase et la pathogénicité du méningocoque.

Tests de résistance au sérum

Fréquemment utilisés dans des laboratoires à vocation plus clinique, des tests de résistance au sérum humain permettent de vérifier le potentiel bactéricide d'anticorps induits par vaccination. Nous pourrions réaliser des tests similaires en comparant la survie de différentes souches de *N. meningitidis* exprimant ou pas SsnA. Différentes combinaisons de sérum d'individus vaccinés ou

non-vaccinés avec le complément actif ou inactivé pourraient être utilisées. Ces tests *in vitro* donneraient des indications de l'influence de SsnA sur le potentiel d'évasion immunitaire de la bactérie.

Tests de compétitivité

Nous avons démontré qu'une infection avec une souche n'exprimant pas SsnA est éliminée plus rapidement et engendre moins de mortalité qu'une souche exprimant la nucléase. Cependant, ceci n'est qu'une preuve indirecte de l'avantage adaptatif que procure SsnA. Pour confirmer que cette protéine est bien bénéfique à la bactérie dans un contexte infectieux, nous pourrions réaliser des tests de compétitivité chez la souris. En coinfectant deux souches exprimant et n'exprimant pas *ssnA*, chacune associée à un marqueur fluorescent différent, nous pourrions suivre la dissémination et la survie de ces souches en temps réel sans grosses modifications de notre modèle murin déjà optimisé. Ces expériences pourraient ensuite être comparées à des tests de co-incubation *in vitro* des mêmes souches, ce qui permettrait de vérifier l'avantage sélectif de SsnA en contexte infectieux et non infectieux.

Tests de portage

Notre laboratoire a récemment mis au point un modèle de portage asymptomatique de *Neisseria meningitidis* chez la souris. Afin de voir l'influence de SsnA dans un contexte d'infection non invasive, ce modèle pourrait être utilisé. Nous avons déjà réalisé un test de ce genre avec des souches de *N. meningitidis* mutées pour SsnA (données non montrées). Cependant, ce dernier comptait trop peu de souris et n'a pas été reproduit. On ne peut donc pas encore en tirer des conclusions, mais cela a au moins pu démontrer qu'une telle approche était faisable et mérite d'être explorée davantage. *Neisseria musculli*, une espèce commensale possédant également le gène *ssnA*, pourrait aussi être utilisée pour étudier l'importance de cette nucléase dans un portage à plus long terme.

Évolution *in vivo* en contexte infectieux

Une hypothèse de longue date concernant les séquences répétées dRS3 et NTS suggère que ces éléments promeuvent la virulence des *Neisseria* pathogènes en favorisant la diversité génétique de plusieurs protéines de surface via la recombinaison avec de l'ADN exogène (Bentley *et al.*, 2007). Nous savons que SsnA clive l'ADN simple-brin à proximité des NTS *in vitro*, mais nous ne connaissons pas l'influence de cette activité sur la stabilité génétique de *N. meningitidis* en contexte infectieux. On pourrait envisager à partir de nos modèles murins déjà optimisés de prélever les bactéries à différents temps post-infection et de les séquencer, puis de

comparer les régions riches en NTS et les gènes de virulence à ceux des souches de départ. Bien que complexes, de telles analyses permettraient de quantifier la variabilité de ces régions en présence et en absence de SsnA, notamment en comparant le nombre de copies de NTS.

Variation de phase

Nos analyses *in vitro* de la variabilité antigénique de Pile n'ont pas révélé de rôle de SsnA dans ce mécanisme important à la virulence du méningocoque via l'évasion immunitaire. Cependant, nous n'avons pas du tout testé la variation de phase des dizaines de gènes codant pour des protéines de surface qui sont également entourés de nombreuses copies de NTS. Grâce à notre expertise et à nos équipements de séquençage, nous pourrions développer une approche similaire pour quantifier la variation de tous ces gènes. Cette approche pourrait être couplée aux infections précédemment décrites pour étudier ce mécanisme en contexte infectieux ou non.

5.4.4 Potentiel biotechnologique des Ssn

Tant que la spécificité exacte des différents membres de la grande famille des Ssn ne sera pas connue en détail, il est difficile de se prononcer sur le potentiel biotechnologique des Ssn. Cependant, notre étude suggère une activité catalytique et une spécificité très inhabituelle ciblant l'ADN simple-brin qui démarque ces enzymes des autres familles de nucléases. Une fois mieux caractérisées, les Ssn pourraient être utilisées comme outils moléculaires dans plusieurs applications biologiques et biotechnologiques. En effet, le manque d'outils modifiant spécifiquement l'ADNsb ralentit considérablement la recherche sur plusieurs mécanismes biologiques essentiels comme la réplication et la réparation d'ADN. Nous avons d'ailleurs déposé un brevet protégeant l'usage de ces nucléases *in vitro* et *in vivo* (**Annexe II**).

L'activité nucléase simple-brin hautement spécifique de SsnA pourrait être exploitée en couplant cette protéine à divers domaines de liaisons d'ADN. En effet, SsnA coupe une dizaine de nucléotides en amont de son site de reconnaissance, ce qui signifie qu'on pourrait potentiellement cliver n'importe quelle séquence d'intérêt en modifiant sa spécificité de liaison. Bien que de nombreuses nucléases chimériques aient été développées pour cliver l'ADN double-brin, comme les CRISPR-Cas, les TALENs et les ZFNs, il n'existe pratiquement aucun outil pour manipuler l'ADN simple-brin (Spåhr *et al.*, 2018). Plusieurs domaines protéiques liant l'ADNsb de façon séquence-spécifique pourraient être envisagés, dont certains domaines *oligonucleotide oligosaccharide-binding* (OB), des domaines *K homology* (KH), et des domaines *whirly* (Dickey *et al.*, 2013). Cependant, ces derniers ont des spécificités définies peu ou pas modulables. Récemment, des chercheurs ont généré les toutes premières protéines programmables liant

l'ADNsb avec une spécificité artificielle, utilisant des répétitions pentricopeptides (PRR) (Spåhr *et al.*, 2018). Ces dernières représentent une avenue prometteuse pour générer des nucléases d'ADNsb programmables à partir de SsnA.

Notre étude suggère que SsnA ne clive pas l'ADN double-brin simplement car elle est incapable de s'y lier. Pourtant, des expériences préliminaires avec cette nucléase ont montré une faible activité nickase capable de cliver un seul brin sur de l'ADNdb (Chenal, 2019). Concrètement, cela suggère que SsnA pourrait potentiellement être couplée à des domaines de liaison d'ADNdb afin de servir comme nickase programmable. Plusieurs nickases programmables ont déjà été développées et sont très prometteuses pour le domaine de l'ingénierie génétique. Contrairement aux systèmes CRISPR-Cas9 couramment utilisés, ces nickases génèrent des mutations « propres », sans aucune insertion aléatoire due à la machinerie de réparation de l'hôte (Kim *et al.*, 2012; Kim & Kim, 2014).

CONCLUSION

Ce projet de recherche entier découle de l'observation intrigante qu'un tout petit gène de *N. meningitidis* encodant pour une nucléase hypothétique (SsnA) soit entouré d'une multitude de copies d'une séquence répétée particulière. Cet élément répété ressemble fortement aux dRS3, des répétitions cryptiques retrouvées majoritairement chez les *Neisseria* et dont les fonctions étaient jusque là méconnues. À l'aide d'analyses *in silico*, nous avons révélé que les dRS3 sont plus longues qu'initialement décrites et possèdent une polarité définie par quelques nucléotides non palindromiques hautement conservés en amont et en aval. Ces répétitions, que l'on a renommées NTS pour *Neisseria transformation sequences*, sont distribuées de façon particulière dans le génome des *Neisseria*. Leur fréquence dans une espèce donnée corrèle fortement avec la phylogénie des *Neisseria*, suggérant leur acquisition dans un ancêtre commun et leur coévolution avec leur hôte. Les NTS sont nettement plus fréquentes chez les *Neisseria* du clade prépathogène et tout particulièrement chez *N. meningitidis*, dans laquelle elles s'accumulent de façon marquante autour de gènes codant pour des protéines de surface impliquées dans la virulence. Nos analyses révèlent également la présence d'une variante de la NTS dans un groupe d'espèces monophylétique des *Neisseria*, supportant davantage leur importance dans l'évolution de ce genre bactérien.

À l'aide de tests biochimiques, nous avons démontré que SsnA interagit physiquement avec les NTS en s'y liant de façon spécifique, et en clivant quelques nucléotides en amont, mais uniquement sous forme d'ADN simple-brin, d'où le nom *Specific single-strand nuclease* (SsnA). Ces expériences, couplées à des modélisations *in silico* de l'interaction entre SsnA et la NTS, suggèrent que cette petite nucléase se fixe au pied de la tige de la structure en épingle à cheveux formée par la NTS sous forme simple-brin.

À l'aide de tests phénotypiques, nous avons confirmé cette interaction *in vivo* chez *N. meningitidis* dans le contexte de transformation d'ADN exogène via la compétence naturelle. Nos résultats révèlent SsnA comme un médiateur important de l'intégration d'ADN transformant par un mécanisme de restriction d'ADN simple-brin dans le cytoplasme, avant l'étape de recombinaison homologue au chromosome de l'hôte. En clivant à proximité des séquences répétées, SsnA réduit considérablement le taux d'intégration de gènes flanqués par des NTS en réduisant la longueur d'homologie avec le génome. Bien que cette hypothèse reste à prouver, nous suggérons que cet effet favorise indirectement la variation de phase de nombreux gènes, soit un mécanisme de diversité génétique indépendant de la transformation et de la recombinaison.

Grâce à un modèle murin d'infection systémique au méningocoque optimisé dans notre laboratoire, nous avons démontré *in vivo* que SsnA est un déterminant majeur de la virulence de *N. meningitidis*. N'agissant ni sur l'adhésion ni sur l'invasion des cellules hôtes, cette nucléase favorise la persistance de l'infection, possiblement en promouvant l'évasion immunitaire. Enfin, nous avons identifié plusieurs milliers de protéines homologues à SsnA dans de nombreux phyla bactériens. Ces homologues se distinguent des autres familles de protéines GIY-YIG d'un point de vue phylogénétique, structural et fonctionnel, soulevant la nécessité de les classer dans une nouvelle famille d'enzymes : les Ssn. Les fonctions de ces homologues restent à identifier, mais il est attendu que ces enzymes clivent spécifiquement l'ADN simple-brin à proximité de séquences répétées entourant leurs gènes. En effet, la majorité des SsnA sont encodées par des gènes flanqués d'éléments répétés palindromiques de la famille des REP, formant des îlots génétiques semblables à des séquences d'insertion (IS). Une telle activité a déjà été confirmée dans cette étude pour les homologues provenant de *Neisseria wadsworthii* et de *Dehalogenimonas lykanthroporeppelens*, suggérant que de nombreuses autres spécificités restent à découvrir.

Impact de cette étude

Les *Neisseria* sont reconnues pour contenir des espèces dont la compétence naturelle est active à haut niveau en tout temps, sans aucune régulation d'expression des gènes impliqués, soulevant de nombreuses questions sur la façon dont ces espèces se protègent des potentiels effets délétères de la transformation. Cette étude révèle que *N. meningitidis* possède un tout nouveau mécanisme de médiation de la compétence naturelle par antagonisme de deux séquences répétées : la DUS favorisant l'import d'ADN, et la NTS inhibant l'intégration d'ADN exogène au chromosome. Cette régulation, médiée par la présence et la disposition de ces éléments répétés dans l'ADN transformant, permettrait de protéger certaines régions génomiques de recombinaisons excessives avec de l'ADN exogène, tout particulièrement en absence de facteurs de sélection naturelle.

Toujours chez les *Neisseria*, notre étude lève enfin le voile sur une séquence répétée cryptique appelée dRS3, soupçonnée depuis longtemps d'être impliquée dans la virulence du méningocoque en raison de sa forte présence autour de gènes codant pour des protéines de surface et de son abondance dans cette espèce par rapport aux espèces commensales. Les NTS, des séquences apparentées aux dRS3, sont spécifiquement ciblées sous forme d'ADN simple-brin par la nucléase SsnA. En montrant que SsnA est un déterminant majeur de la

virulence de *N. meningitidis*, notre étude corrobore enfin de façon expérimentale l'implication des dRS3 dans la pathogénicité de cette espèce.

En s'éloignant du genre *Neisseria*, nous avons identifié une vaste nouvelle famille d'enzymes appartenant à la superfamille des nucléases GIY-YIG : les Ssn. Cette famille regroupe des minuscules protéines avec un unique domaine fonctionnel, confirmées comme étant des endonucléases d'ADN simple-brin avec une panoplie de différentes spécificités, et probablement diverses fonctions biologiques. La diversité de cette famille et son activité catalytique unique en font un excellent candidat comme potentiels outils moléculaires. Ces découvertes font l'objet d'un brevet international (PCT) déposé en 2022, intitulé *Endonucleases that selectively cleave single-stranded nucleic acids and uses thereof* (PCT/CA2022/051668)(**Annexe II**).

Cette étude multidisciplinaire a également contribué au développement de plusieurs méthodologies adaptées aux *Neisseria* qui pourraient se révéler très utiles à d'autres équipes. On y trouve d'abord une approche d'analyse de la variation antigénique des pili par séquençage à haut débit affichant plusieurs avantages par rapport aux méthodologies publiées. On y trouve également l'utilisation d'un modèle d'infection murin par *N. meningitidis* optimisé par notre équipe. Enfin, cette étude exploite une méthodologie de manipulation génétique des *Neisseria* élaborée par mes collègues et moi-même, permettant l'accumulation de nombreuses mutations en limitant les marqueurs de sélection utilisés (Nyongesa *et al.*, 2022a).

BIBLIOGRAPHIE

- Achaz G, Rocha EP, Netter P, Coissac E (2002) Origin and fate of repeats in bacteria. *Nucleic Acids Research* 30(13):2987-2994.
- Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* 46(W1):W537-W544.
- Ailloud F, Estibariz I, Pfaffinger G, Suerbaum S (2022) The *Helicobacter pylori* UvrC nuclease is essential for chromosomal microimports after natural transformation. *Mbio* 13(4):e01811-01822.
- Alonso J-M, Guiyoule A, Zarantonelli ML, Ramisse F, Pires R, Antignac A, Deghmane AE, Huerre M, van Der Werf S, Taha M-K (2003) A model of meningococcal bacteremia after respiratory superinfection in influenza A virus-infected mice. *FEMS Microbiology Letters* 222(1):99-106.
- Ambur OH, Frye SA, Nilsen M, Hovland E, Tønjum T (2012) Restriction and sequence alterations affect DNA uptake sequence-dependent transformation in *Neisseria meningitidis*. *PLoS one* 7(7):e39742.
- ASPC (2019) *Report on Sexually Transmitted Infection Surveillance in Canada, 2019*. 15 p
- Bailey TL & Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology* 2:28-36.
- Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. *Nucleic Acids Research* 43(W1):W39-W49.
- Baltrus DA, Guillemin K, Phillips PC (2008) Natural transformation increases the rate of adaptation in the human pathogen *Helicobacter pylori*. *Evolution* 62(1):39-49.
- Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F (2008) Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell* 132(2):208-220.
- Barrangou R (2015) The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Current Opinions in Immunology* 32:36-41.
- Beauchamp JM, Leveque RM, Dawid S, DiRita VJ (2017) Methylation-dependent DNA discrimination in natural transformation of *Campylobacter jejuni*. *Proceedings of the National Academy of Sciences* 114(38):E8053-E8061.
- Bell JC & Kowalczykowski SC (2016) RecA: regulation and mechanism of a molecular search engine. *Trends in Biochemical Sciences* 41(6):491-507.
- Bentley SD, Vernikos GS, Snyder LAS, Churcher C, Arrowsmith C, Chillingworth T, Cronin A, Davis PH, Holroyd NE, Jagels K (2007) Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genetics* 3(2):e23.
- Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9:1-13.

- Berry J-L, Cehovin A, McDowell MA, Lea SM, Pelicic V (2013) Functional analysis of the interdependence between DNA uptake sequence and its cognate ComP receptor during natural transformation in *Neisseria* species. *PLoS Genetics* 9(12):e1004014.
- Bertani G & Weigle JJ (1953) Host controlled variation in bacterial viruses. *Journal of Bacteriology* 65(2):113-121.
- Bertels F, Gallie J, Rainey PB (2017) Identification and characterization of domesticated bacterial transposases. *Genome Biology and Evolution* 9(8):2110-2121.
- Bertels F & Rainey PB (2011) Curiosities of REPINs and RAYTs. *Mobile genetic elements* 1(4):262-301.
- Bhardwaj A & Nain V (2021) TALENs—an indispensable tool in the era of CRISPR: a mini review. *Journal of Genetic Engineering and Biotechnology* 19(1):1-10.
- Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA (2012) CRISPR interference can prevent natural transformation and virulence acquisition during *in vivo* bacterial infection. *Cell host & microbe* 12(2):177-186.
- Bille E, Zahar J-R, Perrin A, Morelle S, Kriz P, Jolley KA, Maiden MC, Dervin C, Nassif X, Tinsley CR (2005) A chromosomally integrated bacteriophage in invasive meningococci. *The Journal of experimental medicine* 201(12):1905-1913.
- Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Team G (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26(14):1783-1785.
- Blokesch M (2016) Natural competence for transformation. *Current Biology* 26(21):R1126-R1130.
- Boutroux M, Favre-Rochex S, Gorgette O, Touak G, Muhle E, Chesneau O, Clermont D, Rahi P (2023) *Neisseria leonis* sp. nov. isolated from rabbits, reclassification of *Uruburuella suis*, *Uruburuella testudinis*, *Kingella potus*, *Bergeriella denitrificans* and *Morococcus cerebrosus* into *Neisseria* genus and reclassification of *Neisseria shayegani* into *Eikenella* genus. *bioRxiv* :2023.2002. 2024.529859.
- Budroni S, Siena E, Hotopp JCD, Seib KL, Serruto D, Nofroni C, Comanducci M, Riley DR, Daugherty SC, Angiuoli SV (2011) *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proceedings of the National Academy of Sciences* 108(11):4494-4499.
- Burns BL & Rhoads DD (2022) Meningococcal Urethritis: Old and New. *Journal of Clinical Microbiology* 60(11):e00575-00522.
- Callaghan MM, Heilers J-H, van der Does C, Dillard JP (2017) Secretion of chromosomal DNA by the *Neisseria gonorrhoeae* Type IV secretion system. *Type IV Secretion in Gram-Negative and Gram-Positive Bacteria*, Backert S & Grohmann E (Édit.) Springer International Publishing, Cham 10.1007/978-3-319-75241-9_13. p 323-345.
- CDC (2022) CDC assists with meningococcal disease outbreak investigation in Florida. <https://www.cdc.gov/media/releases/2022/p0622-Meningococcal-Outbreak.html> (Consulté le 29 octobre)
- Cehovin A, Simpson PJ, McDowell MA, Brown DR, Noschese R, Pallett M, Brady J, Baldwin GS, Lea SM, Matthews SJ (2013) Specific DNA recognition mediated by a type IV pilin. *Proceedings of the National Academy of Sciences* 110(8):3065-3070.
- Chan S-H, Stoddard BL, Xu S-y (2011) Natural and engineered nicking endonucleases—from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Research* 39(1):1-18.

- Chan YA, Hackett KT, Dillard JP (2012) The lytic transglycosylases of *Neisseria gonorrhoeae*. *Microbial Drug Resistance* 18(3):271-279.
- Chandler M, De La Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature Reviews Microbiology* 11(8):525-538.
- Chandrasegaran S (2017) Recent advances in the use of ZFN-mediated gene editing for human gene therapy. *Cell & gene therapy insights* 3(1):33.
- Chenal M (2019) *Caractérisation d'une protéine hypothétique chez Neisseria meningitidis ; un régulateur important de la compétence naturelle*. Maîtrise en Sciences expérimentales de la santé, Maîtrise en Ingénierie de la santé, diagnostic biomédical (Université du Québec, Institut national de la recherche scientifique). 99 p
- Chenal M (2021) *Résistance aux antibiotiques : l'autre pandémie*. The Conversation, <https://theconversation.com/resistance-aux-antibiotiques-lautre-pandemie-149951> (Consulté le 6 Février)
- Chenal M & Huang E (2023) *DNA Secretion and Uptake Mechanisms in the Neisseria Bacteria*. BioRender, https://app.biorender.com/profile/Martin_Chenal/templates/6209a78761dbff009cb82f50 (Consulté le 29 octobre)
- Cheng K, Xu H, Chen X, Wang L, Tian B, Zhao Y, Hua Y (2016) Structural basis for DNA 5-end resection by RecJ. *Elife* 5:e14294.
- Christodoulides M & Heckels J (2017) Novel approaches to *Neisseria meningitidis* vaccine design. *Pathogens and Disease* 75(3).
- Cohen SN (2013) DNA cloning: a personal view after 40 years. *Proceedings of the National Academy of Sciences* 110(39):15521-15529.
- Colicchio R, Pagliuca C, Pastore G, Cicatiello AG, Pagliarulo C, Talà A, Scaglione E, Sammartino JC, Bucci C, Alifano P (2015) Fitness cost of rifampin resistance in *Neisseria meningitidis*: in vitro study of mechanisms associated with rpoB H553Y mutation. *Antimicrobial Agents and Chemotherapy* 59(12):7637-7649.
- Consortium U (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* 47(D1):D506-D515.
- Cordero M, García-Fernández J, Acosta IC, Yepes A, Avendano-Ortiz J, Lisowski C, Oesterreich B, Ohlsen K, Lopez-Collazo E, Förstner KU (2022) The induction of natural competence adapts staphylococcal metabolism to infection. *Nature communications* 13(1):1525.
- Corneloup A, Caumont-Sarcos A, Kamgoue A, Marty B, Le PTN, Siguier P, Guynet C, Ton-Hoang B (2021) TnpAREP and REP sequences dissemination in bacterial genomes: REP recognition determinants. *Nucleic Acids Research* 49(12):6982-6995.
- Crew PE, McNamara L, Waldron PE, McCulley L, Jones SC, Bersoff-Matcha SJ (2019) Unusual *Neisseria* species as a cause of infection in patients taking eculizumab. *Journal of Infection* 78(2):113-118.
- Criss AK, Kline KA, Seifert HS (2005) The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. *Molecular Microbiology* 58(2):510-519.
- Criss AK & Seifert HS (2012) A bacterial siren song: intimate interactions between *Neisseria* and neutrophils. *Nature Reviews Microbiology* 10(3):178-190.

- Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Research* 14(6):1188-1190.
- Dale AP, Theodosiou AA, Gbesemete DF, Guy JM, Jones EF, Hill AR, Ibrahim MM, de Graaf H, Ahmed M, Faust SN (2022) Effect of colonisation with *Neisseria lactamica* on cross-reactive anti-meningococcal B-cell responses: a randomised, controlled, human infection trial. *The Lancet Microbe* 3(12):e931-e943.
- Dalia AB, McDonough E, Camilli A (2014) Multiplex genome editing by natural transformation. *Proceedings of the National Academy of Sciences* 111(24):8937-8942.
- Davidson T, Rødland EA, Lagesen K, Seeberg E, Rognes T, Tønjum T (2004) Biased distribution of DNA uptake sequences towards genome maintenance genes. *Nucleic Acids Research* 32(3):1050-1058.
- Davidson T & Tønjum T (2006) Meningococcal genome dynamics. *Nature Reviews Microbiology* 4(1):11-22.
- Davies JK, Harrison PF, Lin Y-H, Bartley S, Khoo CA, Seemann T, Ryan CS, Kahler CM, Hill SA (2014) The use of high-throughput DNA sequencing in the investigation of antigenic variation: application to *Neisseria* species. *PLoS one* 9(1):e86704.
- De Gregorio E, Abrescia C, Carlomagno MS, Di Nocera PP (2002) The abundant class of nemis repeats provides RNA substrates for ribonuclease III in *Neisseriae*. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* 1576(1-2):39-44.
- de Vries J & Wackernagel W (2002) Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proceedings of the National Academy of Sciences* 99(4):2094-2099.
- Deghmane A-E, Veckerlé C, Giorgini D, Hong E, Ruckly C, Taha M-K (2009) Differential modulation of TNF- α -induced apoptosis by *Neisseria meningitidis*. *PLoS Pathogens* 5(5):e1000405.
- Dehe PM & Gaillard PHL (2017) Control of structure-specific endonucleases to maintain genome stability. *Nature Reviews in Molecular Cell Biology* 18(5):315-330.
- DeLano WL (2002) Pymol: An open-source molecular graphics tool. *Protein Crystallography* 40(1):82-92.
- Desai NA & Shankar V (2003) Single-strand-specific nucleases. *FEMS Microbiology Reviews* 26(5):457-491.
- Deutscher MP (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Research* 34(2):659-666.
- Deutscher MP & Li Z (2000) Exoribonucleases and their multiple roles in RNA metabolism. *Progress in Nucleic Acid Research and Molecular Biology* 66(2000):67-105.
- Di Felice F, Micheli G, Camilloni G (2019) Restriction enzymes and their use in molecular biology: An overview. *Journal of biosciences* 44(2):38.
- Diallo A, Foster HR, Gromek KA, Perry TN, Dujeancourt A, Krasteva PV, Gubellini F, Falbel TG, Burton BM, Fronzes R (2017) Bacterial transformation: ComFA is a DNA-dependent ATPase that forms complexes with ComFC and DprA. *Molecular Microbiology* 105(5):741-754.
- Dickey TH, Altschuler SE, Wuttke DS (2013) Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure* 21(7):1074-1084.

- Dillard JP & Seifert HS (2001) A variable genetic island specific for *Neisseria gonorrhoeae* is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. *Molecular Microbiology* 41(1):263-277.
- Dittrich CM, Kratz K, Sandoel A, Gruenbaum Y, Jiricny J, Hengartner MO (2012) LEM-3—a LEM domain containing nuclease involved in the DNA damage response in *C. elegans*. *PLoS one* 7(2):e24555.
- Dolange V, Churchward CP, Christodoulides M, Snyder LA (2018) The growing threat of gonococcal blindness. *Antibiotics* 7(3):59.
- Donahue JP, Israel DA, Peek Jr RM, Blaser MJ, Miller GG (2000) Overcoming the restriction barrier to plasmid transformation of *Helicobacter pylori*. *Molecular Microbiology* 37(5):1066-1074.
- Dorey RB, Theodosiou AA, Read RC, Jones CE (2019) The nonpathogenic commensal *Neisseria*: friends and foes in infectious disease. *Current Opinion in Infectious Diseases* 32(5):490-496.
- Dorward DW, Garon CF, Judd RC (1989) Export and intercellular transfer of DNA via membrane blebs of *Neisseria gonorrhoeae*. *Journal of Bacteriology* 171(5):2499-2505.
- Duffin PM & Barber DA (2016) DprA is required for natural transformation and affects pilin variation in *Neisseria gonorrhoeae*. *Microbiology* 162(9):1620-1628.
- Duffin PM & Seifert HS (2012) Genetic transformation of *Neisseria gonorrhoeae* shows a strand preference. *FEMS Microbiology Letters* 334(1):44-48.
- Dunin-Horkawicz S, Feder M, Bujnicki JM (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7(1):1-19.
- Dwivedi GR, Sharma E, Rao DN (2013) *Helicobacter pylori* DprA alleviates restriction barrier for incoming DNA. *Nucleic Acids Research* 41(5):3274-3288.
- Edgell DR, Gibb EA, Belfort M (2010) Mobile DNA elements in T4 and related phages. *Virology journal* 7(1):1-15.
- Epinat JC, Silva GH, Pâques F, Smith J, Duchateau P (2013) Engineered meganucleases for genome engineering purposes. *Site-directed insertion of transgenes* :147-185.
- Evans CJ & Aguilera RJ (2003) DNase II: genes, enzymes and function. *Gene* 322:1-15.
- Evgen'ev M & Arkhipova I (2005) Penelope-like elements—a new class of retroelements: distribution, function and possible evolutionary significance. *Cytogenetic and Genome Research* 110(1-4):510-521.
- Finkel SE & Kolter R (2001) DNA as a nutrient: novel role for bacterial competence gene homologs. *Journal of Bacteriology* 183(21):6288-6293.
- Fiore MA, Raisman JC, Wong NH, Hudson AO, Wadsworth CB (2020) Exploration of the *Neisseria* resistome reveals resistance mechanisms in commensals that may be acquired by *N. gonorrhoeae* through horizontal gene transfer. *Antibiotics* 9(10):656.
- Fricke WM & Brill SJ (2003) Slx1—Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1—Top3. *Genes & Development* 17(14):1768-1778.
- Frickey T & Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20(18):3702-3704.

- Frye SA, Nilsen M, Tønjum T, Ambur OH (2013) Dialects of the DNA uptake sequence in *Neisseriaceae*. *PLoS Genetics* 9(4):e1003458.
- Fukui K, Harada A, Wakamatsu T, Minobe A, Ohshita K, Ashiuchi M, Yano T (2018) The GIY-YIG endonuclease domain of *Arabidopsis* MutS homolog 1 specifically binds to branched DNA structures. *FEBS Letters* 592(24):4066-4077.
- Green MR & Sambrook J (2018) The Hanahan method for preparation and transformation of competent *Escherichia coli*: high-efficiency transformation. *Cold Spring Harbor Protocols* 2018(3):pdb. prot101188.
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The vienna RNA websuite. *Nucleic Acids Research* 36(suppl_2):W70-W74.
- Guynet C, Achard A, Hoang BT, Barabas O, Hickman AB, Dyda F, Chandler M (2009) Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Molecular Cell* 34(5):612-619.
- Hafez M & Hausner G (2012) Homing endonucleases: DNA scissors on a mission. *Genome* 55(8):553-569.
- Hamilton HL & Dillard JP (2006) Natural transformation of *Neisseria gonorrhoeae*: from DNA donation to homologous recombination. *Molecular Microbiology* 59(2):376-385.
- Hebeler BH & Young FE (1976) Mechanism of autolysis of *Neisseria gonorrhoeae*. *Journal of Bacteriology* 126(3):1186-1193.
- Helm RA & Seifert HS (2010) Frequency and rate of pilin antigenic variation of *Neisseria meningitidis*. *Journal of Bacteriology* 192(14):3822-3823.
- Hepp C & Maier B (2016) Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism. *Proceedings of the National Academy of Sciences* 113(44):12467-12472.
- Hilse R, Stoevesandt J, Caugant D, Claus H, Frosch M, Vogel U (2000) Distribution of the meningococcal insertion sequence IS1301 in clonal lineages of *Neisseria meningitidis*. *Epidemiology & Infection* 124(2):337-340.
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35(2):518-522.
- Ibanez de Aldecoa AL, Zafra O, González-Pastor JE (2017) Mechanisms and regulation of extracellular DNA release and its biological roles in microbial communities. *Frontiers in microbiology* 8:1390.
- Ibryashkina EM, Sasnauskas G, Solonin AS, Zakharova MV, Siksnyš V (2009) Oligomeric structure diversity within the GIY-YIG nuclease family. *Journal of Molecular Biology* 387(1):10-16.
- Ibryashkina EM, Zakharova MV, Baskunov VB, Bogdanova ES, Nagornykh MO, Den'mukhamedov MM, Melnik BS, Kolinski A, Gront D, Feder M (2007) Type II restriction endonuclease R. Eco29kl is a member of the GIY-YIG nuclease superfamily. *BMC Structural Biology* 7:1-10.
- Jain S, Zweig M, Peeters E, Siewering K, Hackett KT, Dillard JP, Van Der Does C (2012) Characterization of the single stranded DNA binding protein SsbB encoded in the Gonococcal Genetic Island. *PloS one* 7(4):e35285.

- Jangam D, Feschotte C, Betrán E (2017) Transposable element domestication as an adaptation to evolutionary conflicts. *Trends in Genetics* 33(11):817-831.
- Johnston C, Martin B, Fichant G, Polard P, Claverys J-P (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology* 12(3):181-196.
- Johnston C, Martin B, Granadel C, Polard P, Claverys J-P (2013) Programmed protection of foreign DNA from restriction allows pathogenicity island exchange during pneumococcal transformation. *PLoS Pathogens* 9(2):e1003178.
- Johnston CHG, Hope R, Soulet A-L, Dewailly M, De Lemos D, Polard P (2023) The RecA-directed recombination pathway of natural transformation initiates at chromosomal replication forks in the pneumococcus. *Proceedings of the National Academy of Sciences* 120(8):e2213867120.
- Jones J, Weiss K, Mermin J, Dietz P, Rosenberg ES, Gift TL, Chesson H, Sullivan PS, Lyles C, Bernstein KT (2019) Proportion of incident HIV cases among men who have sex with men attributable to gonorrhea and chlamydia: a modeling analysis. *Sexually Transmitted Diseases* 46(6):357.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583-589.
- Kaminska KH, Kawai M, Boniecki M, Kobayashi I, Bujnicki JM (2008) Type II restriction endonuclease R. Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Structural Biology* 8(1):1-14.
- Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30(4):772-780.
- Kawai M, Uchiyama I, Kobayashi I (2005) Genome *comparison* in silico in *Neisseria* suggests integration of filamentous bacteriophages by their own transposase. *DNA Research* 12(6):389-401.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* 10(6):845-858.
- Kim E, Kim S, Kim DH, Choi B-S, Choi I-Y, Kim J-S (2012) Precision genome engineering with programmable DNA-nicking enzymes. *Genome Research* 22(7):1327-1333.
- Kim H & Kim J-S (2014) A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics* 15(5):321-334.
- Kim WJ, Higashi D, Goytia M, Rendón MA, Pilligua-Lucas M, Bronnimann M, McLean JA, Duncan J, Trees D, Jerse AE (2019) Commensal *Neisseria* kill *Neisseria gonorrhoeae* through a DNA-dependent mechanism. *Cell host & microbe* 26(2):228-239. e228.
- Kishi K, Yasuda T, Takeshita H (2001) DNase I: structure, function, and use in medicine and forensic science. *Leg Med (Tokyo)* 3(2):69-83.
- Kleinstiver BP, Wolfs JM, Edgell DR (2013) The monomeric GIY-YIG homing endonuclease I-Bmol uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Research* 41(10):5413-5427.

- Kleinstiver BP, Wolfs JM, Kolaczyk T, Roberts AK, Hu SX, Edgell DR (2012) Monomeric site-specific nucleases for genome editing. *Proceedings of the National Academy of Sciences* 109(21):8061-8066.
- Kline KA & Seifert HS (2005) Mutation of the *priA* gene of *Neisseria gonorrhoeae* affects DNA transformation and DNA repair. *Journal of Bacteriology* 187(15):5347-5355.
- Knott GJ & Doudna JA (2018) CRISPR-Cas guides the future of genetic engineering. *Science* 361(6405):866-869.
- Koonin EV (2017) Evolution of RNA- and DNA-guided antiviral defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. *Biology direct* 12(1):5.
- Koonin EV, Makarova KS, Wolf YI (2017) Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annual Reviews of Microbiology* 71:233-261.
- Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, Baxter SM, Derbyshire V (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-Tev I: coincidence of computational and molecular findings. *Nucleic Acids Research* 27(10):2115-2125.
- Kramer N, Hahn J, Dubnau D (2007) Multiple interactions among the competence proteins of *Bacillus subtilis*. *Molecular Microbiology* 65(2):454-464.
- Kumar A, Beloglazova N, Bundalovic-Torma C, Phanse S, Deineko V, Gagarinova A, Musso G, Vlasblom J, Lemak S, Hooshyar M (2016) Conditional epistatic interaction maps reveal global functional rewiring of genome integrity pathways in *Escherichia coli*. *Cell reports* 14(3):648-661.
- Kung SH, Retchless AC, Kwan JY, Almeida RP (2013) Effects of DNA size on transformation and recombination efficiencies in *Xylella fastidiosa*. *Applied and Environmental Microbiology* 79(5):1712-1717.
- Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4):357-359.
- Lauková L, Konečná B, Janovičová L, Vlková B, Celec P (2020) Deoxyribonucleases and their applications in biomedicine. *Biomolecules* 10(7):1036.
- Lerminiaux NA & Cameron AD (2019) Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian journal of microbiology* 65(1):34-44.
- Letunic I & Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 49(W1):W293-W296.
- Li JJ, Geyer R, Tan W (2000) Using molecular beacons as a sensitive fluorescence assay for enzymatic cleavage of single-stranded DNA. *Nucleic Acids Research* 28(11):e52-e52.
- Li W & Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658-1659.
- Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Reviews of Biochemistry* 79:181-211.
- Linz B, Schenker M, Zhu P, Achtman M (2000) Frequent interspecific genetic exchange between commensal *Neisseriae* and *Neisseria meningitidis*. *Molecular Microbiology* 36(5):1049-1058.
- Liu G, Tang CM, Exley RM (2015) Non-pathogenic *Neisseria*: members of an abundant, multi-habitat, diverse genus. *Microbiology* 161(Pt_7):1297-1312.

- Liu M, Huang M, Wang M, Zhu D, Jia R, Chen S, Zhang L, Pan L, Cheng A (2020) The Clustered Regularly Interspaced Short Palindromic Repeat System and Argonaute: An Emerging Bacterial Immunity System for Defense Against Natural Transformation? *Frontiers in microbiology* 11:593301.
- Liu Q, Derbyshire V, Belfort M, Edgell DR (2006) Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Research* 34(6):1755-1764.
- Liu SV, Saunders NJ, Jeffries A, Rest RF (2002) Genome analysis and strain comparison of *correa* repeats and *correa* repeat-enclosed elements in pathogenic *Neisseria*. *Journal of Bacteriology* 184(22):6163-6173.
- Liu X, Liu S, Feng Y, Liu J-Z, Chen Y, Pham K, Deng H, Hirschi KD, Wang X, Cheng N (2013) Structural insights into the N-terminal GIY-YIG endonuclease activity of *Arabidopsis* glutaredoxin AtGRXS16 in chloroplasts. *Proceedings of the National Academy of Sciences* 110(23):9565-9570.
- Loenen WA & Raleigh EA (2014) The other face of restriction: modification-dependent enzymes. *Nucleic Acids Research* 42(1):56-69.
- Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Research* 42(1):20-44.
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15(12):1-21.
- Lu Q-F, Cao D-M, Su L-L, Li S-B, Ye G-B, Zhu X-Y, Wang J-P (2019) Genus-wide comparative genomics analysis of *Neisseria* to identify new genes associated with pathogenicity and niche adaptation of *Neisseria pathogens*. *International journal of genomics* 2019.
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* 48(D1):D265-D268.
- Luria SE & Human ML (1952) A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology* 64(4):557-569.
- Madeira F, Pearce M, Tivey AR, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research* 50(W1):W276-W279.
- Mahillon J & Chandler M (1998) Insertion sequences. *Microbiology and Molecular Biology Reviews* 62(3):725-774.
- Maiden MC & Harrison OB (2016) Population and functional genomics of *Neisseria* revealed with gene-by-gene approaches. *Journal of Clinical Microbiology* 54(8):1949-1955.
- Mak AN-S, Lambert AR, Stoddard BL (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R. Eco29kl. *Structure* 18(10):1321-1331.
- Manils J, Marruecos L, Soler C (2022) Exonucleases: Degrading DNA to Deal with Genome Damage, Cell Death, Inflammation and Cancer. *Cells* 11(14):2157.
- Mao J & Lu T (2016) Population-dynamic modeling of bacterial horizontal gene transfer by natural transformation. *Biophysical Journal* 110(1):258-268.
- Marinus MG (2012) DNA Mismatch Repair. *EcoSal Plus* 5(1).

- Marri PR, Paniscus M, Weyand NJ, Rendón MA, Calton CM, Hernández DR, Higashi DL, Sodergren E, Weinstock GM, Rounsley SD (2010) Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. *PLoS one* 5(7):e11835.
- Mazzone M, De Gregorio E, Lavitola A, Pagliarulo C, Alifano P, Di Nocera PP (2001) Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic *Neisseriae*. *Gene* 278(1-2):211-222.
- McMurrough TA, Brown CM, Zhang K, Hausner G, Junop MS, Gloor GB, Edgell DR (2018) Active site residue identity regulates cleavage preference of LAGLIDADG homing endonucleases. *Nucleic Acids Research* 46(22):11990-12007.
- McNamara LA, Potts C, Blain AE, Retchless AC, Reese N, Swint S, Lonsway D, Karlsson M, Lunquist K, Sweitzer JJ (2020) Detection of ciprofloxacin-resistant, β -lactamase-producing *Neisseria meningitidis* serogroup Y isolates—United States, 2019–2020. *Morbidity and Mortality Weekly Report* 69(24):735.
- Mell JC & Redfield RJ (2014) Natural competence and the evolution of DNA uptake specificity. *Journal of Bacteriology* 196(8):1471-1483.
- Minocherhomji S & Hickson ID (2014) Structure-specific endonucleases: guardians of fragile site stability. *Trends in Cell Biology* 24(5):321-327.
- Mol CD, Hosfield DJ, Tainer JA (2000) Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means. *Mutation Research/DNA Repair* 460(3-4):211-229.
- Mullany P (2010) *The dynamic bacterial genome*. Cambridge University Press,
- Nalefski EA, Patel N, Leung PJ, Islam Z, Kooistra RM, Parikh I, Marion E, Knott GJ, Doudna JA, Le Ny A-LM (2021) Kinetic analysis of Cas12a and Cas13a RNA-Guided nucleases for development of improved CRISPR-Based diagnostics. *Science* 24(9):102996.
- Nassif X, Lowy J, Stenberg P, O'Gaora P, Ganji A, So M (1993) Antigenic variation of pilin regulates adhesion of *Neisseria meningitidis* to human epithelial cells. *Molecular Microbiology* 8(4):719-725.
- Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32(1):268-274.
- Nishino T & Morikawa K (2002) Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene* 21(58):9022-9032.
- Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302(1):205-217.
- Nunvar J, Huckova T, Licha I (2010) Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. *BMC Genomics* 11:1-13.
- Nunvar J, Licha I, Schneider B (2013) Evolution of REP diversity: a comparative study. *BMC Genomics* 14(1):1-11.
- Nyongesa S, Chenal M, Bernet È, Coudray F, Veyrier FJ (2022a) Sequential markerless genetic manipulations of species from the *Neisseria* genus. *Canadian journal of microbiology* 68(8):551-560.

- Nyongesa S, Weber PM, Bernet È, Pulido F, Nieves C, Nieckarz M, Delaby M, Viehboeck T, Krause N, Rivera-Millot A (2022b) Evolution of longitudinal division in multicellular bacteria of the *Neisseriaceae* family. *Nature communications* 13(1):4853.
- Oberto J (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* 14(1):1-10.
- Palchevskiy V & Finkel SE (2009) A role for single-stranded exonucleases in the use of DNA as a nutrient. *Journal of Bacteriology* 191(11):3712-3716.
- Palmer GH, Bankhead T, Seifert HS (2016) Antigenic variation in bacterial pathogens. *Virulence Mechanisms of Bacterial Pathogens* :445-480.
- Pandey AK, Cleary DW, Laver JR, Maiden MC, Didelot X, Gorringer A, Read RC (2017) *Neisseria lactamica* Y92–1009 complete genome sequence. *Standards in Genomic Sciences* 12:1-9.
- Parkhill J, Achtman M, James K, Bentley S, Churcher C, Klee S, Morelli G, Basham D, Brown D, Chillingworth T (2000) Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 404(6777):502-506.
- Pearson WR (2001) FASTA search programs. *eLS*.
- Pereira CV, Bacman SR, Arguello T, Zekonyte U, Williams SL, Edgell DR, Moraes CT (2018) mitoTev-TALE: a monomeric DNA editing enzyme to reduce mutant mitochondrial DNA levels. *EMBO molecular medicine* 10(9):e8084.
- Petousis-Harris H & Radcliff FJ (2019) Exploitation of *Neisseria meningitidis* group B OMV vaccines against *N. gonorrhoeae* to inform the development and deployment of effective gonorrhea vaccines. *Frontiers in Immunology* 10:683.
- Pressler T (2008) Review of recombinant human deoxyribonuclease (rhDNase) in the management of patients with cystic fibrosis. *Biologics: Targets and Therapy* 2(4):611-617.
- Priniski LL & Seifert HS (2018) A case for the evolution from commensalism to pathogenicity and possibly back again: lessons learned from the human-adapted *Neisseria* species. *Molecular Mechanisms of Microbial Evolution* :327-370.
- Rao DN, Dryden DT, Bheemanaik S (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Research* 42(1):45-55.
- Redfield R (1993) Genes for breakfast: The have-your-cake and-eat-it-too of bacterial transformation. *Journal of Heredity* 84(5):400-404.
- Reyes-Robles T, Dillard RS, Cairns LS, Silva-Valenzuela CA, Housman M, Ali A, Wright ER, Camilli A (2018) *Vibrio cholerae* outer membrane vesicles inhibit bacteriophage infection. *Journal of Bacteriology* 200(15):10.1128/jb.00792-00717.
- Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16(6):276-277.
- Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research* 43(Database issue):D298-299.
- Roberts SB, Spencer-Smith R, Shah M, Nebel J-C, Cook RT, Snyder LA (2016) Correia repeat enclosed elements and non-coding RNAs in the *Neisseria* species. *Microorganisms* 4(3):31.

- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
- Rostamian M, Lorestani RC, Jafari S, Mansouri R, Rezaeian S, Ghadiri K, Akya A (2022) A systematic review and meta-analysis on the antibiotic resistance of *Neisseria meningitidis* in the last 20 years in the world. *Indian Journal of Medical Microbiology*.
- Rotman E & Seifert HS (2014) The genetics of *Neisseria* species. *Annual Review of Genetics* 48:405-431.
- Rotman E, Webber DM, Seifert HS (2016) Analyzing *Neisseria gonorrhoeae* pilin antigenic variation using 454 sequencing technology. *Journal of Bacteriology* 198(18):2470-2482.
- Rouphael NG & Stephens DS (2012) *Neisseria meningitidis*: biology, microbiology, and epidemiology. *Neisseria meningitidis: advanced methods and protocols* :1-20.
- Rouquette-Loughlin CE, Balthazar JT, Hill SA, Shafer WM (2004) Modulation of the mtrCDE-encoded efflux pump gene complex of *Neisseria meningitidis* due to a Correia element insertion sequence. *Molecular Microbiology* 54(3):731-741.
- Sanders H, Brehony C, Maiden MC, Vipond C, Feavers IM (2012) The effect of iron availability on transcription of the *Neisseria meningitidis* fHbp gene varies among clonal complexes. *Microbiology* 158(Pt 4):869.
- Saramago M, Barria C, Dos Santos RF, Silva IJ, Pobre V, Domingues S, Andrade JM, Viegas SC, Arraiano CM (2014) The role of RNases in the regulation of small RNAs. *Current Opinions in Microbiology* 18:105-115.
- Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9(7):671-675.
- Schoen C, Blom J, Claus H, Schramm-Glück A, Brandt P, Müller T, Goesmann A, Joseph B, Konietzny S, Kurzai O (2008) Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences* 105(9):3473-3478.
- Schoen C, Tettelin H, Parkhill J, Frosch M (2009) Genome flexibility in *Neisseria meningitidis*. *Vaccine* 27:B103-B111.
- Schwechheimer C & Kuehn MJ (2015) Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nature Reviews Microbiology* 13(10):605-619.
- Seib KL & Peak IR (2019) *Neisseria Meningitidis: Methods and Protocols*. Springer,
- Seifert HS (1997) Insertionally inactivated and inducible recA alleles for use in *Neisseria*. *Gene* 188(2):215-220.
- Seifert HS (2019) Location, location, location—commensalism, damage and evolution of the pathogenic *Neisseria*. *Journal of Molecular Biology* 431(16):3010-3014.
- Seitz P & Blokesch M (2013) Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiology Reviews* 37(3):336-363.
- Sharma DK, Misra HS, Bihani SC, Rajpurohit YS (2023) Biochemical Properties and Roles of DprA Protein in Bacterial Natural Transformation, Virulence, and Pilin Variation. *Journal of Bacteriology* :e00465-00422.
- Shen W, Le S, Li Y, Hu F (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PloS one* 11(10):e0163962.

- Shevelev IV & Hübscher U (2002) The 3'–5' exonucleases. *Nature reviews Molecular cell biology* 3(5):364-376.
- Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI (2017) Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology* 15(3):169-182.
- Siddique A, Buisine N, Chalmers R (2011) The transposon-like Correia elements encode numerous strong promoters and provide a potential new mechanism for phase variation in the meningococcus. *PLoS Genetics* 7(1):e1001277.
- Siena E, D'Aurizio R, Riley D, Tettelin H, Guidotti S, Torricelli G, Moxon ER, Medini D (2016) In-silico prediction and deep-DNA sequencing validation indicate phase variation in 115 *Neisseria meningitidis* genes. *BMC Genomics* 17(1):1-13.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7(1):539.
- Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews* 38(5):865-891.
- Singh PK, Bourque G, Craig NL, Dubnau JT, Feschotte C, Flasch DA, Gunderson KL, Malik HS, Moran JV, Peters JE (2014) Mobile genetic elements and genome evolution 2014. (Springer).
- Singh S, Folkers G, Bonvin A, Boelens R, Wechselberger R, Niztayev A, Kaptein R (2002) Solution structure and DNA-binding properties of the C-terminal domain of UvrC from *E. coli*. *The EMBO Journal* 21(22):6257-6266.
- Sinha S, Mell J, Redfield R (2013) The availability of purine nucleotides regulates natural competence by controlling translation of the competence activator Sxy. *Molecular Microbiology* 88(6):1106-1119.
- Spåhr H, Chia T, Lingford JP, Siira SJ, Cohen SB, Filipovska A, Rackham O (2018) Modular ssDNA binding and inhibition of telomerase activity by designer PPR proteins. *Nature communications* 9(1):2212.
- Spencer-Smith R, Roberts S, Gurung N, Snyder LA (2016) DNA uptake sequences in *Neisseria gonorrhoeae* as intrinsic transcriptional terminators and markers of horizontal gene transfer. *Microbial genomics* 2(8).
- Spinosa MR, Progidia C, Tala A, Cogli L, Alifano P, Bucci C (2007) The *Neisseria meningitidis* capsule is important for intracellular survival in human cells. *Infection and Immunity* 75(7):3594-3603.
- Stein DC, Gunn JS, Radlinska M, Piekarowicz A (1995) Restriction and modification systems of *Neisseria gonorrhoeae*. *Gene* 157(1-2):19-22.
- Stohl EA & Seifert HS (2001) The recX gene potentiates homologous recombination in *Neisseria gonorrhoeae*. *Molecular Microbiology* 40(6):1301-1310.
- Su W, Sun J, Shimizu K, Kadota K (2019) TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC research notes* 12(1):1-6.
- Sutherland TC, Quattroni P, Exley RM, Tang CM (2010) Transcellular passage of *Neisseria meningitidis* across a polarized respiratory epithelium. *Infection and Immunity* 78(9):3832-3847.

- Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, Wang Y, Patel DJ, Berenguer J, Brouns SJ (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507(7491):258-261.
- Thakur M, Agarwal A, Muniyappa K (2021) The intrinsic ATPase activity of *Mycobacterium tuberculosis* UvrC is crucial for its damage-specific DNA incision function. *The FEBS Journal* 288(4):1179-1200.
- Thakur M, Badugu S, Muniyappa K (2020) UvrA and UvrC subunits of the *Mycobacterium tuberculosis* UvrABC excinuclease interact independently of UvrB and DNA. *FEBS Letters* 594(5):851-863.
- Thakur M, Kumar MB, Muniyappa K (2016) *Mycobacterium tuberculosis* UvrB is a robust DNA-stimulated ATPase that also possesses structure-specific ATP-dependent DNA helicase activity. *Biochemistry* 55(41):5865-5883.
- Tobes R & Pareja E (2006) Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. *BMC Genomics* 7:1-12.
- Tobiason DM & Seifert HS (2010) Genomic content of *Neisseria* species. *Journal of Bacteriology* 192(8):2160-2168.
- Ton-Hoang B, Siguier P, Quentin Y, Onillon S, Marty B, Fichant G, Chandler M (2012) Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *Nucleic Acids Research* 40(8):3596-3609.
- Ton-Hoang B, Guynet C, Ronning DR, Cointin-Marty B, Dyda F, Chandler M (2005) Transposition of ISHp608, member of an unusual family of bacterial insertion sequences. *The EMBO Journal* 24(18):3325-3338.
- Traglia GM, Quinn B, Schramm ST, Soler-Bistue A, Ramirez MS (2016) Serum albumin and Ca²⁺ are natural competence inducers in the human pathogen *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy* 60(8):4920-4929.
- Treangen TJ, Abraham A-L, Touchon M, Rocha EP (2009) Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiology Reviews* 33(3):539-571.
- Tröder SE & Zevnik B (2022) History of genome editing: From meganucleases to CRISPR. *Laboratory Animals* 56(1):60-68.
- Tsang RS (2021) A narrative review of the molecular epidemiology and laboratory surveillance of vaccine preventable bacterial meningitis agents: *Streptococcus pneumoniae*, *Neisseria meningitidis*, *Haemophilus influenzae* and *Streptococcus agalactiae*. *Microorganisms* 9(2):449.
- Turmel M, Otis C, Côté V, Lemieux C (1997) Evolutionarily conserved and functionally important residues in the I-Ceu I homing endonuclease. *Nucleic Acids Research* 25(13):2610-2619.
- Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM (2015) NPdock: a web server for protein–nucleic acid docking. *Nucleic Acids Research* 43(W1):W425-W430.
- Unemo M & Shafer WM (2014) Antimicrobial resistance in *Neisseria gonorrhoeae* in the 21st century: past, evolution, and future. *Clinical Microbiology Reviews* 27(3):587-613.
- Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics* 11(9):636-646.

- Van Houten B, Eisen JA, Hanawalt PC (2002) A cut above: discovery of an alternative excision repair pathway in bacteria. *Proceedings of the National Academy of Sciences* 99(5):2581-2583.
- van Passel MW, van der Ende A, Bart A (2006) Plasmid diversity in *Neisseriae*. *Infection and Immunity* 74(8):4892-4899.
- Vanbaelen T, Van Dijck C, Laumen J, Gonzalez N, De Baetselier I, Manoharan-Basil SS, De Block T, Kenyon C (2022) Global epidemiology of antimicrobial resistance in commensal *Neisseria* species: a systematic review. *International Journal of Medical Microbiology* :151551.
- Vasu K & Nagaraja V (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and Molecular Biology Reviews* 77(1):53-72.
- Verma P, Kumari P, Negi S, Yadav G, Gaur V (2022) Holliday junction resolution by At-HIGLE: an SLX1 lineage endonuclease from *Arabidopsis thaliana* with a novel in-built regulatory mechanism. *Nucleic Acids Research* 50(8):4630-4646.
- Veyrier F, Pletzer D, Turenne C, Behr MA (2009) Phylogenetic detection of horizontal gene transfer during the step-wise genesis of *Mycobacterium tuberculosis*. *BMC Evolutionary Biology* 9(1):1-14.
- Veyrier FJ, Biais N, Morales P, Belkacem N, Guilhen C, Ranjeva S, Sismeiro O, Péhau-Arnaudet G, Rocha EP, Werts C (2015) Common cell shape evolution of two nasopharyngeal pathogens. *PLoS Genetics* 11(7):e1005338.
- Vink C, Rudenko G, Seifert HS (2012) Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiology Reviews* 36(5):917-948.
- Völkel P, Le Faou P, Angrand P-O (2010) Interaction proteomics: characterization of protein complexes using tandem affinity purification–mass spectrometry. *Biochemical Society Transactions* 38(4):883-887.
- Vorkapic D, Pressler K, Schild S (2016) Multifaceted roles of extracellular DNA in bacterial physiology. *Current Genetics* 62:71-79.
- Wakamatsu T, Kim K, Uemura Y, Nakagawa N, Kuramitsu S, Masui R (2011) Role of RecJ-like protein with 5'-3' exonuclease activity in oligo (deoxy) nucleotide degradation. *Journal of Biological Chemistry* 286(4):2807-2816.
- Wang W, Ding J, Zhang Y, Hu Y, Wang D-C (2014) Structural insights into the unique single-stranded DNA-binding mode of *Helicobacter pylori* DprA. *Nucleic Acids Research* 42(5):3478-3491.
- Warfel JD & LiCata VJ (2015) Enhanced DNA binding affinity of RecA protein from *Deinococcus radiodurans*. *DNA Repair* 31:91-96.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.
- Weyand NJ, Ma M, Phifer-Rixey M, Taku NA, Rendón MA, Hockenberry AM, Kim WJ, Agellon AB, Biais N, Suzuki TA (2016) Isolation and characterization of *Neisseria musculi* sp. nov., from the wild house mouse. *International Journal of Systematic and Evolutionary Microbiology* 66(9):3585.
- WHO (2021a) *Defeating meningitis by 2030: a global road map*. p

- WHO (2021b) *Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021: accountability for the global health sector strategies 2016–2021: actions for impact*. (9240027076) p
- Willkomm S, Makarova KS, Grohmann D (2018) DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. *FEMS Microbiology Reviews* 42(3):376-387.
- Wilson RC & Doudna JA (2013) Molecular mechanisms of RNA interference. *Annual review of biophysics* 42:217-239.
- Wolfs JM, Hamilton TA, Lant JT, Laforet M, Zhang J, Salemi LM, Gloor GB, Schild-Poulter C, Edgell DR (2016) Biasing genome-editing events toward precise length deletions with an RNA-guided TevCas9 dual nuclease. *Proceedings of the National Academy of Sciences* 113(52):14988-14993.
- Wyatt HD & West SC (2014) Holliday junction resolvases. *Cold Spring Harbour Perspectives in Biology* 6(9):a023192.
- Xu SY (2015) Sequence-specific DNA nicking endonucleases. *Biomol Concepts* 6(4):253-267.
- Yadav T, Carrasco B, Myers AR, George NP, Keck JL, Alonso JC (2012) Genetic recombination in *Bacillus subtilis*: a division of labor between two single-strand DNA-binding proteins. *Nucleic Acids Research* 40(12):5546-5559.
- Yadav T, Carrasco B, Serrano E, Alonso JC (2014) Roles of *Bacillus subtilis* DprA and SsbA in RecA-mediated genetic recombination. *Journal of Biological Chemistry* 289(40):27640-27652.
- Yan Y, Zhang D, Zhou P, Li B, Huang S-Y (2017) HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research* 45(W1):W365-W373.
- Yang J, Cho WC, Zheng Y (2020) Argonaute proteins: Structural features, functions and emerging roles. *Journal of Advanced Research* 24:317-324.
- Yang W (2011) Nucleases: diversity of structure, function and mechanism. *Quarterly Reviews of Biophysics* 44(1):1-93.
- Yu Y, Ouyang Y, Yao W (2018) shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* 34(7):1229-1231.
- Zarantonelli M, Lancellotti M, Deghmane A-E, Giorgini D, Hong E, Ruckly C, Alonso J-M, Taha M-K (2008) Hyperinvasive genotypes of *Neisseria meningitidis* in France. *Clinical Microbiology and Infection* 14(5):467-472.
- Zhang Y (2017) The CRISPR-Cas9 system in *Neisseria* spp. *Pathogens and Disease* 75(4):ftx036.
- Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ (2013) Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Molecular Cell* 50(4):488-503.
- Zhang Y, Rajan R, Seifert HS, Mondragón A, Sontheimer EJ (2015) DNase H activity of *Neisseria meningitidis* Cas9. *Molecular Cell* 60(2):242-255.
- Zhu L, Lin J, Kuang Z, Vidal JE, Lau GW (2015) Deletion analysis of *S treptococcus pneumoniae* late competence genes distinguishes virulence determinants that are dependent or independent of competence induction. *Molecular Microbiology* 97(1):151-165.

ANNEXE I – MODÉLISATIONS *IN SILICO*

En attendant de déterminer expérimentalement l'interaction SsnA-ADNsb par co-cristallisation, des modélisations *in silico* ont été réalisées par notre collaborateur à l'INRS, Yossef Lopez de los Santos. Bien qu'il s'agisse de prédictions, la simplicité relative de SsnA et de son ligand, couplée à plusieurs informations clés comme le site de clivage et les résidus essentiels à l'activité catalytique, a permis d'inférer un modèle assez fiable (**Figure A.I**). Ce dernier a été établi par une approche hybride utilisant plusieurs algorithmes, dont HDock et NPDock (Tuszynska *et al.*, 2015; Yan *et al.*, 2017).

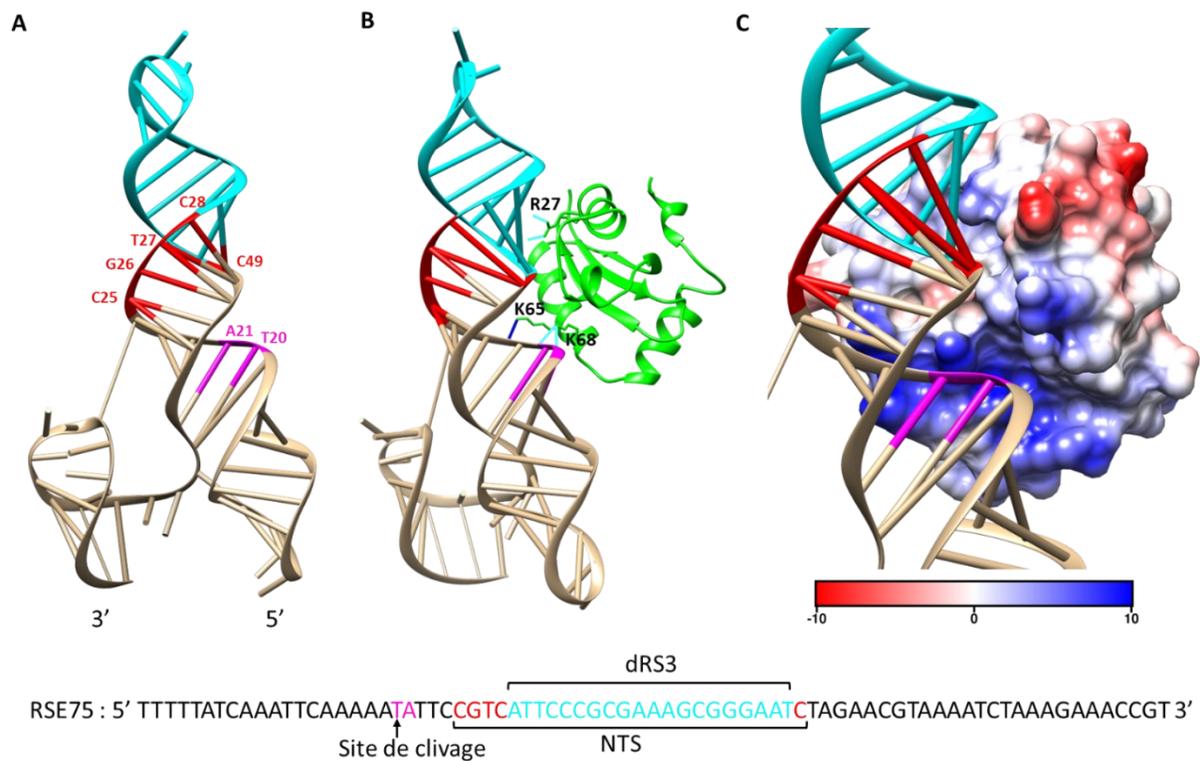


Figure A.I : Modélisation *in silico* de l'interaction SsnA-NTS

A. Structure secondaire prédite de l'oligonucléotide RSE75 utilisé dans les tests de nucléase. La séquence palindromique correspondant à l'élément répété dRS3 est représentée en cyan, tandis que les extensions caractéristiques de l'élément NTS sont en rouge. Le site de clivage est coloré en magenta. **B.** Modèle le plus probable d'interaction entre SsnA (vert) et la NTS. Les résidus protéiques prédits comme interagissant directement avec l'ADNsb sont identifiés. **C.** Représentation de la surface de SsnA liée à la NTS, colorée selon le potentiel électrostatique Coulombien (Bleu = charge positive, rouge = charge négative). En bas, la séquence complète de RSE75 contenant la NTS est affichée, avec le même code couleur qu'en A-B-C. Modélisations générées par notre collaborateur Yossef de los Santos, assemblées et visualisées avec l'outil Chimera v1.16.

En premier lieu, la structure secondaire de l'oligonucléotide RSE75, utilisé dans la majorité des tests de nucléases de cette étude, a été modélisée (**Figure A.I A**). Cette séquence d'ADN simple-brin de 75 nt provenant du génome de *N. meningitidis* contient une copie entière de la NTS ainsi que le site de clivage de SsnA, en plus d'extensions en 5' et en 3'. De façon intéressante, ce modèle révèle une structure en tige-boucle beaucoup plus longue que prévu, allant bien au-delà des régions palindromiques représentant la séquence répétée dRS3 (en cyan), et même des extensions non palindromiques composant la NTS (rouge). Selon ce modèle, le site de clivage de SsnA (rose) est sous forme double-brin, ce qui pourrait expliquer la perte d'activité de clivage lorsque la séquence en amont de la NTS est modifiée, puisque cette structure serait probablement perturbée. Ensuite, un modèle d'arrimage (*docking*) a été calculé avec SsnA de *N. meningitidis* (**Figure A.I B-C**). Dans ce dernier, la nucléase interagit avec plusieurs nucléotides allant du site de clivage jusqu'au début du palindrome, ce qui implique plusieurs acides aminés dont R27, K65 et K68. Lorsque le potentiel électrostatique de SsnA est illustré, on constate que cette protéine exhibe un sillon chargé positivement qui s'imbrique sur le ligand d'ADN au niveau du site de clivage (**Figure A.I C**). Ce sillon abrite vraisemblablement le site catalytique de la nucléase.

Ces prédictions bioinformatiques renforcent nos résultats expérimentaux montrant que SsnA se lie spécifiquement à la NTS et clive quelques nucléotides en amont de celle-ci, probablement sous forme monomérique. Cependant, la fiabilité des modèles obtenus ne permet pas de conclure sur la façon précise dont SsnA reconnaît et clive son substrat. Étant donné que cette étude révèle une famille entière de protéines jusque-là non caractérisées, il sera absolument essentiel de déterminer expérimentalement plusieurs structures en présence et en absence de substrat par cristallographie, afin d'associer cette famille à un mécanisme d'action précis.

ANNEXE II – BREVET

Le document ci-dessous représente un extrait du brevet pour lequel je figure comme co-inventeur, portant sur les résultats de ce projet de thèse. Ce dernier, déposé à l'international le 11 novembre 2022, porte le numéro PCT/CA2022/051668 et accessible en suivant le lien suivant :

<https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2023082011>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization

International Bureau

(43) International Publication Date
19 May 2023 (19.05.2023)



(10) International Publication Number
WO 2023/082011 A1

(51) International Patent Classification:

CI2N 9/22 (2006.01) CI2N 15/55 (2006.01)
CI2N 15/10 (2006.01) CI2P 19/30 (2006.01)

(21) International Application Number:

PCT/CA2022/051668

(22) International Filing Date:

11 November 2022 (11.11.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/263,896 11 November 2021 (11.11.2021) US

(71) Applicant: **INSTITUT NATIONAL DE LA RECHERCHE SCIENTIFIQUE** [CA/CA]; 490 rue de la Couronne, Québec, Québec G1K 9A9 (CA).

(72) Inventors: **VEYRIER, Frédéric**; 7075 rue de Lanaudière, Montréal, Québec H2E 1Y1 (CA). **CHENAL, Martin**; 9935 rue Lajeunesse, App. 213, Montréal, Québec H3L 2C9 (CA).

(74) Agent: **LAVERY, DE BILLY, LLP**; 1, Place Ville Marie, Bureau 4000, Montréal, Québec H3B 4M4 (CA).

(81) Designated States (unless otherwise indicated, for every kind of national protection available):

AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available):

ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: ENDONUCLEASES THAT SELECTIVELY CLEAVE SINGLE-STRANDED NUCLEIC ACIDS AND USES THEREOF

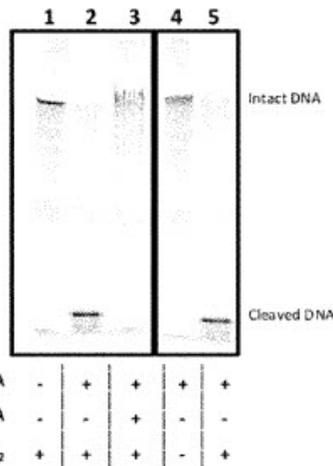


FIG. 2A

(57) Abstract: The present application relates to endonucleases specific for single-stranded nucleic acid molecules. These endonucleases having a length of 60 to 150 amino acids or less and comprises a single GIY-YIG domain. The present application also relates to compositions and cells comprising such endonucleases as well as to methods of cleaving single-stranded nucleic acid molecules comprising a nucleotide sequence recognized by the endonucleases.

[Continued on next page]

TITLE OF INVENTION

ENDONUCLEASES THAT SELECTIVELY CLEAVE SINGLE-STRANDED NUCLEIC ACIDS AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

5 The present application claims the benefit of U.S. provisional patent application serial No. 63/263,896, filed on November 11, 2021, which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

The present disclosure generally relates to the fields of enzymology and molecular biology, and more particularly to endonucleases.

10 BACKGROUND ART

Endonucleases are enzymes that cleave the phosphodiester bond within a polynucleotide chain. Two types of endonucleases are restriction nucleases and homing endonucleases.

15 Restriction endonucleases are enzymes that recognize a specific nucleotide sequence in a double-stranded nucleic acid called a restriction site. Upon binding to the restriction site, the restriction endonucleases cleave within or near the restriction site. These enzymes are routinely used for DNA modification in laboratories, such as for genetic engineering and molecular cloning. For example, they are used to assist insertion of genes into plasmid vectors, to distinguish gene alleles by specifically recognizing single base changes in DNA known as single-nucleotide polymorphisms (SNPs), to digest genomic DNA for gene analysis, and to insert nucleic acid molecules within the genome of an organism. Most of the known restriction enzymes recognize a restriction site, typically comprising from 4 to 8 nucleotides that are often palindromic, within a double-stranded DNA molecule, and produce a double-stranded cut in the DNA.

20 Homing endonucleases are double-stranded DNases that have large, asymmetric recognition sites (12-40 base pairs). However, unlike restriction endonucleases, homing endonucleases tolerate some sequence degeneracy within their recognition sequence, which means that single base changes do not abolish cleavage but reduce its efficiency to variable extents. As a result, their observed sequence specificity is typically in the range of 10-12 base pairs.

30 In contrast to restriction and homing endonucleases that cleaves double-stranded nucleic acids at specific sites, endonucleases that cleave single-stranded nucleic acid molecules are usually non-specific, *i.e.*, they do not recognize a specific sequence (*e.g.*, restriction site) within the single-stranded nucleic acid molecules but rather cleave at various sequences to degrade the nucleic acid molecules in several fragments.

Thus, there is a need for the identification of novel endonucleases that are able to recognize and cleave at specific sequences within single-stranded nucleic acid molecules such as single-stranded DNA molecules.

5 The present description refers to a number of documents, the content of which is herein incorporated by reference in their entirety.