



**HAL**  
open science

# L'îlot génomique pks chez *Escherichia coli* : structure-fonction de la protéine ClbP et études épidémiologiques

Damien Dubois

► **To cite this version:**

Damien Dubois. L'îlot génomique pks chez *Escherichia coli* : structure-fonction de la protéine ClbP et études épidémiologiques. Médecine humaine et pathologie. Université d'Auvergne - Clermont-Ferrand I, 2011. Français. NNT : 2011CLF1MM04 . tel-00612631

**HAL Id: tel-00612631**

**<https://theses.hal.science/tel-00612631>**

Submitted on 29 Jul 2011

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

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

UNIVERSITE BLAISE PASCAL

UNIVERSITE D'Auvergne

Année 2011

N° d'ordre

***ECOLE DOCTORALE  
DES SCIENCES DE LA VIE ET DE LA SANTE***

**THESE**

Présentée à l'Université d'Auvergne  
Pour l'obtention du grade de Docteur d'Université  
Spécialité : **Microbiologie**

par

**DUBOIS Damien**

**L'îlot génomique *pks* chez *Escherichia coli* :  
structure-fonction de la protéine ClbP et études épidémiologiques**

Soutenue le 11 mars 2011

**Président**

**F. DELBAC, Professeur, Université Blaise Pascal (Clermont-Ferrand)**

**Rapporteurs**

**C. GUILHOT, Directeur de recherche au CNRS, IPBS, Toulouse**

**W. SOUGAKOFF, Maître de Conférences, Université Pierre et Marie Curie (Paris VI)**

**Membres**

**E. OSWALD, Professeur, Université Paul Sabatier (Toulouse III)**

**R. BONNET, Professeur, Université d'Auvergne – Directeur de thèse**

JE2526, usc-INRA 2018

Evolution des bactéries pathogènes et susceptibilité génétique de l'hôte  
Clermont Université

28 place H. Dunant – 63001 Clermont-Ferrand Cedex



## DUBOIS Damien

L'îlot génomique *pks* chez *Escherichia coli* : structure-fonction de la protéine ClbP et études épidémiologiques

### Résumé

L'îlot génomique *pks* de *Escherichia coli* et d'autres *Enterobacteriaceae* code des synthèses de polycétides et de peptides non ribosomiaux qui permettent l'assemblage d'un composé hybride polycétide-peptide non ribosomal putative. Ce composé nommé colibactine induit des cassures double-brin de l'ADN des cellules eucaryotes.

La machinerie enzymatique codée par l'îlot *pks* comporte une protéine essentielle ClbP, atypique dans ce type d'îlot. Nous avons montré que ClbP possède une partie N-terminale catalytique et périplasmique, et une partie C-terminale associée à la membrane cytoplasmique. La structure cristalline de ClbP et des expériences de mutagenèse ont révélé un site actif à sérine et des caractéristiques structurales originales, qui sont compatibles avec une activité peptidase, confirmée par des analyses biochimiques. Dix homologues de ClbP ont été identifiés *in silico* dans des îlots génomiques de synthèses de peptides non ribosomiaux d'espèces bactériennes proches et éloignées. Tous les homologues testés ont présenté une promiscuité fonctionnelle avec ClbP. ClbP est donc le prototype d'une nouvelle sous-famille de peptidases, qui sont probablement impliquées dans la maturation de composés peptidiques non ribosomiaux.

Par ailleurs, nous avons réalisé deux études épidémiologiques sur la prévalence de l'îlot *pks* dans l'espèce *E. coli* dans deux contextes physiopathologiques, l'urosepsis et les cancers coliques et rectaux. L'îlot *pks* était significativement associé aux souches issues d'urosepsis comparé à des souches commensales, et aux souches issues de biopsies de tumeurs coliques comparé à des souches commensales ou issues de biopsies de tumeurs rectales, de diverticuloses et de lésions iléales de maladie de Crohn.

**Mots-clés** : colibactine, peptidase, peptides non ribosomiaux, urosepsis, cancer colorectal

### Summary

The *pks* genomic island of *Escherichia coli* and other *Enterobacteriaceae* encodes polyketide and nonribosomal peptide synthases that build a putative hybrid PK-NRP compound. This compound designated Colibactin induces DNA double-strand breaks in eukaryotic cells.

The *pks*-encoded enzymatic machinery comprises an essential protein ClbP, atypical for this type of genomic islands. We report that ClbP harbors a catalytic and periplasmic N-terminal part, and a C-terminal part associated to the cytoplasmic membrane. ClbP crystal structure and mutagenesis experiments revealed a serine-active site and original structural features, which are compatible with peptidase activity confirmed by biochemical assays. Ten ClbP homologs were identified *in silico* in NRPS-encoding genomic islands of close and distant-related bacterial species. All tested ClbP homologs showed functional promiscuity with ClbP. ClbP is therefore a prototype of a new subfamily of peptidases, which are probably involved for the maturation of NRP compounds.

Furthermore, we undertook two epidemiological studies on the prevalence of *pks* island in *E. coli* in two pathophysiology contexts; urosepsis and colorectal cancers. The *pks* island was significantly associated with urosepsis strains compared to commensal strains, and strains isolated from biopsies of colon tumors compared with commensal strains or strains isolated from biopsies of rectal tumors, diverticulosis and ileal lesions of Crohn disease.

**Key-words** : colibactin, peptidase, nonribosomal peptides, urosepsis, colorectal cancer

<b>Jury :</b>	Président :	Mr DELBAC Frédéric
	Rapporteurs :	Mr GUILHOT Christophe Mr SOUGAKOFF Wladimir
	Membres :	Mr OSWALD Eric Mr BONNET Richard – Directeur de thèse

**Date de soutenance** : 11 mars 2011

**Adresse de l'auteur** : 4 allée de Layrac – 31700 Blagnac





*A Frédérique,*

*A Anaé, puis-je te transmettre le sens de la curiosité*

*A Olivier, pour notre collaboration qui s'est arrêtée trop tôt*

*A tous ceux qui me sont chers*



## **Remerciements**

*Je remercie sincèrement,*

*Messieurs les Docteurs Guilhot et Sougakoff, d'avoir aimablement accepté d'être les rapporteurs de cette thèse. Je suis fier de vous compter parmi mes juges et vous assure ma reconnaissance et mon respect.*

*Monsieur le Professeur Delbac, d'avoir aimablement accepté de juger ce travail et de présider notre jury.*

*Monsieur le Professeur Oswald, pour votre collaboration, d'avoir aimablement accepté de juger cette thèse et de me permettre de poursuivre dans votre équipe.*

*Monsieur le Professeur Bonnet pour m'avoir accueilli comme Assistant et Doctorant dans son laboratoire, pour sa disponibilité, son écoute, sa confiance, sa patience, son soutien et ses conseils amicaux. Cela a été un plaisir de travailler dans ton équipe et sous ta direction. Trouve ici toute ma gratitude et mon respect sincère.*

*Je tiens à remercier ceux sans qui ce travail n'aurait pas été possible,*

*Rolande, pour son dynamisme, son optimisme et son aide,  
Marlène, pour sa gentillesse, ses conseils et son aide,  
cela a été un plaisir de travailler avec vous, au-delà de mes espérances en arrivant au fond du couloir,*

*Julien, pour les diverses collaborations et les nombreux échanges qui n'ont peut-être pas révolutionné la bactériologie mais qui ont fait avancer mes projets et mes idées, et « Viva San Diego ! »,*

*Nathalie, pour sa rigueur scientifique, son dynamisme et son franc-parler,  
Olivier, qui nous a quitté prématurément, pour notre collaboration amicale,  
Bernadette, pour sa collaboration et son enthousiasme pour la problématique,  
Pierre pour sa disponibilité, sa bonne humeur et nos échanges dans notre galère « colorectale »,*

*Marie-Agnès pour ses conseils et son aide,  
Antony, pour nos échanges et sa bonne humeur constante ; je te souhaite de réussir avec ClbP et ses semblables.*

*Marie-Hélène et Claudine pour leur gentillesse et leur disponibilité.*

*Je remercie également,*

*Emilie et Caroline, mes successives charmantes collègues de bureau,  
Tous ceux de l'équipe « Evolution des bactéries pathogènes et susceptibilité de l'hôte », en particulier Madame le Professeur Arlette Darfeuille-Michaud pour m'avoir accueilli dans son unité,  
Alexandra, pour son implication, et Sabah, pour son enthousiasme,  
Jean-Philippe pour ses conseils,*

*Tous ceux du laboratoire de bactériologie du CHU de Clermont-Ferrand pour leur accueil et leur bonne humeur,*

*Frédérique, mon épouse, pour sa compréhension et son soutien ces dernières années, pour avoir accepté les nombreuses absences,*

*Reçois ma profonde gratitude,*

*Anaé, ma fille, pour ton sourire constant, avec toi, c'est facile d'être un papa qui « cherche ».*



# SOMMAIRE

LISTE DES FIGURES.....	12
LISTE DES TABLEAUX .....	14
LISTE DES ABREVIATIONS .....	15
INTRODUCTION.....	17

## CHAPITRE I – Revue bibliographique

<b>I. LE CYCLE CELLULAIRE EUCARYOTE.....</b>	<b>23</b>
<b>A. STRATEGIE GENERALE DU CYCLE CELLULAIRE .....</b>	<b>23</b>
<b>B. LES PRINCIPAUX ACTEURS DE LA REGULATION DU CYCLE CELLULAIRE .....</b>	<b>24</b>
1. LES PROTEINES KINASES DEPENDANTES DES CYCLINES (CDK) .....	24
2. LES CYCLINES .....	24
3. LA PHOSPHORYLATION DES CDK .....	25
4. LES INHIBITEURS DE CDK .....	26
5. LA REGULATION DE LA LOCALISATION INTRACELLULAIRE DES CDK .....	26
<b>C. LES PRINCIPAUX EVENEMENTS DE REGULATION DU CYCLE .....</b>	<b>27</b>
<b>D. LES POINTS DE CONTROLE DE LA QUALITE DU CYCLE CELLULAIRE .....</b>	<b>28</b>
1. LA SIGNALISATION DES ALTERATIONS DE L'ADN.....	28
2. LES POINTS DE CONTROLE G1/S ET G2/M .....	30
3. LES POINTS DE CONTROLE EN PHASE S.....	32
4. LE POINT DE CONTROLE DU FUSEAU MITOTIQUE .....	32
<b>II. LES CYCLOMODULINES DE <i>ESCHERICHIA COLI</i> .....</b>	<b>33</b>
<b>A. LE FACTEUR CYTOTOXIQUE NECROSANT CNF .....</b>	<b>34</b>
1. LES CNF VARIANTS.....	34
2. STRUCTURE.....	34
3. SUPPORT GENETIQUE .....	34
4. MODE D'ACTION .....	35
<b>a) Sécrétion .....</b>	<b>35</b>
<b>b) Translocation dans les cellules eucaryotes.....</b>	<b>36</b>
<b>c) Modification de la cible moléculaire .....</b>	<b>36</b>
5. EXPRESSION PHENOTYPIQUE.....	38
6. DISTRIBUTION.....	39
7. ROLE DANS LE PROCESSUS INFECTIEUX.....	39
8. DES AGENTS CARCINOGENES POTENTIELS.....	40
<b>B. LA TOXINE DE DISTENSION CYTOLETHALE CDT .....</b>	<b>43</b>
1. LES VARIANTS CDT ET LEUR SUPPORT GENETIQUE .....	43
2. STRUCTURE DES CDT .....	43
3. MODE D'ACTION .....	44
<b>a) Sécrétion .....</b>	<b>44</b>

b)	<b>Translocation dans les cellules eucaryotes</b> .....	<b>44</b>
c)	<b>Altération de l'ADN et expression phénotypique</b> .....	<b>46</b>
4.	DISTRIBUTION ET ROLE DANS LE PROCESSUS INFECTIEUX.....	46
5.	DES AGENTS CARCINOGENES POTENTIELS.....	47
<b>C.</b>	<b>LE FACTEUR D'INHIBITION DU CYCLE Cif</b> .....	<b>48</b>
1.	DISTRIBUTION ET SUPPORT GENETIQUE.....	48
2.	STRUCTURE.....	48
3.	MODE D'ACTION ET EXPRESSION PHENOTYPIQUE.....	49
4.	UN AGENT CARCINOGENE POTENTIEL.....	49
<b>D.</b>	<b>LA COLIBACTINE</b> .....	<b>51</b>
1.	SUPPORT GENETIQUE.....	51
2.	TRANSCRIPTION DE L'ILOT <i>PKS</i> .....	51
3.	DISTRIBUTION DE L'ILOT <i>PKS</i> AU SEIN DES ENTEROBACTERIES.....	53
4.	ENVIRONNEMENT GENETIQUE DE L'ILOT <i>PKS</i> .....	53
5.	EXPRESSION PHENOTYPIQUE ET MODE D'ACTION.....	54
6.	UN AGENT CARCINOGENE POTENTIEL.....	56
<b>III.</b>	<b>LES POLYCETIDES ET LES PEPTIDES NON RIBOSOMAux</b> .....	<b>58</b>
<b>A.</b>	<b>GENERALITES</b> .....	<b>58</b>
1.	LES METABOLITES SECONDAIRES.....	58
2.	INTERET DE L'ETUDE DES COMPOSES D'ORIGINE NATURELLE.....	59
<b>B.</b>	<b>LA SYNTHÈSE DES COMPOSES PK ET NRP</b> .....	<b>61</b>
1.	LA LOGIQUE CHIMIQUE.....	62
a)	<b>L'activation des monomères</b> .....	<b>62</b>
b)	<b>L'élongation des oligomères</b> .....	<b>64</b>
c)	<b>La modification des oligomères</b> .....	<b>64</b>
2.	LES MACHINERIES ENZYMATIQUES.....	65
a)	<b>Les chaînes d'assemblage des composés PK</b> .....	<b>65</b>
b)	<b>Les chaînes d'assemblage des composés NRP</b> .....	<b>70</b>
c)	<b>Les chaînes d'assemblages hybrides PKS-NRPS</b> .....	<b>75</b>
d)	<b>Les modifications enzymatiques après la chaîne d'assemblage</b> .....	<b>75</b>
e)	<b>Les enzymes de réparation des chaînes d'assemblage</b> .....	<b>75</b>
f)	<b>Les domaines manquants et autres violations de colinéarité</b> .....	<b>75</b>
g)	<b>Les fragments de chaîne d'assemblage</b> .....	<b>76</b>
3.	LA TAILLE DES CHAINES D'ASSEMBLAGE.....	76
4.	STRUCTURE DES CHAINES D'ASSEMBLAGE PKS ET NRPS.....	77
a)	<b>L'architecture modulaire</b> .....	<b>77</b>
b)	<b>Reconnaissance entre les sous-unités protéiques</b> .....	<b>79</b>
c)	<b>Autres interactions</b> .....	<b>80</b>
<b>C.</b>	<b>FONCTIONS NATURELLES DES PK ET DES NRP</b> .....	<b>81</b>
1.	LA COLONISATION.....	81
a)	<b>Armes de destruction massive ou molécules de signalisation ?</b> .....	<b>81</b>
b)	<b>La protection contre les prédateurs eucaryotes</b> .....	<b>82</b>
c)	<b>La mobilité, l'adhérence et la formation de biofilm</b> .....	<b>82</b>
d)	<b>La chélation du fer et des autres ions métalliques</b> .....	<b>83</b>
2.	LA SYMBIOSE.....	85
3.	LA VIRULENCE.....	85

## **CHAPITRE II – Résultats expérimentaux**

I.	ARTICLE 1 – STRUCTURE-FONCTION DE LA PEPTIDASE CLBP .....	89
II.	ARTICLE 2 – LES CYCLOMODULINES DES <i>E. COLI</i> D'UROSEPSIS .....	133
III.	ARTICLE 3 - LES CYCLOMODULINES DES <i>E. COLI</i> DE TUMEURS .....	145

## **CHAPITRE III – Discussion et Perspectives**

I.	FONCTION BIOCHIMIQUE DE CLBP ET DE SES ORTHOLOGUES .....	169
A.	CLBP, UNE PEPTIDASE DES COMPOSES NRP .....	169
B.	RECHERCHE D'UN MOTIF CLIVE PAR CLBP ET SES ORTHOLOGUES .....	169
C.	LOCALISATION DE CLBP, DE SES ORTHOLOGUES ET DE LA COLIBACTINE .....	170
II.	ROLE BIOLOGIQUE DE CLBP ET DE SES ORTHOLOGUES .....	173
A.	CLBP ET LA SYNTHÈSE D'UN COMPOSE NRP BIOACTIF .....	173
B.	CLBP CLIVE UNE SEQUENCE D'ADRESSAGE ? .....	174
C.	CLBP ET PROTECTION DE LA CELLULE PRODUCTRICE .....	174
D.	CLBP ET PROTECTION CONTRE DES NRP EXOGENES.....	175
III.	DISTRIBUTION DES GENES CODANT LES PANRP .....	176
IV.	FONCTIONS BIOLOGIQUES DE LA COLIBACTINE .....	177
V.	ROLE DE LA COLIBACTINE DANS LE PROCESSUS INFECTIEUX .....	178
VI.	ROLE DE LA COLIBACTINE DANS LE CANCER COLORECTAL.....	179
VII.	AUTRES PERSPECTIVES RELATIVES A L'ETUDE DE L'ILOT <i>PKS</i> .....	185
VIII.	CONCLUSION .....	186

## **CHAPITRE IV – Autres publications**

I.	ARTICLE 4 – CTX-M ET VIRULENCE DE <i>E. COLI</i> K1 .....	189
II.	ARTICLE 5 – IDENTIFICATION DES <i>STAPHYLOCOCCUS</i> PAR SPECTROMÉTRIE DE MASSE .....	193

## **CHAPITRE V – Références bibliographiques**



## Liste des figures

<u>Figure 1 :</u>	Contribution des éléments génétiques mobiles à l'évolution des <i>E. coli</i> pathogènes	18
<u>Figure 2 :</u>	Cladogramme montrant les phylogroupes A, B1, B2, D et E à partir de 20 génomes de <i>E. coli</i> et de <i>Shigella</i>	18
<u>Figure 3 :</u>	Modèle classique du contrôle du cycle cellulaire chez les mammifères	23
<u>Figure 4 :</u>	Régulation de CDK1 par les kinases CDK7 et Wee1/Myt1, la phosphatase Cdc25c et l'inhibiteur de CDK p21 <sup>Waf1/Cip1</sup>	25
<u>Figure 5 :</u>	Exemples de substrats des CDK	27
<u>Figure 6 :</u>	Modèle des voies de signalisation des altérations de l'ADN	29
<u>Figure 7 :</u>	Le point de contrôle G <sub>1</sub> /S	31
<u>Figure 8 :</u>	Le point de contrôle G <sub>2</sub> /M	31
<u>Figure 9 :</u>	Organisation génétique et modèle d'expression de la région <i>hlyCABD-cnfl</i> du PAI-II de la souche UPEC J96	35
<u>Figure 10 :</u>	Modèle d'action des CNF sur les cellules hôtes	37
<u>Figure 11 :</u>	Organisation génétique et modèle d'action des CDT sur les cellules hôtes	45
<u>Figure 12 :</u>	Localisation génétique et modèle d'action de Cif sur les cellules hôtes	50
<u>Figure 13 :</u>	Carte schématique de l'îlot <i>pks</i>	52
<u>Figure 14 :</u>	Unités transcriptionnelles de l'îlot <i>pks</i>	52
<u>Figure 15 :</u>	Exposition de cellules HeLa à une infection transitoire à <i>E. coli pks+</i>	55
<u>Figure 16 :</u>	Modèle d'activation du point de contrôle G <sub>2</sub> /M dans les cellules HeLa exposées aux souches de <i>E. coli pks+</i>	55
<u>Figure 17 :</u>	Ponts anaphasiques et aberrations chromosomiques induits par une infection à <i>E. coli pks+</i>	57
<u>Figure 18 :</u>	Exemples de métabolites secondaires de type PK, NRP et hybrides utilisés en médecine	60
<u>Figure 19 :</u>	Utilisation de monomères thioesters par les PKS et les NRPS	63

<u>Figure 20 :</u>	Transfert et condensation des aminoacyl-thioesters lors de la formation d'un NRP	64
<u>Figure 21 :</u>	Domaines essentiels constituant les modules des PKS et des NRPS	65
<u>Figure 22 :</u>	Comparaison des PKS de type I et de type II	66
<u>Figure 23 :</u>	Modification post-traductionnelle des domaines de Thiolation T par les PhosphoPantéthéinyl-Transférases	66
<u>Figure 24 :</u>	Activités des domaines AT et KS des PKS	67
<u>Figure 25 :</u>	Elongation de la chaîne dans les systèmes PKS de type I	67
<u>Figure 26 :</u>	Transformation du carbone $\beta$ dans les PKS	69
<u>Figure 27 :</u>	Organisation des PKS de type I ou modulaires synthétisant le 6-dEB	69
<u>Figure 28 :</u>	Fonctionnement du domaine T et activités des domaines A et C des NRPS	71
<u>Figure 29 :</u>	Les quatre processus de macrocyclisation des NRP	72
<u>Figure 30 :</u>	Organisation des NRPS linéaires synthétisant la tyrocidine	74
<u>Figure 31 :</u>	Modèles structuraux des modules de PKS de type I	78
<u>Figure 32 :</u>	Structure cristalline du module de la dernière NRPS de la surfactine	79
<u>Figure 33 :</u>	Domaines d'ancrage entre les sous-unités protéiques des systèmes de biosynthèse mixtes PKS-NRPS	80
<u>Figure 34 :</u>	Acquisition du fer par <i>E. coli</i> chez les mammifères	84
<u>Figure 35 :</u>	Exemples de composés NRPK impliqués dans la colonisation et la survie bactériennes	84
<u>Figure 36 :</u>	Arbre phylogénétique des sous-types de domaines de condensation C de NRPS	171
<u>Figure 37 :</u>	Modèle hypothétique d'initiation de la tumorigenèse colique par la colibactine	183

## Liste des tableaux

<u>Tableau 1</u> :	Exemples de cyclomodulines bactériennes	33
<u>Tableau 2</u> :	Comparaison entre les effets cellulaires de CNF1 de <i>E. coli</i> et des facteurs de virulence majeurs (VacA et CagA) de <i>H. pylori</i> .	42
<u>Tableau 3</u> :	Caractéristiques des métabolismes primaires et secondaires	58

## Liste des abréviations

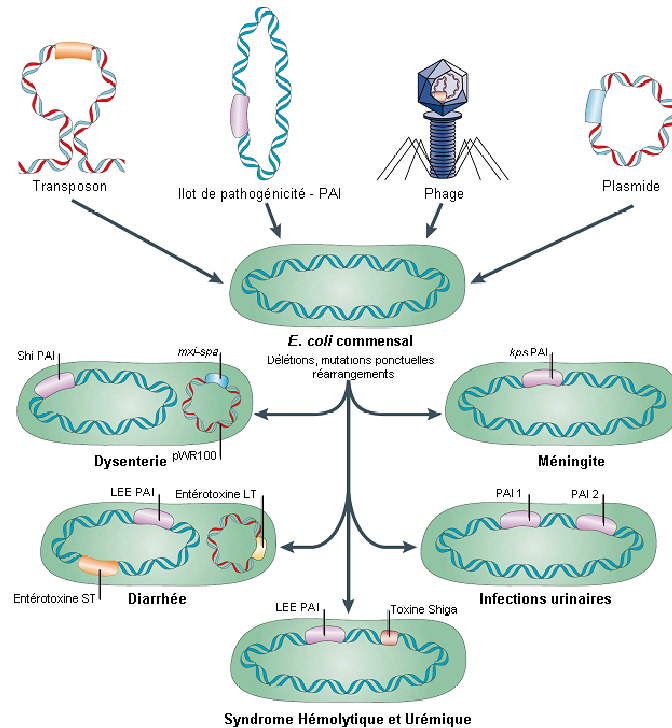
6-dEB	précurseur de l'érythromycine	ExPEC	<i>E. coli</i> à l'origine de pathologies extra-intestinales
A	domaine d'Adénylation	FAS	Synthase d'acide gras
ABC	ATP Binding Cassette	GFP	Green Fluorescent Protein
ACP	Acyl Carrier Protein	KR	Ketoacyl Réductase
ADN	Acide DésoxyriboNucléique	KS	Ketoacyl Synthase
AIEC	<i>E. coli</i> Adhérent et Invasif	LEE	Locus d'effacement des entérocytes
AMP	Adénosine MonoPhosphate	MATE	Multidrug And Toxic compound Extrusion
AmpC	$\beta$ -lactamase de classe C/de type AmpC	NF-kB	Nuclear Factor-kappaB
APC	<i>adenomatous polyposis coli</i>	NRP	Peptide Non Ribosomal
AT	AcylTransférase	NRPK	Ensemble des PK, NRP et PK-NRP
ATM	Ataxia Telangiectasia Mutated protein	NRPS	Synthase de Peptides Non Ribosomaux
ATP	Adénosine TriPhosphate	NTEC	Necrotoxic <i>E. coli</i>
ATR	ATM and Rad3-related protein	OX	Domaine d'Oxidation
ATRIP	ATR Interacting Protein	PAI	Ilot de pathogénicité
C	domaine de Condensation	PaNRP	Peptidases associées aux NRP
CagA	Cytotoxin-associated protein A	PCP	Peptide Carrier Protein
CAK	Cdk-Activating Kinase	PDIM	Phthiocérol dimycocérosates
COX2	Cyclo-OXYgénase2	PK	Polycétide
CCR	Cancer ColoRectal	PK-NRP	Polycétide-Peptide non ribosomal
Cdc42	Rho GTPase	PKS	Synthase de Polycétides
CDK	Cyclin-Dependent Kinase	PKS-NRPS	Synthase de PK-NRP
CDT	Cytolethal Distending Toxin	pRb	Protéine du gène suppresseur de tumeur du rétinoblastome
Chk1	Checkpoint kinase 1	PPTase	PhosphoPantéthéinyl-Transférase
Chk2	Checkpoint kinase 2	Rac	Rho GTPase
Cif	Cycle Inhibition Factor	Rho-GAP	Rho-GTPase Activating Protein
CKI	Cycline-dependent Kinase Inhibitor	RhoA	GTPase
CM	Cyclomoduline	ROS	Espèces réactives de l'oxygène
CNF	Cytotoxic Necrotizing Factor	SST3	Système de Sécrétion de type III
CoA	Coenzyme A	T	domaine de Thiolation
Cy	domaine de Cyclodéshydratation	TE	domaine ThioEstérase
DH	Déshydratase	UPEC	Urinary Pathogenic <i>E. coli</i>
DNT	Dermonecrotic Toxin	VacA	Vacuolating cytotoxin
<i>E. coli</i>	<i>Escherichia coli</i>	ZmA	Zwittermicine A
EHEC	<i>E. coli</i> EntéroHémorragique		
EPEC	<i>E. coli</i> EntéroPathogène		
ER	Enoyl Réductase		



## Introduction

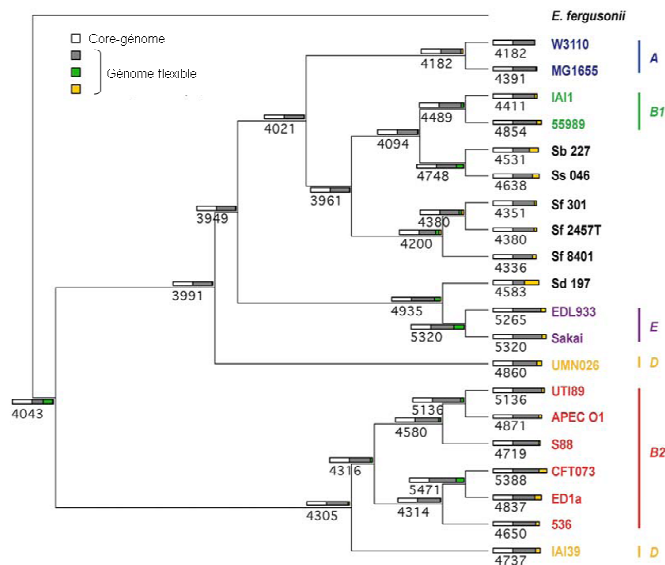
Découvert en 1885 par Théodore Escherich dans des selles de nourrissons, *Escherichia coli* fait partie du microbiote intestinal, écosystème composé de plus de 5600 taxons différents [1]. Il colonise généralement le tractus gastro-intestinal des nourrissons dans les heures qui suivent la naissance, puis établit une relation symbiotique qui perdure pendant des décennies dans la couche muqueuse du côlon [2]. Compétiteur très performant sur ce site surpeuplé, *E. coli* est l'anaérobie facultatif le plus abondant du microbiote intestinal humain, qui assure des fonctions indispensables, comme la modulation du système immunitaire, la dégradation des aliments, la synthèse de vitamines et la protection vis-à-vis de microorganismes pathogènes. De plus, il participe au développement, à la différenciation, à l'homéostasie des muqueuses et jouerait un rôle dans l'étiologie du cancer sporadique du côlon [3-5].

*E. coli* peut devenir pathogène lors de l'affaiblissement des défenses de l'hôte et/ou suite à l'acquisition d'attributs de virulence. Ces facteurs de virulence confèrent une capacité à coloniser de nouvelles niches écologiques et sont codés sur des éléments génétiques mobiles ou anciennement mobiles (**figure 1**) [2]. Ainsi, le génome de *E. coli* comprend le core-génome constitué d'environ 2000 gènes essentiels aux fonctions cellulaires de base (par exemple la traduction et le métabolisme) ainsi qu'un ensemble génique flexible. Ce dernier est composé de 2068 à 3379 gènes par souche, issus d'un pool d'environ 15800 gènes codant des fonctions optionnelles parfois impliquées dans la pathogénèse [6, 7]. Par ailleurs, l'analyse du core-génome de *E. coli* a révélé cinq groupes phylogénétiques A, B1, B2, D et E (**figure 2**) [8-15]. La distribution des facteurs de virulence varie selon les phylogroupes, indiquant un rôle du « fond » génétique dans l'acquisition, le maintien et l'expression des gènes de virulence [16-19].



**Figure 1 - Contribution des éléments génétiques mobiles à l'évolution des *E. coli* pathogènes.** D'après Kaper *et al* [2].

Les facteurs de virulence de *E. coli* peuvent être codés par plusieurs éléments génétiques mobiles, incluant les transposons, les plasmides, les bactériophages et les îlots de pathogénicité (PAI). Les *E. coli* commensaux peuvent aussi subir des délétions, des mutations ponctuelles et d'autres réarrangements d'ADN. Toutes ces modifications génétiques peuvent engendrer des *E. coli* pouvant infecter l'hôte sain.



**Figure 2 - Cladogramme montrant les phylogroupes A, B1, B2, D et E à partir de 20 génomes de *E. coli* et de *Shigella*.** D'après Touchon *et al* [7].

Les groupes phylogénétiques majeurs sont indiqués par les lignes verticales. Chaque souche et nœud interne sont étiquetés avec le nombre de gènes présents. *Shigella*, qui a été élevé au rang de genre avec quatre espèces (*S. dysenteriae* (Sd), *S. flexneri* (Sf), *S. boydii* (Sb), *S. sonnei* (Ss)), en se basant sur leur capacité à générer une diarrhée invasive chez l'homme et sur leurs caractéristiques biochimiques, appartient en fait à l'espèce *E. coli*.

Les souches de *E. coli* pathogènes sont responsables d'atteintes intestinales ou extra-intestinales et sont classées en pathotypes selon les manifestations cliniques engendrées, les facteurs de virulence hébergés et certaines interactions cellulaires [2]. Parmi les agents pathogènes intestinaux responsables de diarrhées, six catégories sont bien définies : les *E. coli* entéro-pathogènes (EPEC), les *E. coli* entéro-hémorragiques (EHEC), les *E. coli* entéro-toxinogènes (ETEC), les *E. coli* entéro-agrégatifs (EAEC), les *E. coli* entéro-invasifs (EIEC) et les *E. coli* à adhésion diffuse (DAEC). D'autres pathotypes intestinaux ont été décrits, notamment les *E. coli* adhérent et invasif (AIEC) associés à la maladie de Crohn, qui se répliquent dans les cellules épithéliales, les macrophages et induisent la sécrétion de cytokines pro-inflammatoires [20]. Les *E. coli* à l'origine de pathologies extra-intestinales (ExPEC) sont encore mal définis. Ils appartiennent majoritairement aux phylogroupes B2 et dans une moindre mesure D. Ils ont acquis la capacité à dépasser les défenses immunitaires de leur hôte et à se propager dans l'organisme. Ils sont responsables de méningites, d'abcès, de péritonites, de septicémies et surtout d'infections du tractus urinaire. Trois pathotypes d'ExPEC sont couramment individualisés : les *E. coli* uropathogènes (« Urinary Pathogenic *E. coli* » ou UPEC) responsables d'infections urinaires, les *E. coli* associés aux méningites néonatales (« Neonatal Meningitis *E. coli* » ou NMEC) et les *E. coli* responsables d'infections systémiques en élevage avicole (« Avian Pathogenic *E. coli* » ou APEC) [21, 22].

Parmi les facteurs de virulence les plus importants de *E. coli*, on retrouve des adhésines, des invasines, des sidérophores, le flagelle, les antigènes O, la capsule et des toxines. Ces facteurs contribuent à la colonisation et à la pathogénèse en bloquant ou en détournant des processus cellulaires normaux, comme l'assemblage du cytosquelette ou l'apoptose [2, 21]. Récemment, le groupe du Pr. Eric Oswald a défini un nouveau type d'effecteurs bactériens activant ou inhibant le cycle cellulaire eucaryote. Ces toxines, appelées cyclomodulines, sont identifiées dans un nombre croissant de bactéries [23]. Quatre types ont été identifiées chez *E. coli* : CNF, CDT, Cif et une toxine récemment décrite, la colibactine [24].

La structure de la colibactine est inconnue à ce jour. Cependant, l'ilot génomique responsable de sa production, appelé *pks*, code des synthèses de peptides non ribosomiaux (« Non-Ribosomal Peptide Synthases » ou NRPS), des synthèses de polycétides (« PolyKetide Synthases » ou PKS) et des enzymes hydrides PKS-NRPS. Ces enzymes sont connues pour être codées par des blocs de gènes et pour constituer des chaînes métaboliques



synthétisant des composés chimiques de type polycétide-peptide non ribosomal (« PolyKetide-Non-Ribosomal Peptide » ou PK-NRP), souvent toxiques et parfois utilisés en thérapeutique [25]. Le rôle de la colibactine chez *E. coli* reste cependant à éclaircir. Par ailleurs, l'îlot génomique *pks* produit une protéine essentielle appelée ClbP, qui est inhabituelle dans ce type d'îlot et dont le rôle reste également à préciser.

L'objectif de ce travail était de caractériser sur le plan structural et fonctionnel la protéine ClbP et d'explorer le rôle de l'îlot *pks* au travers d'une étude épidémiologique des cyclomodulines de *E. coli*. Les résultats sont présentés sur articles, précédés par une revue bibliographique portant sur le cycle cellulaire eucaryote, les cyclomodulines de *E. coli*, et les composés PK-NRP. Les résultats sont ensuite discutés. Enfin, les publications réalisées en marge de ce travail sont annexées à la fin de ce manuscrit.

**Chapitre I**

**Revue**

**bibliographique**

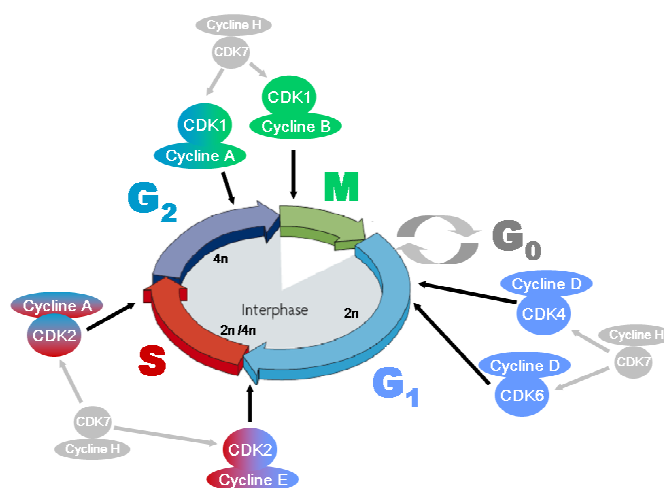


# I. Le cycle cellulaire eucaryote

Le cycle cellulaire est un ensemble d'événements conservés et ordonnés, qui régissent la croissance et la division cellulaire [26].

## A. Stratégie générale du cycle cellulaire

Lorsqu'elles ne se divisent pas, les cellules sont dites quiescentes ou en phase  $G_0$ . Elles représentent une partie importante des cellules du corps humain. Le passage de ces cellules dans une phase de prolifération nécessite des signaux mitogènes qui déclenchent le cycle cellulaire. Ce dernier comprend une période de croissance cellulaire appelée interphase, suivie de la mitose (**figure 3**) [26, 27]. L'interphase comprend trois phases : (i) la phase  $G_1$  (« Gap 1») qui prépare la synthèse de l'ADN et assure le passage d'un point de non-retour rendant le processus de division indépendant des signaux mitogènes, (ii) la phase S (« Synthesis») de réplication de l'ADN et (iii) la phase  $G_2$  (« Gap 2») de préparation à la mitose. La phase M de mitose correspond à la ségrégation des chromosomes répliqués et à la répartition des organites dans les cellules filles. Elle est classiquement subdivisée en 4 phases appelées prophase, métaphase, anaphase et télophase. Ces processus sont étroitement régulés par des mécanismes qui permettent ou limitent sa progression afin d'assurer l'intégrité des cellules filles.



**Figure 3 - Modèle classique du contrôle du cycle cellulaire chez les mammifères.**

Modifié d'après Malumbres et Barbacid, et Vermeulen *et al* [28, 29].

Les cyclines D se lient à CDK4 puis CDK6 pour former des complexes essentiels pour l'entrée en phase  $G_1$ . Le complexe CDK2-cycline E déclenche la progression de la phase  $G_1$  en phase S. Le complexe CDK2-cycline A régule la phase S et son achèvement, relayé par le complexe CDK1-cycline A en fin de phase  $G_2$ . Le complexe CDK1-cycline B déclenche ensuite la mitose. Le complexe CDK7-cycline H (complexe kinase activatrice) et les protéines Cdc 25 (phosphatases activatrices) activent les différentes CDKs.

## **B. Les principaux acteurs de la régulation du cycle cellulaire**

La régulation du cycle cellulaire est assurée par les variations de l'activité des kinases dépendantes des cyclines (« Cyclin-dependent Kinase » ou CDK). Elles-mêmes sont modulées à quatre niveaux par : (i) l'association à des protéines activatrices, les cyclines, (ii) des modifications de leur état de phosphorylation, (iii) la fixation de protéines inhibitrices, et (iv) la régulation de leur localisation cellulaire.

### **1. Les protéines Kinases Dépendantes des Cyclines (CDK)**

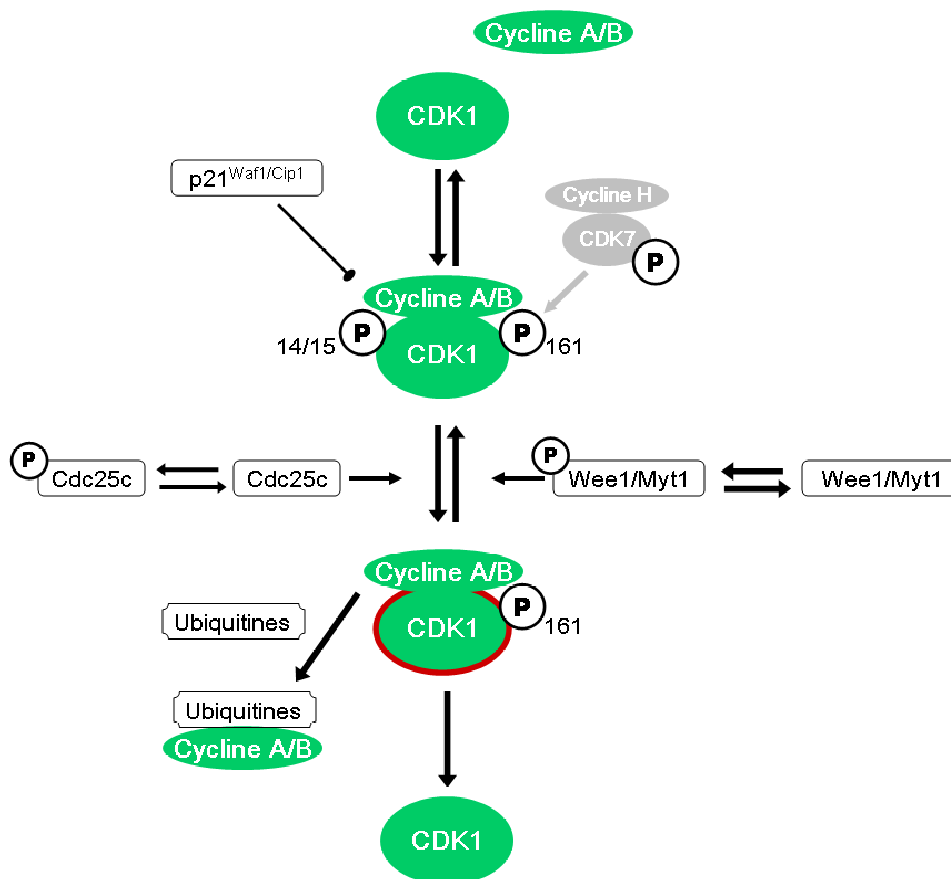
Les CDK forment une famille de sérine/thréonine kinases essentielles à la progression du cycle cellulaire [30]. Globalement, l'activité CDK augmente à partir de la phase G<sub>1</sub> pour culminer et s'effondrer lors de la mitose [31]. Leur rôle est de phosphoryler, sur des sites consensus, des protéines clés de la progression du cycle cellulaire. Leur niveau de production dans la cellule est stable au cours du cycle cellulaire. Cependant, elles sont activées à des points spécifiques du cycle cellulaire. Six CDK sont actives au cours du cycle cellulaire ; CDK4, CDK6 et CDK2 interviennent pendant la phase G<sub>1</sub>, CDK2 en phase S, CDK1 en phase G<sub>2</sub> et M, et CDK7 à toutes les phases du cycle (**figure 3**) [28, 31].

### **2. Les Cyclines**

Les CDK sont actives sous forme de complexe avec des protéines régulatrices spécifiques, les cyclines [32, 33]. Vingt-cinq loci codent les cyclines chez l'homme [28]. Dix cyclines appartenant à quatre classes différentes (A, B, D et E) sont nécessaires au cycle cellulaire (**figure 3**). Leur concentration varie en fonction des phases du cycle cellulaire. Elles portent des séquences d'ubiquitination permettant une dégradation efficacement par le protéasome [34]. Les cyclines D et E sont principalement exprimées lors des phases G<sub>1</sub> et sont dégradées en phase S [31, 35, 36]. Les cyclines A et B sont majoritairement synthétisées lors de la phase S et G<sub>2</sub> respectivement et sont dégradées au début et en fin de mitose respectivement [31, 34, 37].

### 3. La phosphorylation des CDK

L'activité CDK est aussi régulée par des phosphorylations activatrices ou inhibitrices de résidus thréonine et tyrosine [29, 38, 39]. Elles seront actives ou non selon le profil des phosphorylations portées. Le complexe CDK7-Cycline H, appelé aussi CAK (« Cdk-Activating Kinase ») active les principaux complexes CDK-cycline par phosphorylation (**figure 3**) [40]. L'activation passe également par la déphosphorylation d'autres sites par les trois isoformes (a, b et c) des phosphatases Cdc25, qui jouent un rôle essentiel dans les transitions G<sub>1</sub>/S et G<sub>2</sub>/M [41]. Par exemple, les kinases Wee1 et Myt1 agissent sur les résidus 14 et/ou 15 pour maintenir inactive CDK1 [42, 43]. Pour présenter une activité kinase, CDK1 doit d'une part, être phosphorylée sur le résidu 161 par CAK et d'autre part, être déphosphorylée sur les résidus 14 et/ou 15 par Cdc25c (**figure 4**).



**Figure 4 - Régulation de CDK1 par les kinases CDK7 et Wee1/Myt1, la phosphatase Cdc25c et l'inhibiteur de CDK p21<sup>Waf1/Cip1</sup>.** La forme active de CDK1 est entourée en rouge. D'après Vermeulen *et al* [29].

## 4. Les inhibiteurs de CDK

L'activité des CDK peut être bloquée par des protéines inhibitrices du cycle cellulaire, appelées inhibiteurs de CDK (« Cycline-dependent Kinase Inhibitor » ou CKI). Elles se lient aux CDK ou aux complexes CDK-cycline. Deux familles de CKI ont été décrites, la famille INK4 et la famille Cip/Kip [44].

La famille INK4 (Inhibiteur de CDK4) comprend les protéines p15<sup>INK4b</sup>, p16<sup>INK4a</sup>, p18<sup>INK4c</sup> et p19<sup>INK4d</sup>. Elles inactivent spécifiquement les CDK de la phase G<sub>1</sub>, CDK4 et CDK6 [45].

La deuxième famille d'inhibiteurs, la famille Cip/Kip comprend les protéines p21<sup>Waf1/Cip1</sup>, p27<sup>Kip1</sup> et p57<sup>Kip2</sup> [46-48]. Ces inhibiteurs inactivent les complexes CDK-cycline de la phase G<sub>1</sub> et dans une moindre mesure, le complexe CDK1-cycline B de la phase M (**figure 4**) [49]. La protéine p21<sup>Waf1/Cip1</sup> induite par p53 inhibe également la synthèse de l'ADN en se liant à l'antigène nucléaire de prolifération cellulaire PCNA (« Proliferation Cell Nuclear Antigen ») [50].

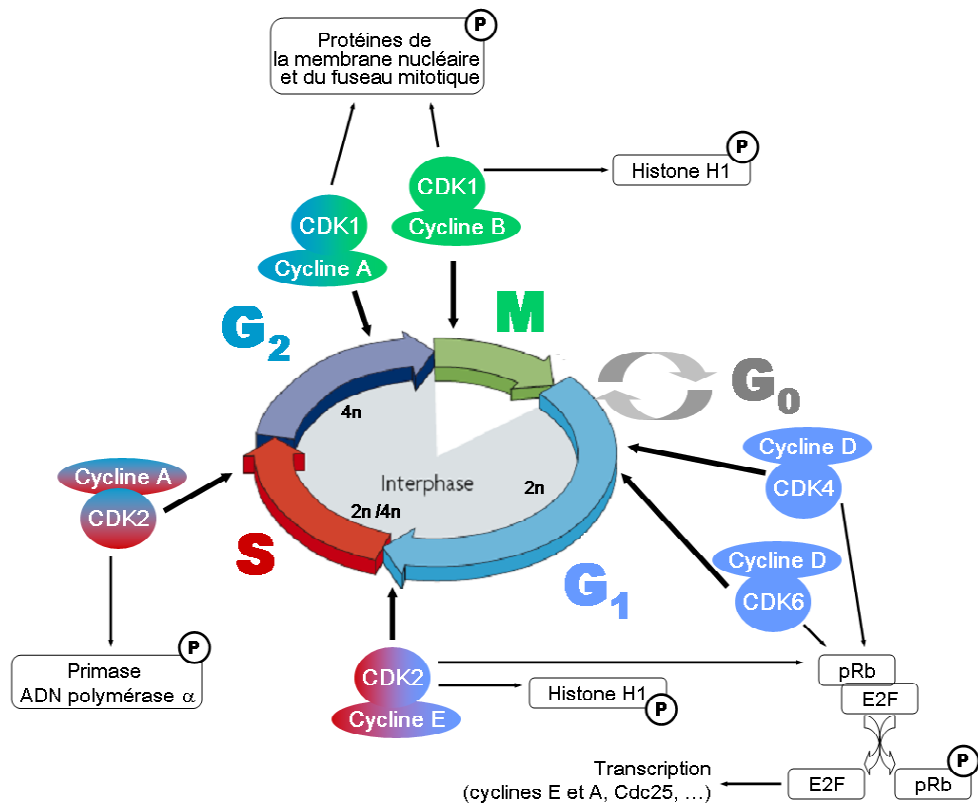
## 5. La régulation de la localisation intracellulaire des CDK

La localisation intracellulaire des protéines effectrices du cycle cellulaire contribue également à sa régulation. La cycline B contient un signal d'exclusion nucléaire, qui assure une exportation active de la protéine jusqu'au début de la prophase [51]. Les kinases Wee1 et Myt1 qui inactivent CDK1 sont situées respectivement dans le noyau et au niveau du réticulum endoplasmique, et protègent la cellule d'une mitose prématurée [52, 53]. Le groupe de protéines 14-3-3 régule le trafic intracellulaire de différentes protéines dont la phosphatase Cdc25c. Elles maintiennent cette dernière dans le cytoplasme pendant l'interphase afin d'éviter l'activation de CDK1 [54, 55].

### C. Les principaux évènements de régulation du cycle

En phase  $G_1$  précoce, les complexes CDK4/6-cycline D phosphorylent différentes protéines comme la protéine du gène suppresseur de tumeur du rétinoblastome (pRb), qui est complexée à l'histone déacétylase et aux facteurs de transcription de la famille E2F [56, 57]. Cette phosphorylation libère les facteurs E2F qui régulent positivement la transcription des gènes nécessaires pour la progression en phase S, tels que ceux de la cycline E et des Cdc25 (**figure 5**) [58].

Au cours de la transition de phase  $G_1/S$ , le complexe CDK2-cycline E phosphoryle pRb sur des sites additionnels afin d'assurer la libération complète des facteurs E2F, qui induisent la transcription du gène codant la cycline A, nécessaire à la phase S [58-60]. Le complexe CDK2-cycline E phosphoryle également l'histone H1 et ainsi participe à la condensation des chromosomes nécessaire à la réplication de l'ADN (**figure 5**) [61, 62].



**Figure 5 - Exemples de substrats des CDK.** Modifié d'après Vermeulen *et al* [29].



En phase S, les CDK2 et CDK1 complexées à la cycline A régulent notamment l'initiation de la réplication de l'ADN par la phosphorylation de la primase ADN polymérase alpha (**figure 5**) [63].

Afin d'assurer la mitose, CDK1 complexée à la cycline A en phase G<sub>2</sub> et à la cycline B1 en fin de phase G<sub>2</sub> et en début de mitose, phosphoryle de nombreuses protéines telles que l'histone H1 et des protéines nécessaires au désassemblage de la membrane nucléaire et à l'assemblage du fuseau mitotique (**figure 5**) [64].

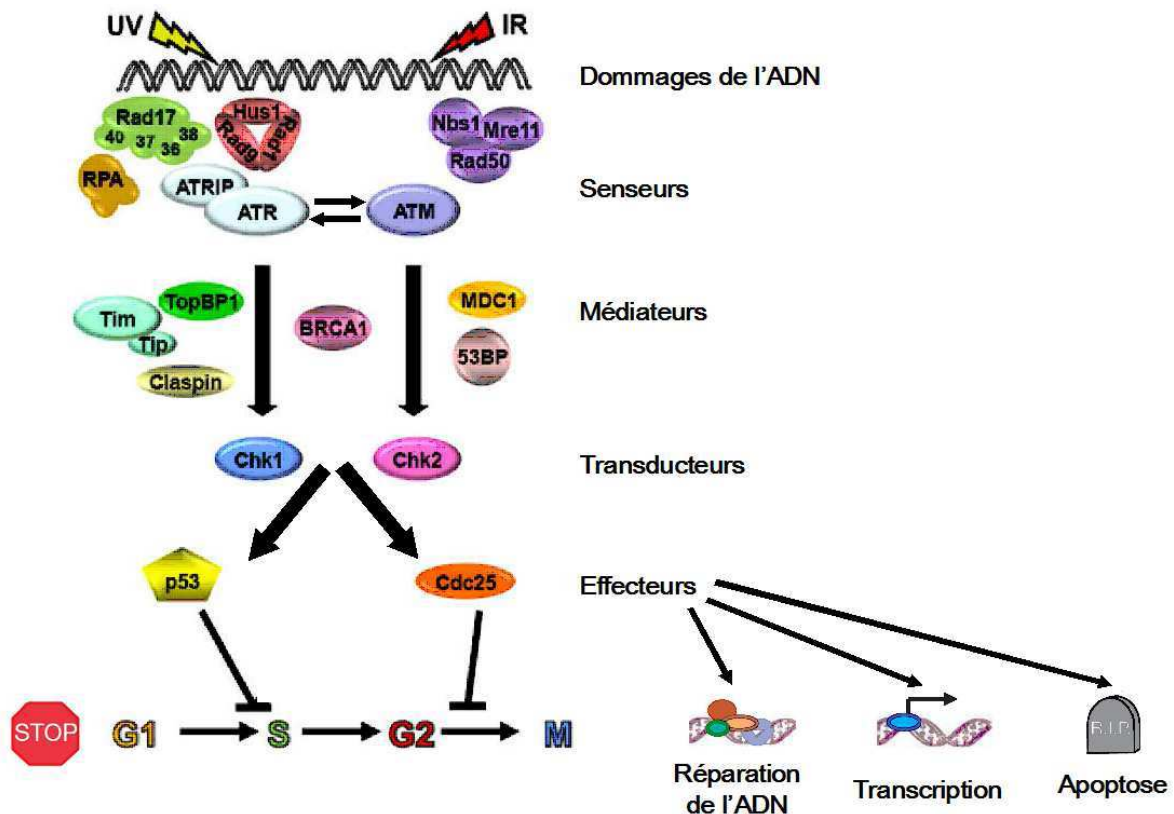
## ***D. Les points de contrôle de la qualité du cycle cellulaire***

Pendant le cycle cellulaire, l'ADN doit être parfaitement conservé, dupliqué et réparti entre les cellules filles. Pour assurer le bon déroulement de ces phénomènes, des points de contrôle sont positionnés aux points critiques du cycle cellulaire : lors des transitions G<sub>1</sub>/S et G<sub>2</sub>/M, et durant la phase S et la mitose.

### **1. La signalisation des altérations de l'ADN**

Les points de contrôle sont activés par des voies de détection spécifiques des dommages de l'ADN faisant intervenir quatre niveaux de signalisation (les protéines senseurs, médiatrices, transductrices et effectrices) et arrêtent le cycle cellulaire afin de permettre la réparation de l'ADN avant d'avancer dans le cycle ou d'activer la mort cellulaire par apoptose en cas de dommages irréversibles (**figure 6**) [65-71].

La phosphorylation de l'histone H2AX par le senseur ATM (« Ataxia Telangiectasia Mutated protein ») est un régulateur maître du recrutement des protéines de réparation aux sites de coupure double-brin de l'ADN. Cette phosphorylation qui s'étend jusqu'à un mégabase du site de coupure et permet la formation de foci de réparation aisément visualisables en imagerie de fluorescence [72]. De plus, la dynamique des protéines de signalisation et de réparation des altérations de l'ADN est assurée par un processus d'ubiquitination influençant l'efficacité et la spécificité de la réparation de l'ADN [73].



**Figure 6 - Modèle des voies de signalisation des altérations de l'ADN.**

Modifié d'après Sancar *et al*, et Zhou et Elledge [65, 66].

Dans les cellules de mammifères, la signalisation des points de contrôle est initiée notamment par les deux phospho-inositide kinases senseurs : ATM et ATR (« ATM and Rad3-related protein»), en complexe avec leurs partenaires de liaison respectifs MRN (Mre11-Rad50-Nbs1) et ATRIP (« ATR interacting protein »).

Les cassures double-brin de l'ADN générées par les agents génotoxiques tels que les radiations ionisantes sont perçues par le complexe MRN, qui recrute alors ATM aux sites de dommage et favorise l'activation de cette kinase [70]. Lorsque les fourches de réplication arrivent aux cassures double-brin de l'ADN au cours des phases S et G<sub>2</sub> du cycle cellulaire, ATR est activée, en aval d'ATM [67]. Inversement, lors d'arrêt des fourches de réplication ou d'exposition à une irradiation UV (cassures simple-brin) ATR est activée en amont d'ATM : les cassures simple-brin non réparées au cours des phases G<sub>1</sub> et S génèrent lors du passage des fourches de réplication, des cassures double-brin qui activent ATM [71].

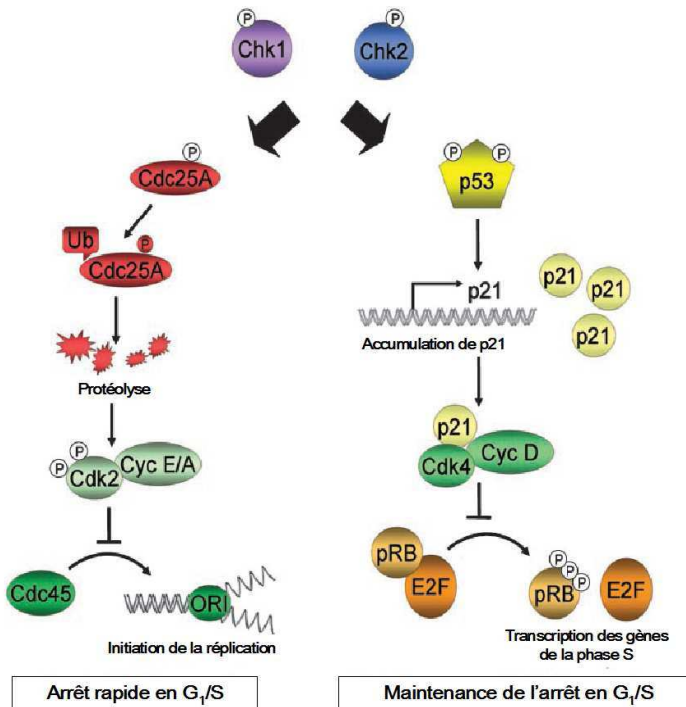
Les complexes senseurs d'ATM et d'ATR, avec l'aide de médiateurs, activent respectivement les kinases transductrices Chk1 et Chk2 [68]. Celles-ci inhibent ou activent en retour les effecteurs impliqués dans la réparation de l'ADN, la transcription, l'apoptose ou l'arrêt du cycle cellulaire. p53 et les Cdc25 participent directement à l'inhibition des transitions de phase G<sub>1</sub>/S et G<sub>2</sub>/M et la progression de la phase S. Selon le contexte, certaines molécules peuvent avoir des fonctions multiples dans cette voie de transduction du signal.

## 2. Les points de contrôle G<sub>1</sub>/S et G<sub>2</sub>/M

Selon la nature des lésions et la phase du cycle cellulaire, les kinases transductrices Chk1 et Chk2 vont réguler les protéines effectrices dont p53 et Cdc25, qui participent directement à l'inhibition des transitions G<sub>1</sub>/S (**figures 7**) et G<sub>2</sub>/M [65, 74-77].

Lors de dommages de l'ADN en phase G<sub>1</sub>, les transducteurs Chk1 et Chk2 activés phosphorylent Cdc25a, favorisant sa liaison aux protéines 14-3-3 et son exclusion nucléaire consécutive, ainsi que sa dégradation via la voie de l'ubiquitine [76]. CDK2 phosphorylée et inactivée s'accumule et ne peut pas phosphoryler Cdc45 pour initier la réplication de l'ADN, induisant un arrêt rapide du cycle cellulaire. Contrairement aux autres points de contrôle, le maintien de l'arrêt du cycle cellulaire au point de contrôle G<sub>1</sub>/S est dépendant de la protéine p53 [77]. p53 qui est phosphorylée par les senseurs ATM/ATR et par les transducteurs Chk1/Chk2, induit la transcription de p21<sup>Waf-1/Cip1</sup>, qui se lie au complexe CDK4-cycline D (ou Cdk4-Cyc D), l'empêchant de phosphoryler pRb [78]. pRb ne peut pas libérer le facteur de transcription E2F nécessaire à la transcription des gènes de la phase S. p21<sup>Waf-1/Cip1</sup> inactive également le complexe CDK2-cycline E, assurant le maintien au point de contrôle G<sub>1</sub>/S.

Lors de dommages de l'ADN en phase G<sub>2</sub>, les cellules sont capables d'initier un arrêt du cycle cellulaire en présence ou en absence de p53. Les transducteurs Chk1 et Chk2 activés phosphorylent Cdc25a, qui ne peut plus activer le complexe CDK1-cycline B (Cdc2-cycline B), bloquant l'entrée en mitose [75]. De même, Chk1 et Chk2 activés phosphorylent Wee1 qui effectue une phosphorylation inhibitrice de CDK1 (**figures 8**). L'augmentation de p53 dépendante des dommages de l'ADN induit comme au cours du point de contrôle G<sub>1</sub>/S, la transcription accrue de p21<sup>Waf-1/Cip1</sup> et de protéines 14-3-3. p21<sup>Waf-1/Cip1</sup> se lie au complexe CDK1-cycline B, empêchant son activité. L'augmentation de la liaison de la cycline B aux protéines 14-3-3 l'exclut activement du noyau [74, 79].



**Figure 7 - Le point de contrôle G<sub>1</sub>/S.** D'après Sancar *et al.* [65].

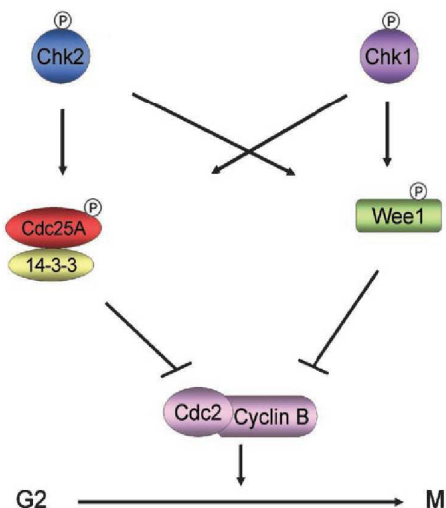
Les transducteurs Chk1 et Chk2 activés phosphorylent Cdc25a, favorisant sa liaison aux protéines 14-3-3 et son exclusion nucléaire consécutive, ainsi que sa dégradation via la voie de l'ubiquitine [76]. CDK2 phosphorylée et inactivée s'accumule et ne peut pas phosphoryler Cdc45 pour initier la réplication de l'ADN, induisant un arrêt rapide du cycle cellulaire (Cyc E/A : cyclines E ou A).

Le maintien de l'arrêt du cycle cellulaire au point de contrôle G<sub>1</sub>/S est dépendant de la protéine p53 [77]. p53 qui est phosphorylée par les senseurs ATM/ATR et par les transducteurs Chk1/Chk2, induit la transcription de p21<sup>Waf-1/Cip1</sup>, qui se lie au complexe CDK4-cycline D (ou Cdk4-Cyc D), l'empêchant de phosphoryler pRb [78]. pRb ne peut alors pas libérer le facteur de transcription E2F nécessaire à la transcription des gènes de la phase S. p21<sup>Waf-1/Cip1</sup> inactive également le complexe CDK2-cycline E, assurant le maintien au point de contrôle G<sub>1</sub>/S.

**Figure 8 - Le point de contrôle G<sub>2</sub>/M.**

D'après Sancar *et al.* [65].

Les transducteurs Chk1 et Chk2 activés phosphorylent Cdc25a, favorisant sa liaison aux protéines 14-3-3 et son exclusion nucléaire consécutive. Cdc25a ne peut plus activer le complexe CDK1-cycline B (Cdc2-cycline B), bloquant l'entrée en mitose [75]. De même, Chk1 et Chk2 activés phosphorylent Wee1 qui effectue une phosphorylation inhibitrice de CDK1.



### **3. Les points de contrôle en phase S**

La phase S du cycle cellulaire comprend trois points de contrôle qui ne requièrent pas la protéine p53 : le point de contrôle de réplication, le point de contrôle intra-S et le point de contrôle S/M [80]. Ces points de contrôle sont de découverte plus récente. Le point de contrôle de réplication fait intervenir le complexe senseur ATR-ATRIP et inhibe l'initiation de la réplication de l'ADN quand la progression des fourches de réplication est arrêtée en réponse à différents stress. Le point de contrôle intra-S est activé via ATM par des cassures double-brin de l'ADN. Les transducteurs Chk1 et Chk2 activés induisent alors la phosphorylation et la protéolyse de la phosphatase Cdc25a, conduisant à l'inhibition des complexes CDK2-cyclines E/A. Le point de contrôle S/M vérifie la duplication complète du génome avant d'initier la division cellulaire. La cible clé de ce processus est le complexe CDK1-cycline B.

### **4. Le point de contrôle du fuseau mitotique**

Les protéines du point de contrôle du fuseau mitotique ou point de contrôle métaphase-anaphase détectent un mauvais attachement des kinétochores des chromosomes sur les microtubules des fuseaux mitotiques. Ce point de contrôle induit alors l'arrêt du cycle cellulaire en métaphase afin d'éviter la mauvaise ségrégation des chromosomes et l'aneuploïdie des cellules filles. L'arrêt du cycle cellulaire passe par un blocage de l'ubiquitination et donc de la dégradation de la cycline B, évitant ainsi l'inactivation du complexe CDK1-cycline B et donc la sortie de mitose [81].

## II. Les cyclomodulines de *Escherichia coli*

Comme les virus qui perturbent le cycle cellulaire eucaryote pour leur réplication [82], des bactéries pathogènes sont capables d'interférer activement avec le cycle cellulaire. Cette activité repose sur la synthèse de toxines, appelées cyclomodulines [23] qui stimulent ou inhibent la progression du cycle cellulaire (**tableau 1**). Quatre cyclomodulines ont été identifiées chez *E. coli* : CNF, CDT, Cif et la colibactine.

**Tableau 1 : Exemples de cyclomodulines bactériennes.** D'après Nougayrède *et al.* [23].

Bactérie	Effecteur *	Mécanisme
<b>Inhibe la prolifération</b>		
<i>Helicobacter pylori</i>	VacA	Arrêt du cycle cellulaire des lymphocytes T en G <sub>1</sub> -S, inhibition de l'activation des lymphocytes T
<i>Mycobacterium ulcerans</i>	Mycolactone	Arrêt du cycle cellulaire en G <sub>0</sub> -G <sub>1</sub>
<i>Fusobacterium nucleatum</i>	FIP	Arrêt du cycle cellulaire des lymphocytes T en milieu de phase G <sub>1</sub>
<b>Favorise la prolifération</b>		
<i>Helicobacter pylori</i>	CagA	Active la voie Ras-MEK-ERK, conduisant à une dispersion et une prolifération cellulaire
<i>Bordetella</i> spp.	DNT	Déamide la glutamine de Rho GTPases Forme des cellules géantes multinucléées Augmente la synthèse d'ADN
<i>Pasteurella multocida</i>	PMT	Active de multiples voies cellulaires de transduction du signal Puissant mitogène de divers types cellulaires

\* effecteurs protéiques, sauf la mycolactone qui est un PK.

FIP : « *Fusobacterium* immunosuppressive protein » ; DNT : « dermonecrotic toxin » ; PMT : « *P. multocida* toxin ».

## **A. Le facteur cytotoxique nécrosant CNF**

Le facteur cytotoxique nécrosant 1 ou CNF1 (pour « Cytotoxic Necrotizing Factor 1 ») a été décrit pour la première fois en 1983 par Caprioli *et al.* chez *E. coli*. Il appartient au groupe des cyclomodulines activant le cycle cellulaire. [83].

### **1. Les CNF variants**

Trois variants de CNF1 ont été caractérisés : CNF2 et CNF3 chez *E. coli*, CNFY chez *Yersinia pseudotuberculosis* [84-86]. CNF1 et CNF2 partagent 90% d'identité. CNF3 et CNFY présentent respectivement 70% d'identité avec CNF1 et CNF2, et 65% d'identité avec CNF1. Ces protéines s'apparentent par ailleurs à la toxine DNT de *Bordetella* par leur domaine catalytique [87].

### **2. Structure**

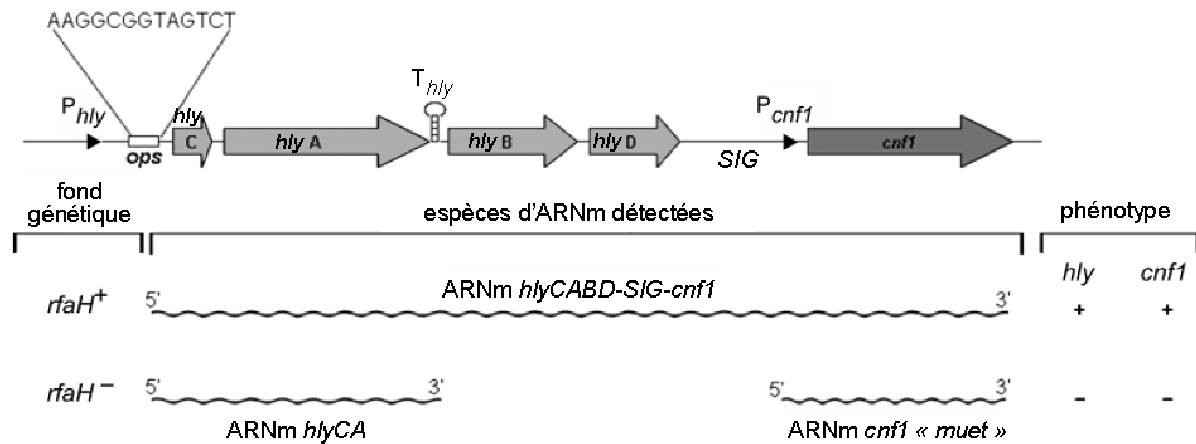
Les toxines de type CNF comprennent une chaîne polypeptidique unique d'environ 115 kDa organisée en trois domaines fonctionnels [87]. La région N-terminale de la protéine contient un domaine de liaison aux récepteurs cellulaires [88, 89]. La partie médiane formerait un domaine central constitué par deux hélices hydrophobes pouvant initier la translocation du domaine enzymatique dans le cytosol des cellules [90]. La partie carboxy-terminale contient le domaine enzymatique qui porte la triade catalytique Val833, Cys866 et His881 [88, 91, 92].

### **3. Support génétique**

Les gènes *cnf1* et *cnfy* sont chromosomiques, alors que *cnf2* est plasmidique (plasmide Vir) [85, 93-95]. L'essentiel des études a porté sur le gène *cnf1*.

Le gène *cnf1* est situé dans l'îlot de pathogénicité II (PAI-II) fréquemment observé chez les *E. coli* du phylogroupe B2 [87]. Il se situe juste en aval de l'opéron *hlyCABD* codant pour la production et la sécrétion de l'hémolysine  $\alpha$  (**figure 9**) [87, 93, 94]. Une corégulation transcriptionnelle de leurs gènes est proposée via 3 mécanismes : une zone promotrice commune, l'action répressive de la protéine histone-like H-NS et une activité activatrice de l'antitermineur transcriptionnel RfaH. Cependant, un promoteur putatif indépendant a été

identifié dans les 60 nucléotides situés en amont du codon d'initiation de *cnf1*. Enfin, une séquence anti-Shine-Dalgarno située à 150 paires de bases en aval du codon d'initiation de *cnf1* réprime la traduction de *cnf1*, probablement par la formation d'une structure en tige-boucle qui séquestre le site de liaison au ribosome [89, 94, 96, 97].



**Figure 9 - Organisation génétique et modèle d'expression de la région *hlyCABD-cnf1* du PAI-II de la souche UPEC J96. D'après Lemonnier *et al* [87].**

L'antiterminateur transcriptionnel protéique RfaH interagit directement avec les complexes de transcription arrêtés sur les sites *ops* afin d'augmenter l'élongation de la transcription par un mécanisme qui empêche la pause et la fin de transcription. Dans le cas de l'opéron *hlyCABD*, les complexes de transcription amorcés par RfaH peuvent contourner l'arrêt de transcription au terminateur  $T_{hly}$  situé entre les gènes *hlyA* et *hlyB*, permettant ainsi la transcription des gènes distaux *hlyB*, *hlyD* et *cnf1*. En plus de RfaH, la protéine histone-like H-NS, co-régule négativement les gènes de l'hémolysine et de *cnf1*.

#### 4. Mode d'action

Le mode d'action des toxines CNF repose sur trois étapes : la sécrétion de la toxine, sa translocation dans le cytosol des cellules cibles et la modification enzymatique des molécules cibles.

##### a) Sécrétion

Les CNF sont sécrétées dans le milieu extérieur. Cependant, aucun peptide signal détectable ou système de transporteur spécifique n'a été caractérisé. La sécrétion de CNF1 à travers la membrane cytoplasmique implique la ferredoxine [98]. Elle est par ailleurs associée



aux vésicules de la membrane externe libérées par les bactéries [97, 99]. L'importance physiologique de cette voie d'adressage reste toutefois à démontrer.

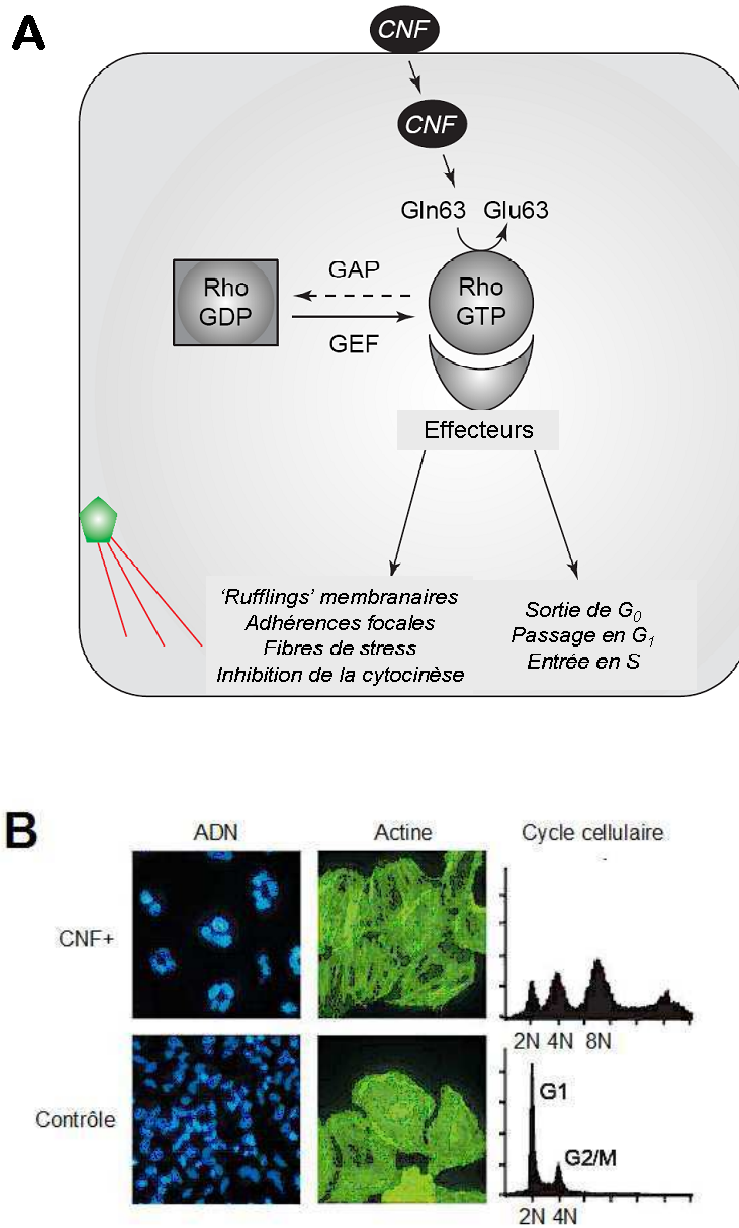
#### b) Translocation dans les cellules eucaryotes

CNF1 se lie à la protéine précurseur du récepteur de la laminine (37 kDa) à la surface de la cellule hôte [100-102]. Deux régions en N-terminale (résidus 135 à 164) et C-terminale (résidus 683 à 730) sont impliquées dans la fixation au récepteur [103, 104]. CNF3 et CNFY se lieraient à des récepteurs cellulaires de surface différents [105, 106].

Après fixation à son récepteur, CNF1 entre dans les cellules hôtes via une endocytose indépendante de la clathrine pour atteindre le compartiment endosomal tardif, permettant la libération du domaine enzymatique et son transfert dans le cytosol grâce aux 2 hélices hydrophobes du domaine médian [90, 102, 105, 107, 108].

#### c) Modification de la cible moléculaire

La cible cytoplasmique des CNF sont les GTPases Rho (**figure 10A**) [106, 108, 109], des interrupteurs moléculaires de la superfamille Ras [110]. Le domaine enzymatique de CNF1 assure la désamidation d'un résidu glutamine spécifique des protéines Rho ; le résidu Gln63 de RhoA ou Gln61 de Rac et Cdc42 [111-114]. Cette modification compromet la formation du complexe Rho-GAP (« GTPase Activating Protein ») nécessaire à l'inactivation des protéines Rho et induit par conséquent une activation permanente des protéines Rho sous forme de complexe avec le GTP. Ainsi activées en permanence, les protéines Rho sont rapidement ubiquitinées et dégradées par le protéasome. Au total, CNF1 induit une activation modérée et transitoire des protéines GTPases Rho [115-117]. Par ailleurs, les toxines CNF ont un spectre d'action particulier ; CNF1 et CNF3 activent RhoA, Rac1 et Cdc42 [106] alors que CNFY active sélectivement RhoA [109]. Les substrats préférentiels de CNF1 sont Rac et Cdc42 et celui de CNF3, RhoA [106].



**Figure 10 - Modèle d'action des CNF sur les cellules hôtes.** D'après Nougayrède *et al* [23].

A. CNF se lie au récepteur cellulaire puis est internalisé par endocytose pour atteindre le compartiment endosomal tardif acide, permettant le transfert de son domaine enzymatique dans le cytosol. Il déamide ensuite un résidu glutamine des GTPases Rho, bloquant celles-ci dans un état activé leur permettant d'activer de manière persistante leurs effecteurs. Ces effecteurs modifient le cytosquelette et favorisent la prolifération cellulaire.

B. Effet de CNF1 sur des cellules HeLa après 72 heures de contact.

A gauche - Multinucléation des cellules révélée par marquage au DAPI.

Au milieu - Fibres de stress révélée par la fluorescéine-phalloïdine.

A droite - Analyse du cycle cellulaire montrant des cellules ayant un contenu en ADN élevé (4n, 8n, 16n ...).

## 5. Expression phénotypique

Les GTPases Rho contribuent à la régulation d'un large panel de processus cellulaires en interagissant avec des effecteurs protéiques spécifiques [110]. Elles régulent l'organisation et la dynamique des microtubules (ségrégation des chromosomes) et du cytosquelette d'actine (polarité, migration et division cellulaires), et des événements de la réponse immunitaire (synapses immunologiques, phagocytose). Ces Rho GTPases contribuent à la progression du cycle cellulaire en influençant le niveau de cycline D1 et d'inhibiteurs de CDK (p27<sup>Kip1</sup> et p21<sup>Waf1/Cip1</sup>) et prennent part à la transcription de nombreux gènes, tels que ceux intervenant dans l'inflammation et l'apoptose. Les effets cytopathogènes induits par les CNF sont la formation de fibres de stress d'actine<sup>1</sup> (**figure 10B**), de plaques d'adhérence focale<sup>2</sup>, de lamellipodes, de filipodes<sup>3</sup> et de « rufflings » membranaires<sup>4</sup> [106, 108, 118, 119]. L'importante activité de « ruffling » permet aux cellules épithéliales de se comporter comme des phagocytes. Elles développent une activité macropinocytaire permettant la phagocytose de différents types de particules, dont les bactéries [117, 120-122].

Les toxines CNF déclenchent la transition de phase G<sub>1</sub>/S et induisent la réplication de l'ADN dans différentes cellules de mammifères. Probablement en raison d'altérations du fuseau mitotique, la ségrégation chromosomique et la cytokinèse<sup>5</sup> sont altérées. Des aberrations nucléaires (multinucléation, segmentations nucléaires et mitoses multipolaires) conduisant à des aneuploïdies sont associées à une augmentation de la teneur cellulaire en ADN (4n, 8n, 16n) (**figure 10B**) [118, 119, 123-125].

CNF1, via les protéines Rho, active plusieurs voies de signalisation dont celles mettant en jeu les kinases Jun, MAP p38 et NF-κB, des facteurs de transcription majeurs de la réponse inflammation et de la différenciation cellulaire [114, 126].

---

<sup>1</sup> Faisceaux filamentaires organisés dans lesquels des faisceaux de courts filaments d'actine de polarité alternée sont entremêlés avec des filaments bipolaires de myosine II, retrouvés généralement dans les cellules mobiles.

<sup>2</sup> Structures d'adhérence caractéristiques des cellules en culture, correspondant aux sites dispersés et discontinus de contact très étroit avec la surface de la boîte.

<sup>3</sup> Respectivement larges et fines extensions membranaires rigides faites de polymères d'actine utilisées par la cellule pour explorer son environnement à la recherche de matrice extra-cellulaire et à l'origine du mouvement amiboïde des cellules.

<sup>4</sup> Surfaces cellulaires motiles contenant un maillage de filaments d'actine nouvellement polymérisés.

<sup>5</sup> Phase finale de la division cellulaire qui correspond à la séparation physique des deux cellules filles.

## 6. Distribution

Les souches de *E. coli* *cnf1*<sup>+</sup> (« Necrotoxic *E. coli* 1 » ou NTEC-1) sont issues de l'homme et de nombreuses espèces de mammifères domestiques. Ces souches NTEC-1 appartiennent majoritairement au phylogroupe B2 [127-130]. Les souches *cnf2*<sup>+</sup> (NTEC-2) ont été isolées le plus souvent d'animaux [86, 131-134]. Récemment décrit, CNF3 n'a été mise en évidence que dans des souches d'origine animale [84].

Les souches NTEC-1 sont fréquemment responsables de bactériémies [135] et d'infections urinaires, avec une prévalence croissante allant de 21 à 78 % en fonction des affections suivantes : cystites, pyélonéphrites et prostatites [127-129, 136-144]. Les souches de *E. coli* responsables de méningites néonatales sont *cnf1*<sup>+</sup> dans 10 à 13% des cas [130, 145, 146].

La participation des souches NTEC aux syndromes diarrhéiques de l'homme n'est pas clairement établie, bien que les souches NTEC semblent exercer un effet pathogène sur l'intestin, au moins chez les nouveau-nés et les jeunes enfants [133, 147-150]. L'implication des souches NTEC-1 et -2 dans la survenue de syndromes dysentériques chez les animaux est suggérée par quelques études épidémiologiques et un modèle expérimental [86, 134, 151, 152]. Cependant la pathogénie intestinale des NTEC repose probablement sur la production combinée de plusieurs toxines (hémolysine, CNF, CDT, ...).

## 7. Rôle dans le processus infectieux

La forte prévalence de CNF1 et CNF2 dans les souches ExPEC suggère un rôle dans la pathogénie des infections extra-intestinales. CNF1 contribue à l'invasion du système nerveux central par les souches de *E. coli* K1. Elle favorise l'internalisation de *E. coli* dans les cellules endothéliales microvasculaires du cerveau humain *in vitro* et ainsi contribuerait à la traversée de la barrière hémato-encéphalique [100, 153]. De même, CNF1 aiderait l'internalisation des bactéries dans les cellules épithéliales, notamment de la vessie [117, 120-122]. Ainsi, abritées des effecteurs du système immunitaire, les bactéries productrices peuvent persister et former des réservoirs intracellulaires responsables des infections urinaires récurrentes [154]. L'intoxication par CNF1 altère les jonctions intercellulaires et induit l'apoptose de cellules épithéliales vésicales en culture, ce qui faciliterait l'invasion des tissus sous-jacents par les bactéries [155].

CNF1 augmente la production de médiateurs de l'inflammation sur des cellules uro-épithéliales en culture [156]. Des études menées à l'aide de modèles animaux d'infections urinaires et de prostatites montrent que cette toxine contribue aussi aux dommages tissulaires et à l'inflammation [156-159]. Des niveaux plus élevés des facteurs pro-inflammatoires MCP-1 et IL-8 ont été mesurés dans l'urine de patients infectés par des UPEC *cnf1*-positives [140] ; néanmoins, ceci n'exclut pas la participation d'autres facteurs bactériens, parmi lesquels l'hémolysine  $\alpha$  et le P-fimbriae, qui déclenchent également la production de médiateurs inflammatoires par les cellules épithéliales intoxiquées [160, 161]. CNF1 pourrait donc majorer les dommages de l'épithélium *via* l'inflammation.

CNF1 améliore aussi la survie bactérienne dans un modèle murin [158]. Plusieurs études suggèrent que CNF1 retarde la transmigration des neutrophiles à travers l'épithélium et réduit leurs fonctions antibactériennes (régulation négative de la phagocytose, altération de la distribution des récepteurs CR3 du complément) [99, 162-164]. Bien qu'il favorise un état inflammatoire, CNF1 faciliterait cependant la survie des bactéries face aux effecteurs du système immunitaire inné en modulant leur fonction. CNF1 induirait une réponse inflammatoire contrôlée ou en partie inopérante et permettrait ainsi une plus grande persistance des bactéries [164].

En conclusion, les études *in vitro* et *in vivo* convergent vers un modèle dans lequel CNF1 est un facteur de virulence favorisant l'internalisation des bactéries et modulant les réponses inflammatoires. Il est donc concevable que les mécanismes d'action de CNF1 permettent d'augmenter la gravité et la chronicité des infections bactériennes.

## 8. Des agents carcinogènes potentiels

Plusieurs arguments soutiennent le rôle potentiel des CNF dans la transformation cellulaire<sup>6</sup>.

La modification biochimique des protéines Rho induite par les CNF correspond à la modification des protéines Ras d'origine génétique, observée dans de nombreuses tumeurs [165]. Les toxines CNF dérégulent l'activité des GTPases Rho comme les effecteurs SopE et SptP de *Salmonella* spp., bactéries dont le portage est associé au cancer de la vésicule biliaire [166, 167].

---

<sup>6</sup> Capacité acquise d'une cellule à proliférer de manière monoclonale et excessive, en échappant à l'homéostasie cellulaire.

Dans des cellules épithéliales, CNF1 induit une activation prolongée du facteur de transcription NF- $\kappa$ B [126], un acteur reconnu de l'oncogenèse [168]. Cette activation induit l'expression de *COX2* [169], un gène majeur de la cascade des prostaglandines pro-inflammatoires dont la surexpression est observée précocement dans de nombreux cancers comme le cancer colorectal. NF- $\kappa$ B induit également l'expression et la libération de cytokines pro-inflammatoires (IL-6, IL-8 et TNF- $\alpha$ ) à partir des cellules épithéliales et endothéliales [156, 164]. CNF1 en activant de façon prolongée la voie NF- $\kappa$ B favorise donc un état inflammatoire chronique, qui est reconnu pour favoriser la carcinogenèse [170].

L'activité de CNF1 protège les cellules épithéliales de l'apoptose en augmentant l'expression des facteurs anti-apoptotiques Bcl-2 et Bcl-X<sub>L</sub>, module l'autophagie, induit des aneuploïdies et la formation de réseaux complexes de mitochondries ; des altérations pouvant conduire à la transformation cellulaire [123, 171-173].

CNF1 promeut la rupture des jonctions cellulaires et la motilité cellulaire, par un mécanisme dépendant de RhoA [117, 172, 174] ; un mécanisme proche de la transition épithéliale-mésenchymateuse<sup>7</sup> [175].

Enfin, l'activité de CNF1 s'apparente par de nombreux aspects à celle des protéines CagA et VacA de *H. pylori*, un des meilleurs exemples de toxines carcinogènes (**tableau 2**).

---

<sup>7</sup> Ensemble d'événements permettant aux cellules épithéliales de surmonter les contraintes physiques, telles que celles imposées par les jonctions intercellulaires et d'adopter le phénotype migratoire intervenant dans l'embryogénèse ou le processus métastatique des cellules cancéreuses.

**Tableau 2 - Comparaison entre les effets cellulaires de CNF1 de *E. coli* et des facteurs de virulence majeurs (VacA et CagA) de *H. pylori*. D'après Travaglione *et al* [176].**

	<b>CNF1</b>	<b>CagA</b>	<b>VacA</b>
<b>Effets sur le cytosquelette et la morphologie cellulaire</b>	- polymérisation de l'actine - diffusion cellulaire - multinucléation	- réarrangement de l'actine - perte de polarité cellulaire - dispersion/migration cellulaire	- perturbations de l'actine - vacuolisation
<b>Régulation de l'apoptose</b>	Effet anti-apoptotique	Effet anti-apoptotique	Effet pro-apoptotique
<b>Production de médiateurs de l'inflammation</b>	TNF- $\alpha$ IL-6 IL-8 ROS	IL-8 ROS	TNF- $\alpha$ IL-6 IL-1 $\beta$ IL-8
<b>Activation de facteurs de transcription</b>	NF-kB	NF-kB NFAT SRF AP-1	NF-kB NFAT ATF-2
<b>Effets sur le cycle cellulaire</b>	Endoreduplication Arrêt en G <sub>2</sub> /M	Inhibition de la transition G <sub>1</sub> /S	Induction de la transition G <sub>1</sub> /S
<b>Effets sur les mitochondries</b>	Inhibition de la dépolarisation membranaire induite par les UVB		Réduction du potentiel de membrane, libération de cytochrome c

ROS : Espèces réactives de l'oxygène

## **B. La toxine de distension cytoléthale CDT**

La toxine de distension cytoléthale (« Cytolethal Distending Toxin » ou CDT) a été caractérisée en 1997 comme la première toxine bactérienne bloquant le cycle cellulaire en transition de phase G<sub>2</sub>/M [177]. Ce type de toxine a été caractérisé chez de nombreuses bactéries à Gram négatif pathogènes, comme *E. coli*, *Shigella dysenteriae*, *Salmonella Typhi*, *Haemophilus ducreyi*, *Helicobacter hepaticus*, et *Aggregatibacter* (anciennement *Actinobacillus) actinomycetemcomitans* [178]. CDT est constituée de trois sous-unités CdtA, CdtB et CdtC.

### **1. Les variants CDT et leur support génétique**

Cinq variants, CDT-I à CDT-V ont été caractérisés chez *E. coli*. Les sous-unités CdtB sont plus conservées que les sous-unités CdtA et CdtC (>56% versus >36% et >33% d'identité) [179-181].

Elles sont codées par trois gènes *cdtA*, *cdtB* et *cdtC* organisés en un opéron le plus souvent chromosomique, parfois plasmidique (plasmide Vir codant CDT-III et CNF2). Leur expression semble constitutive. Chez *E. coli*, ces gènes sont souvent flanqués de séquences phagiques [95, 182-184].

### **2. Structure des CDT**

Les toxines CDT appartiennent aux toxines holoprotéiques de type AB<sub>2</sub>. Les trois sous-unités CdtA, CdtB et CdtC forment un complexe ternaire [185, 186]. CdtA et CdtC partagent un motif lectine et sont responsables de la liaison de la toxine à des récepteurs cellulaires membranaires [185, 187, 188]. CdtB, la sous-unité la plus conservée, présente une homologie structurale et fonctionnelle avec les membres d'une superfamille de métallo-enzymes incluant des nucléases et diverses phosphatases de seconds messagers lipidiques. Les résidus His154, Asp229, His261 et Asp260 des sous-unités CdtB sont impliqués dans la liaison de l'ion métallique Mg<sup>2+</sup> et le processus enzymatique catalysé par la toxine [186, 189, 190].



### 3. Mode d'action

Le mode d'action des toxines CDT repose sur trois étapes : l'assemblage et la sécrétion du complexe protéique ternaire, la translocation de la sous-unité CdtB dans le cytosol des cellules cibles et l'altération de la cible, l'ADN (**figure 11**).

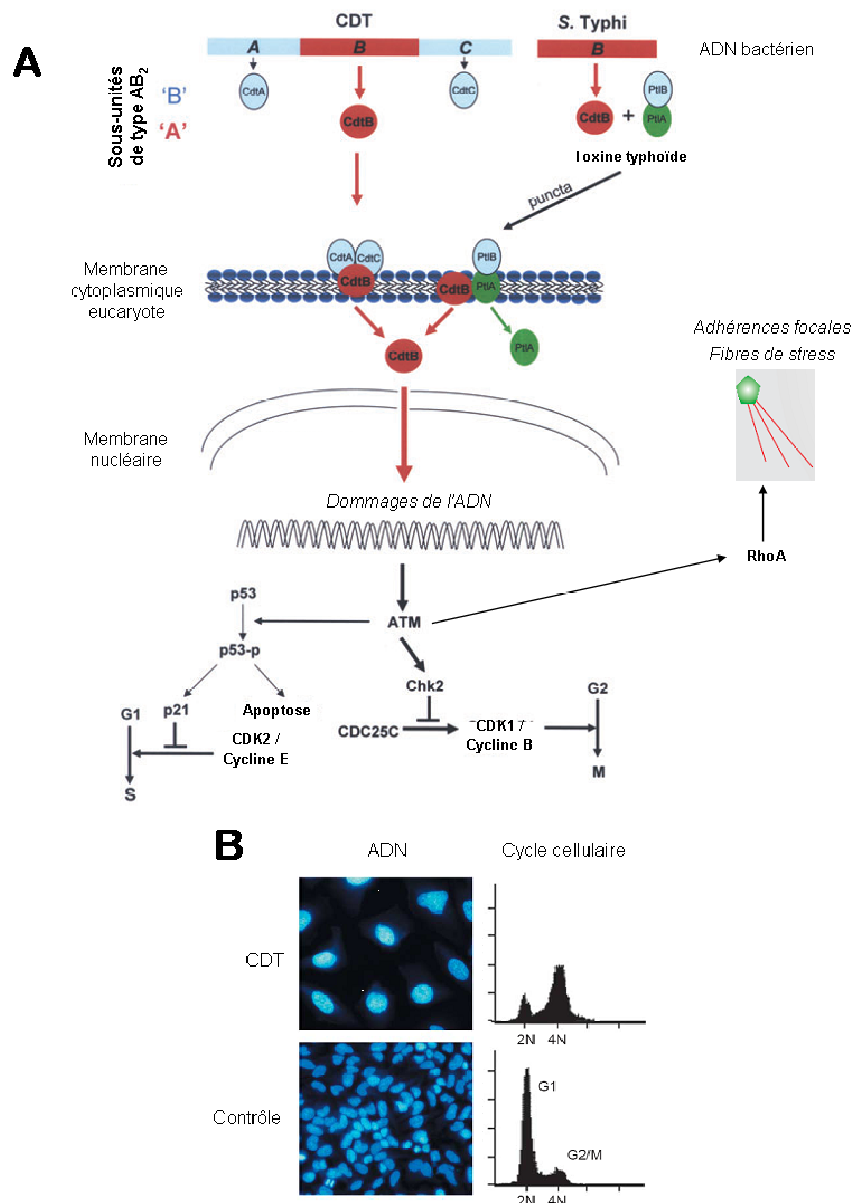
#### a) Sécrétion

Les trois protéines CdtA, CdtB et CdtC sont transloquées de façon indépendante dans le périplasme par une séquence signal d'adressage [191, 192]. La sous-unité CdtA y subit un greffage lipidique au niveau d'un domaine consensus appelé « lipobox », important pour une production efficace de l'holotoxine [193]. Les trois sous-unités forment alors un complexe périplasmique ternaire. Dans le surnageant de culture, la majorité des CDT reste étroitement associée aux vésicules de la membrane externe [194].

#### b) Translocation dans les cellules eucaryotes

Deux types de récepteurs cellulaires ont été impliqués dans la liaison de CDT à la surface cellulaire : des glycoprotéines contenant du fucose pour CDT-II [195] et des glycosphingolipides GM3 pour la toxine CDT de *A. actinomycetemcomitans* [196]. Les radeaux lipidiques riches en cholestérol de la membrane cytoplasmique sont également impliqués dans l'adressage de CdtB dans la cellule cible [197]. Lors de la translocation de CdtB dans la cellule cible, CdtA et CdtC demeureraient associées à la membrane (**figure 11A**). La sous-unité CdtB est internalisée par endocytose, transportée par voie rétrograde vers l'appareil de Golgi et le réticulum endoplasmique, puis transloquée dans le noyau, son site d'action [198-200]. L'entrée de CdtB dans le noyau repose sur un domaine de translocation nucléaire [201].

La toxine CDT de *Salmonella* Typhi présente un comportement particulier. L'expression de sa sous-unité CdtB est induite par l'invasion des cellules hôtes. Les gènes *cdtA* et *cdtC* étant absents, CdtB s'associe alors avec les protéines PltA (toxine de type pertussique A) et PltB (toxine de type pertussique B). Ce complexe désignée « toxine typhoïde » (**figure 11A**) est empaqueté dans une structure membranaire appelée «puncta», et transportée hors des cellules infectées. CdtB induit alors l'intoxication des cellules cibles par son adressage aux cellules infectées par *S. Typhi* (mode autocrine) ou à des cellules non infectées (mode paracrine) [202, 203].



**Figure 11 - Organisation génétique et modèle d'action des CDT sur les cellules hôtes.** D'après Ge *et al* et Nougayrède *et al* [23, 178].

- A. CDT est une toxine hétéro-trimérique dont les gènes codant les sous-unités sont organisés en opéron. CdtA et CdtC sont nécessaires à la liaison de CDT à la surface de la cellule hôte. L'internalisation de CdtB est suffisante pour l'intoxication cellulaire. CdtB est internalisée par endocytose, transportée par voie rétrograde vers l'appareil de Golgi et le réticulum endoplasmique, puis transloquée dans le noyau où elle induit des altérations de l'ADN. La production et translocation de CdtB chez *Salmonella Typhi* fait intervenir les protéines PtlA et PtlB. Les dommages de l'ADN activent les voies de réponse ATM dépendante, entraînant un arrêt du cycle cellulaire en transitions de phase G<sub>1</sub>/S ou G<sub>2</sub>/M, voire une apoptose selon le modèle cellulaire. L'activation de RhoA par ATM peut conduire à la formation de fibres de stress et d'adhérences focales.
- B. Effet d'une CDT purifiée de *E. coli* sur des cellules HeLa 72 heures après le début de l'exposition.  
 A gauche - Elargissement des noyaux et absence de figures mitotiques révélés par marquage au DAPI.  
 A droite - Analyse du cycle cellulaire montrant une majorité de cellules bloquées en phase G<sub>2</sub>.

### c) Altération de l'ADN et expression phénotypique

Une fois dans le noyau, CdtB induit des dommages subtils de l'ADN détectables par la phosphorylation des histones H2AX, un évènement de réponse caractéristique des cassures double-brin de l'ADN [190, 204-208]. Cependant, l'action directe des CDT sur l'ADN reste à démontrer. Deux études suggèrent des mécanismes par lesquels CdtB induirait des dommages de l'ADN autres, notamment grâce à une activité phosphatidylinositol 3,4,5-triphosphate phosphatase [209, 210].

Les dommages de l'ADN associés aux CDT de différentes espèces bactériennes induisent l'activation de la kinase ATM, dans des cellules en prolifération ou non. Selon le modèle cellulaire utilisé, un blocage du cycle cellulaire est observé en transitions de phase G<sub>1</sub>/S ou G<sub>2</sub>/M (**figure 11**) [177, 178, 190, 205, 208, 211-214]. Le sort des cellules après intoxication par CDT dépendrait aussi de la spécificité de liaison des différentes sous-unités CdtA-CdtC aux récepteurs [215]. L'activation d'ATM par CDT active la protéine RhoA, conduisant à la formation de fibres de stress et d'adhérences focales [207, 216].

## 4. Distribution et rôle dans le processus infectieux

Le rôle potentiel des CDT dans la pathogénicité de *E. coli* est uniquement étayé par des études épidémiologiques. L'association préférentielle des gènes *cdt* au phylogroupe B2 de *E. coli* a été rapportée dans une étude [217]. Cependant, la faible prévalence des gènes *cdt* chez les UPEC (2 à 21%) et les souches responsables de bactériémies (10%) suggère que CDT n'est pas un facteur de virulence majeur dans ces contextes [127, 128, 135, 144, 217]. La prévalence de *cdtB* chez les *E. coli* fécaux issus de patients présentant ou non des signes digestifs est faible (1 et 5%), de même que chez les pathotypes intestinaux EPEC et EHEC [128, 131, 179, 217]. Une étude cas-témoin portant sur les souches de *E. coli cdt+* isolées d'enfants n'a pas montré d'association entre ces souches et un épisode diarrhéique chez l'enfant [218]. L'association *cdt3/cnf2* a été observée essentiellement dans des souches de *E. coli* isolées d'animaux, présentant fréquemment des troubles digestifs et des sepsis [132, 134].

Les toxines CDT joueraient un rôle important dans le pouvoir pathogène d'autres espèces bactériennes. La toxine CDT de *H. hepaticus* (un pathogène murin entéro-hépatique) est impliquée dans l'induction d'une inflammation et la persistance de l'infection au niveau intestinal et hépatique [219-222]. Des rôles similaires ont été attribués aux toxines CDT de *C. jejuni* [223-225]. Cependant les résultats diffèrent selon les modèles animaux utilisés [226]. La production de toxines CDT par *H. ducreyi* aggrave les chancres chez le lapin [227].

Les toxines CDT induisent la nécrose des cellules épithéliales et des fibroblastes, cellules impliquées dans la réparation des lésions produites par les pathogènes [215]. Elles ciblent aussi les cellules de la réponse immunitaire. Elles provoquent un arrêt du cycle cellulaire suivi de l'apoptose des lymphocytes humains [213, 228, 229] et inhibent la production d'oxyde nitrique par les macrophages murins d'une manière dose-dépendante [230]. Aussi, il a été proposé que ces toxines agissent sur les cellules du système immunitaire des muqueuses et donc altèrent la capacité de l'hôte à éliminer les bactéries productrices.

## **5. Des agents carcinogènes potentiels**

La toxine CDT produite par la bactérie hépato-carcinogène *H. hepaticus* favorise le développement de dysplasie hépatique dans un modèle murin [219]. L'activité génotoxique de CDT favoriserait l'acquisition d'évènements mutationnels pouvant activer des gènes oncogènes et/ou inactiver des gènes suppresseurs de tumeurs.

## **C. Le facteur d'inhibition du cycle Cif**

Le facteur d'inhibition de cycle ou Cif («Cycle Inhibition Factor») est une cyclomoduline protéique inhibant le cycle cellulaire. Elle est observée chez les pathotypes intestinaux EPEC et EHEC [231-233]. Quatre protéines homologues de Cif ont été mises en évidence chez *Yersinia pseudotuberculosis*, *Phototribadus luminescens*, *Phototribadus asymbiotica* et *Burkholderia pseudomallei* [234, 235].

### **1. Distribution et support génétique**

La prévalence du gène *cif* a été évaluée à moins de 2% chez des souches de *E. coli* d'origine environnementale et animale (n=5 001) [233]. Ce gène est inactif dans 66% des souches suite à un décalage du cadre de lecture ou à la présence d'une séquence d'insertion. Un point commun des souches porteuses du gène est la présence du locus d'attachement et d'effacement des entérocytes (locus LEE), qui est observé chez de nombreux EHEC et caractérise le pathotype EPEC [233]. Ce locus code un système de sécrétion de type III (SST3) permettant l'injection dans la cellule hôte d'effecteurs bactériens codés ou non par ce locus. Bien que retrouvé principalement dans le groupe phylogénétique B1, le gène *cif* est retrouvé dans les 4 principaux groupes phylogénétiques de *E. coli*. Il est cependant absent du groupe E, incluant la plupart des souches d'EHEC de sérotype O157:H7 [233].

La toxine Cif n'est pas codée par LEE, mais par un prophage lambdaïde présent dans la plupart des souches EHEC et EPEC (**figure 12A**) [232, 233]. Le prophage peut être induit et former des particules phagiques infectieuses. La région d'insertion du prophage portant le gène *cif* est très variable et porte de nombreuses traces de remaniements génétiques liés à l'activité de séquences d'insertion [233]. Les gènes homologues de *cif* sont aussi présents dans des régions génomiques fortement réarrangées et les bactéries les hébergeant possèdent les gènes codant un SST3 fonctionnel [234].

### **2. Structure**

Cif porte en C-terminal un domaine effecteur et en N-terminal, une séquence de sécrétion et de translocation via le SST3 (**figure 12A**). Cette séquence d'adressage SST3 est échangeable avec celles d'autres protéines effectrices utilisant le SST3 [236]. La structure cristalline de Cif a révélé qu'elle est un membre divergent d'une superfamille d'enzymes

incluant des protéases à cystéine active, des acétyltransférases et des transglutaminases avec qui elle partage 23 à 56% d'identité. Ces protéines présentent un repliement similaire et la triade catalytique Cys109, His165 et Gln185 caractéristique de cette superfamille. Ce motif est crucial pour l'activité biologique de Cif [235, 237, 238]. Cif est insensible à l'action des inhibiteurs irréversibles des cystéine protéases mais sensibles à l'action des inhibiteurs spécifiques des cystéine protéases de type papaïne [235, 238].

### 3. Mode d'action et expression phénotypique

Cif induit un effet cytopathique progressif caractérisé par le recrutement de nombreuses plaques d'adhérences focales, l'assemblage de fibres de stress d'actine et l'élargissement du noyau dû à l'inhibition de la mitose (**figure 12B**) [232, 239]. Le cycle cellulaire s'arrête de façon irréversible en G<sub>2</sub>/M ou G<sub>1</sub>/S selon le stade du cycle lors de l'intoxication [232, 240]. Les cellules synthétisent de l'ADN sans division cellulaire (endoréduplication<sup>8</sup>). La teneur en ADN des cellules augmente ainsi et peut atteindre 16n (**figure 12B**) [23, 232, 239]. Enfin, Cif peut provoquer une mort cellulaire retardée, 48 heures après l'infection [241].

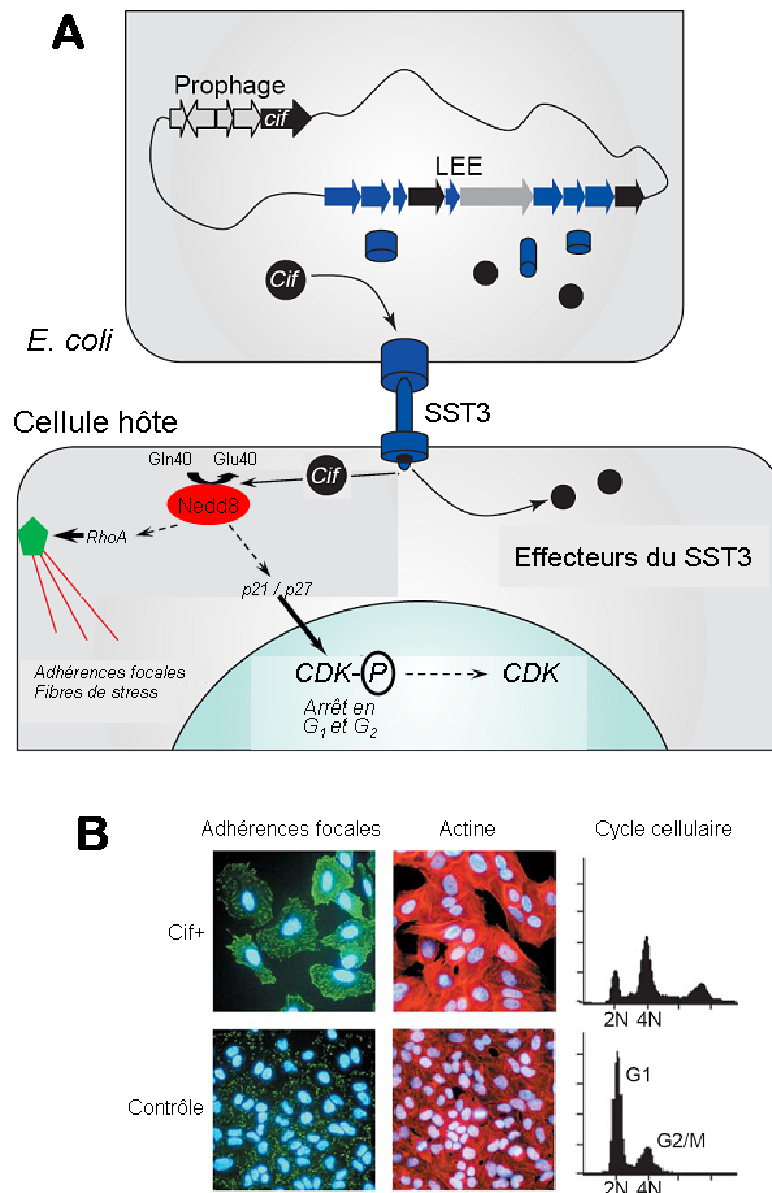
Le blocage du cycle cellulaire de l'hôte induit par Cif est lié à l'accumulation des CKI p21<sup>Waf1/Cip1</sup> et p27<sup>Kip1</sup>, suite à l'inhibition de leur dégradation par la voie ubiquitine/protéasome (**figure 12A**) [240]. Cette accumulation entraîne le maintien des CDK dans un état inactif [231, 232, 239]. Récemment, Cui *et al.* ont montré que Cif et son homologue CHBP de *B. pseudomallei* sont de puissants inhibiteurs de la voie d'ubiquitination eucaryote [242]. CHBP déamide la glutamine 40 de l'ubiquitine et de la protéine ubiquitine-like Nedd8, alors que Cif déamide spécifiquement Nedd8. Ces toxines altèrent ainsi le processus d'ubiquitination et donc la dégradation de protéines comme p21<sup>Waf1/Cip1</sup>, p27<sup>Kip1</sup> et RhoA. L'accumulation de ces dernières rend compte de l'effet cytopathique progressif observé après intoxication.

### 4. Un agent carcinogène potentiel

L'activité de Cif sur l'ubiquitination suggère une activité pléiotropique, qui peut affecter des protéines clés de la cancérogenèse comme p53 et les ubiquitine ligases [243]. Les endoréductions induites par Cif peuvent aussi engendrer des anomalies génétiques susceptibles de favoriser le développement tumoral [244].

---

<sup>8</sup> Duplication de l'ADN sans la séparation des chromosomes ni celle des cellules filles (cycle cellulaire sans phases G<sub>2</sub> et M).



**Figure 12 - Localisation génétique et modèle d'action de Cif sur les cellules hôtes.** Modifié d'après Nougayrède *et al* [23].

A. Les souches de *E. coli* possédant un système de sécrétion de type III (SST3) codé par le locus d'attachement et d'effacement des entérocytes (locus LEE), injectent directement dans la cellule hôte des effecteurs protéiques, dont Cif codé sur un prophage lambdaïde. Cif déamide la protéine ubiquitine-like Nedd8 sur la glutamine 40. Ces altérations affectent le processus d'ubiquitination par Nedd8 et donc la dégradation de protéines comme  $p21^{Waf1/Cip1}$ ,  $p27^{Kip1}$  et RhoA. L'accumulation de ces dernières rend compte de l'effet cytopathogène progressif.

B. Les cellules HeLa ont été infectées pendant 4 heures avec une souche de *E. coli cif+* et analysées 72 heures après infection.

A gauche – Plaques d'adhérence focales révélées par la taline.

Au milieu – Fibres de stress révélées par la rhodamine-phalloïdine.

A droite – Analyse du cycle cellulaire montrant des cellules arrêtées en  $G_2/M$  (majoritaires) et en  $G_1$ , et des cellules ayant poursuivi leur synthèse d'ADN sans mitose (endoréduplication).

## **D. La colibactine**

La colibactine est une cyclomoduline récemment et fréquemment observée chez les Entérobactéries dont *E. coli* [24]. Sa structure est inconnue à ce jour. Cependant, son support génétique a été caractérisé ainsi que son activité biologique.

### **1. Support génétique**

Le support génétique permettant la synthèse de la colibactine est un îlot génomique de 54 kb (**figure 13**) appelé *pks* [24]. Il contient 23 gènes putatifs dont 16 sont requis pour l'effet toxique sur le cycle cellulaire.

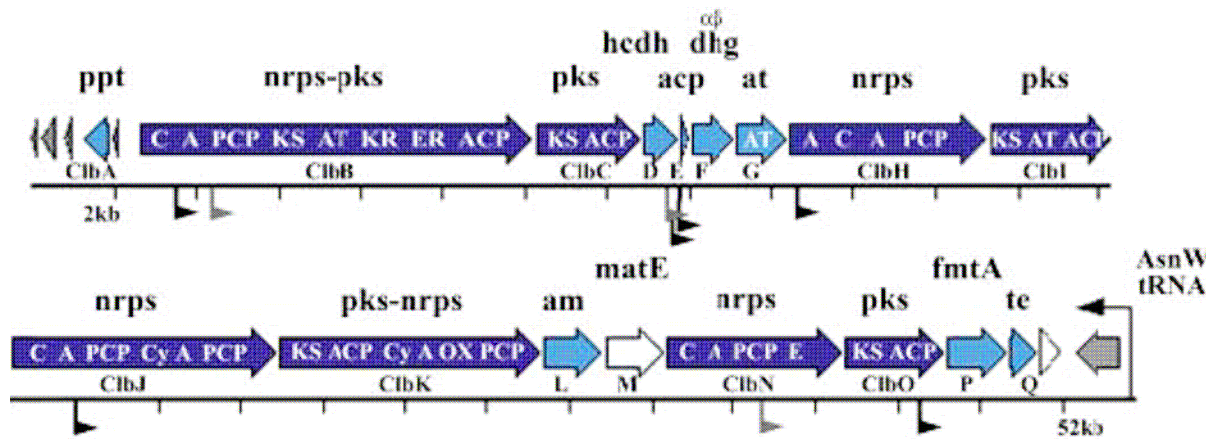
L'analyse *in silico* des gènes essentiels a permis d'identifier leur fonction putative [24] : trois gènes codent des PKS (*clbC*, *clbI*, *clbO*), trois gènes codent des NRPS (*clbH*, *clbJ*, *clbN*) deux codent des PKS-NRPS (*clbB* et *clbK*) et six gènes codent des enzymes accessoires. La coopération de ces enzymes codées par des blocs de gène est connue pour produire un composé chimique hybride de type PK-NRP (cf chapitre III). La colibactine serait donc un composé PK-NRP responsable de l'activité cyclomoduline. Parmi les protéines essentielles, on dénombre aussi ClbL annoté « amidase » et ClbP annoté « protéine FmtA-like ». Ce type de protéines n'est pas rapporté comme intervenant dans la biosynthèse de composés PK, NRP ou PK-NRP.

L'îlot *pks* code sept protéines putatives non requises pour l'activité toxique [24] : la protéine ClbR possédant un motif de liaison à l'ADN hélice-tour-hélice et pouvant intervenir dans la régulation de cet îlot génomique, une pompe d'efflux ClbM, une petite protéine de 170 acides aminés (ORF2) et quatre protéines (une intégrase et trois transposases) probablement impliquées dans l'histoire mobile de l'îlot.

### **2. Transcription de l'îlot *pks***

L'étude de la transcription de l'îlot *pks* par RT-PCR a révélé sept transcripts, dont quatre sont polycistroniques. Ces polycistrons comprennent les gènes *clbC* à *clbG* (6,3 kb), *clbI* à *clbN* (23,3 kb), *clbO* à *clbP* (3,9 kb) et *clbR* à *clbA* (0,9 kb) (**figure 14**) [245].



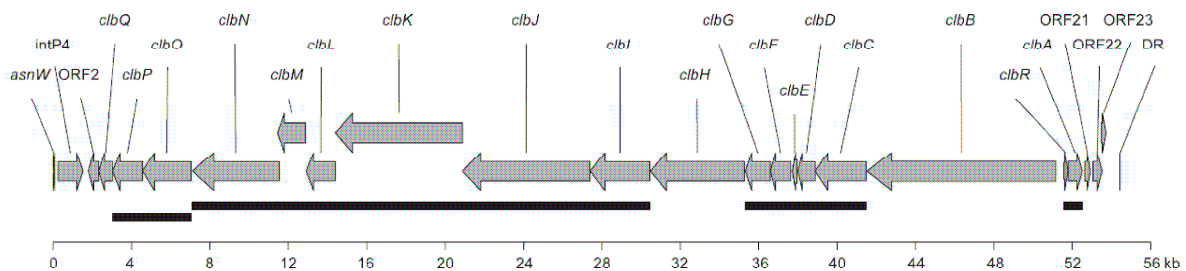


**Figure 13 - Carte schématique de l'îlot *pks* (54kb).** D'après Nougayrède *et al* [24].

Les cadres ouverts de lecture (ORFs) dont les gènes sont impliqués dans la synthèse de PK-NRP et l'effet cytopathique sont indiqués dans différentes nuances de bleu (NRPS et PKS en bleu foncé ; les autres en bleu clair). Les ORFs qui ne sont pas strictement nécessaires à l'effet cytopathique sont en blanc. Les ORFs des transposases et des intégrases sont en gris. Les fonctions prédites sont indiquées au-dessus des ORFs : ppt, phosphopantéthéinyl transférase ; nrps, « nonribosomal peptide synthetase » ; pks, « polyketide synthase » ; hedh, hydroxyl acyl CoA déhydrogénase ; acp, « acyl carrier protein » ; dhg,  $\alpha$   $\beta$  déhydrogénase ; at, acyl-transférase ; am, amidase ; te, thioestérase.

L'organisation prédite des domaines des NRPS et PKS est indiquée comme suit : A, adénylation; ACP/PCP, « acyl carrier protein/peptide carrier protein » ; AT, acyltransférase; C, condensation; Cy, cyclization; ER, enoyl réductase; KR, ketoacyl réductase; KS, ketoacyl synthase; OX, oxydation.

La localisation des insertions de transposons entraînant la perte de l'effet cytopathique sont indiqués par des drapeaux noirs et gris.



**Figure 14 - Unités transcriptionnelles de l'îlot *pks* déterminées par RT-PCR.** D'après Homburg *et al* [245].

Les unités transcriptionnelles opéroniques sont indiquées par des barres noires horizontales. DR, répétition directe ; *asnW*, gène de l'ARNt *asnW* ; intP4, intégrase ; ORF2, protéine hypothétique conservée ; ORF21 et ORF22, transposases ; ORF23, fragment de transposase.

Les niveaux de transcription en culture sous agitation sont supérieurs à ceux obtenus en culture statique. La seconde partie de l'îlot *pks* (*clbJ* à *clbQ*) est transcrite à des niveaux plus faibles quelles que soient les conditions de culture testées. La coculture des bactéries avec des cellulaires eucaryotes n'a pas d'influence sur la transcription des gènes. Au cours de la croissance bactérienne, une augmentation continue de l'expression des gènes est observée en milieu minimum, LB ou DMEM avec un maximum en fin de phase exponentielle et en début de phase stationnaire. Une influence de la source de carbone et d'énergie est observée sur les niveaux de transcription en milieu minimum [245].

### 3. Distribution de l'îlot *pks* au sein des Entérobactéries

L'îlot *pks* fait partie du pool de gènes flexible de *E. coli*. Il est presque exclusivement observé dans les isolats du groupe phylogénétique B2 ; seulement 3 isolats du groupe phylogénétique B1 ont été rapportés. Sa prévalence est de 20 à 34% dans les isolats fécaux (n=236) et 36 à 53% dans les ExPEC (n =215) [24, 246, 247]. L'îlot *pks* est statistiquement associé aux souches de *E. coli* isolées lors de bactériémies et produisant de nombreux facteurs de virulence, mais n'est pas le trait le plus prédictif d'une origine septique [247]. Il est cependant absent des souches K12 de laboratoire et des souches de *E. coli* de pathotypes intestinaux (n =737), de *Shigella* spp. (n=10) et de *Salmonella enterica* (n=114) [24, 246, 247].

L'îlot *pks* est observé dans d'autres genres de la famille des *Enterobacteriaceae* : *Klebsiella pneumoniae* (3,5%, n=141), *Enterobacter aerogenes* (27,3%, n=11) et *Citrobacter koseri* (n=1) [246]. La séquence de l'îlot *pks* est très conservée chez ces espèces d'*Enterobacteriaceae* (>98% d'identité de séquences nucléotidiques). Des variations de séquence ont été uniquement observées dans la région intergénique *clbR-clbB*. Cette région contient un nombre variable (2 à 20) de répétitions en tandem d'octanucléotides. Aucune corrélation entre la taille de la région, le nombre pair ou impair de répétitions, la parenté des souches, le pathotype et l'activité cytopathogène de *pks* n'a pu être mise en évidence [246].

### 4. Environnement génétique de l'îlot *pks*

Chez les *E. coli* du groupe B2, l'îlot *pks* est inséré au niveau du locus de l'ARNt *AsnW* [24], un « point chaud » d'intégration des éléments mobiles [248]. Fait intéressant, l'îlot génomique *pks* chez *K. pneumoniae*, *E. aerogenes* et *C. koseri*, et les trois souches de *E. coli*

du groupe B1 est associé à des séquences mobiles apparentées aux éléments d'intégration et de conjugaison (ICE), *ICEEc1* et *ICEKp1*. Différents gènes *asn*-ARNt ont servi de sites d'insertion chromosomique à ces éléments chez les souches de *E. coli* du groupe B1 et de *K. pneumoniae*. Dans la souche *C. koseri* étudiée, l'îlot *pks* est inséré dans l'environnement proche d'un *asn*-ARNt [246].

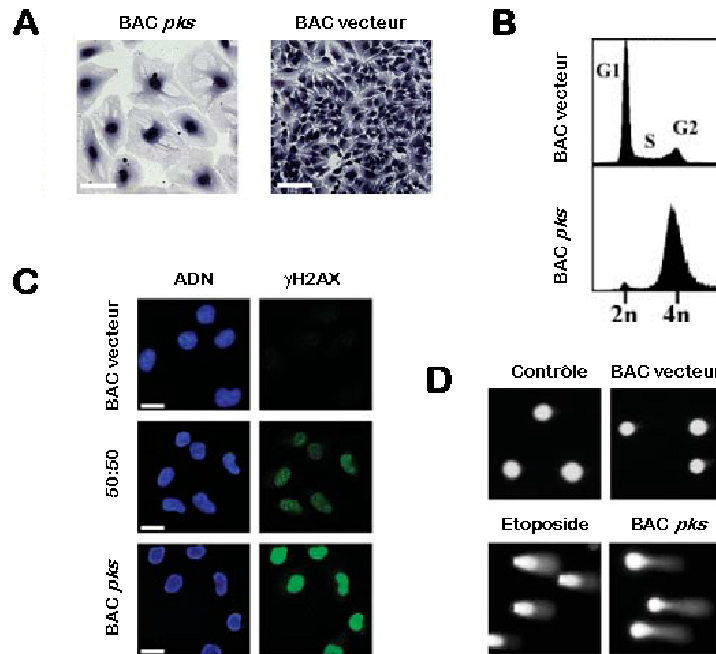
Chez *E. coli*, l'îlot *pks* est toujours retrouvé à proximité de l'îlot génomique HPI (« High Pathogenicity Island »). La réciproque n'est pas vraie. L'îlot HPI est considéré comme un facteur de virulence des ExPEC et permet la biosynthèse de la yersiniabactine, un sidérophore de type NRP. La production de la colibactine est cependant indépendante de celle de la yersiniabactine [24, 246].

## 5. Expression phénotypique et mode d'action

L'infection transitoire de lignées cellulaires de mammifères (cellules tumorales humaines du col de l'utérus HeLa, cellules ovariennes de hamster CHO, cellules de mélanomes humains A375 et cellules non tumorales des cryptes intestinales de rat IEC-6) par des *E. coli* porteuses de l'îlot *pks* induit une mégaloctose caractérisée par un élargissement progressif du corps cellulaire et du noyau, et l'absence de mitose (**figure 15**) [24, 249]. L'activité cytopathique nécessite un contact intime entre la cellule hôte et la bactérie, mais ne requiert pas l'internalisation de la bactérie. Tuées par la chaleur, la sonication ou les antibiotiques, les bactéries ne sont pas cytopathogènes [24].

L'analyse du cycle cellulaire de cellules tumorales infectées montre que le noyau des cellules géantes a une teneur en ADN de 4n (**figure 15**). Cette observation, conjuguée à l'absence de cellules en division indique que les cellules géantes sont bloquées à la transition de phase G2/M (Fig. 3A).

Le blocage du cycle cellulaire est associé à une activation de la protéine ATM et une phosphorylation dose-dépendante de l'histone H2AX (**figure 15**). Ces phénomènes correspondent à l'activation du point de contrôle G2/M en réponse à des cassures double-brin de l'ADN, qui ont été mises en évidence par le test de la comète (**figure 15**). La voie de signalisation conduisant au blocage cellulaire en phase G2/M passe par l'activation de la voie ATM-Chk2 (**figure 16**). De plus, la lignée CHO déficiente pour la protéine Ku80 (protéine impliquée dans la voie de réparation de l'ADN par le système de « Non-Homologue End-Joining ») et donc hypersensible aux cassures doubles brins de l'ADN meurt massivement [250]. Cependant, l'action directe ou indirecte de la colibactine sur l'ADN reste à démontrer.



**Figure 15 - Exposition de cellules HeLa à une infection transitoire à *E. coli pks*+**.

D'après Nougayrède *et al* [24].

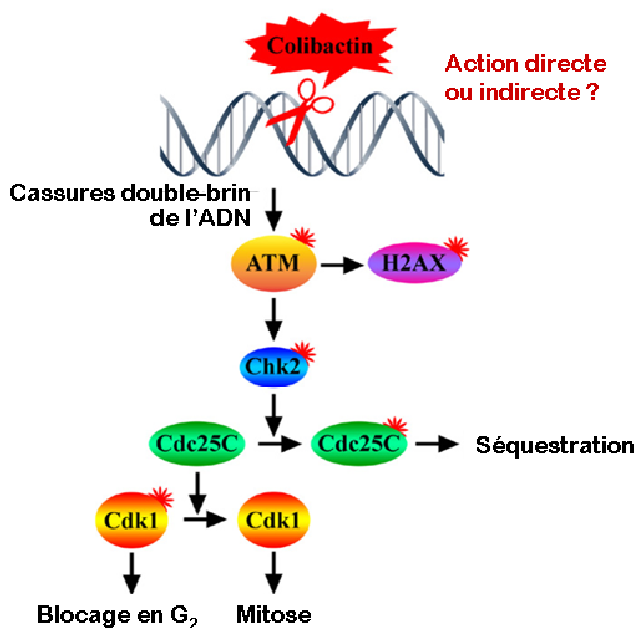
Les cellules HeLa ont été infectées pendant 4 heures avec une souche de *E. coli* K12 portant un BAC contenant l'îlot *pks* ou non (multiplicité d'infection de 100) ou un mélange de chaque souche (50:50).

A. Mégalocytose des cellules révélée par coloration de Giemsa 72 h après infection.

B. Analyse du cycle cellulaire 72 h après infection montrant un arrêt des cellules en G<sub>2</sub>/M.

C. Phosphorylation des histones H2AX, témoin de cassures double-brin de l'ADN, analysée 4 heures après infection.

D. Test de la comète après infection ou traitement avec l'étoposide, un inhibiteur de la topoisomérase II humaine qui induit des cassures double-brin de l'ADN.



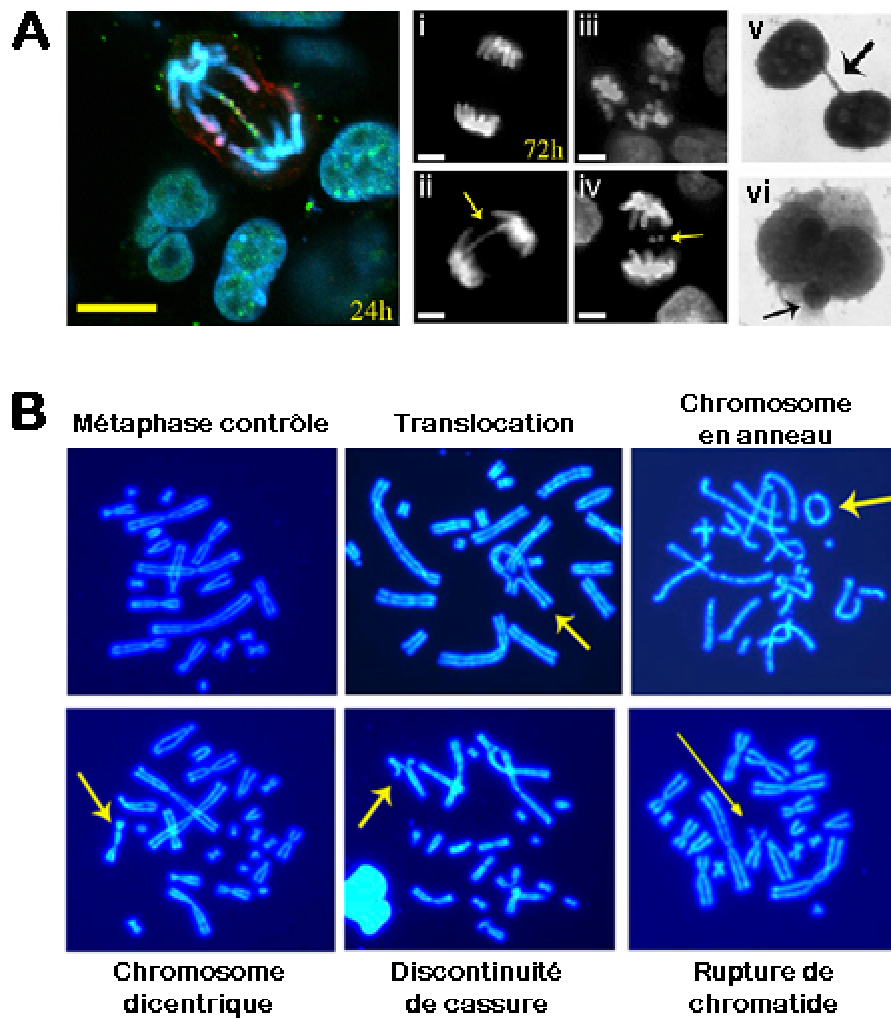
**Figure 16 - Modèle d'activation du point de contrôle G<sub>2</sub>/M dans les cellules HeLa exposées aux souches de *E. coli pks*+**. D'après Nougayrède *et al* [24].

Le complexe enzymatique codé par l'îlot *pks* synthétise un ou des composés de type PK-NRP, désigné « colibactin », qui induit, directement ou indirectement des cassures double brin de l'ADN dans les cellules hôtes eucaryotes. Ces dommages activent la voie de transduction du signal ATM-Chk2, conduisant à la séquestration cytoplasmique de Cdc25C et au défaut de déphosphorylation de Cdk1, induisant finalement un blocage du cycle cellulaire en phase G<sub>2</sub>. Les phosphorylations sont représentées par des astérisques rouges.

## 6. Un agent carcinogène potentiel

Une seule et courte exposition de cellules épithéliales de mammifère en culture (lignée CHO) à de faibles doses infectieuses de *E.coli pks+* induit une réponse cellulaire aux dommages de l'ADN, marquée par un blocage transitoire du cycle. Les cellules reprennent leur division et présentent des signes de réparation incomplets de l'ADN, conduisant à des ponts anaphasiques et des aberrations chromosomiques (**figure 17**) [250]. Dans des cultures de différentes lignées cellulaires, des anomalies du nombre (aneuploïdie) ou de la structure des chromosomes (micronoyaux, chromosomes en anneau et ponts anaphasiques) persistent dans les cellules en division jusqu'à 21 jours après l'infection, indiquant l'apparition de cycles de cassure-fusion-pont ayant un rôle majeur dans l'instabilité chromosomique [250].

L'infection d'entérocytes dans un modèle d'anse intestinale ligaturée de souris par une souche de *E. coli pks+* induit la formation de foyers d'histones H2AX phosphorylés correspondant à une irradiation de 0,5 Grays [250]. Les cellules en culture montrent aussi une augmentation significative de la fréquence des mutations et la formation de colonies cellulaires indépendantes de l'ancrage [250]. Ces données suggèrent une activité mutagène et de transformation cellulaire des *E. coli pks+*.



**Figure 17 - Ponts anaphasiques et aberrations chromosomiques induits par une infection à *E. coli pks+*. D'après Cuevas-Ramos *et al* [250].**

A. Pont anaphasique d'une cellule CHO 24 heures après infection. 72 heures après infection (i) anaphase normale, (ii) pont anaphasique générant des ponts nucléoplasmiques (v), (iii) mitose multipolaire, et (iv) chromosome retardé générant des micronoyaux (vi). Barres : 10  $\mu$ m.

B. Aberrations chromosomiques des cellules métaphasiques 24 heures après infection.

### III. Les polycétides et les peptides non ribosomaux

#### A. Généralités

##### 1. Les métabolites secondaires

Le métabolisme d'un organisme comprend le métabolisme primaire, essentiel à la croissance et au développement, et le métabolisme secondaire, qui est souvent nécessaire à l'interaction de l'organisme avec son environnement (**Tableau 3**) [251]. Chaque voie de biosynthèse du métabolisme secondaire dérive d'un métabolisme primaire et implique des réactions complexes et souvent hautement spécifiques qui aboutissent aux produits finis. Les métabolites secondaires des micro-organismes, des plantes et des organismes marins forment un immense réservoir de diversité chimique naturelle.

Les six voies principales de métabolisme secondaire synthétisent les terpénoïdes, les alcaloïdes, les peptides, les oligosaccharides, les PK et les NRP [251, 252]. Des combinaisons entre ces six voies principales de métabolisme secondaire peuvent se produire et permettent la synthèse de nouveaux composés sans créer de nouvelles voies de biosynthèses comme par exemples les molécules hybrides PK-NRP [253, 254].

L'analyse de l'îlot *pks* montre que la colibactine est un métabolite secondaire hybride de type PK-NRP. La suite de ce chapitre se focalisera donc sur la biosynthèse et les fonctions des composés PK, NRP et des hybrides PK-NRP.

**Tableau 3 : Caractéristiques des métabolismes primaires et secondaires.**  
D'après Hutter [251].

Métabolisme primaire	Métabolisme secondaire
Solution générale à un problème biologique	Solution spécifique à un problème biologique
Nécessaire à la croissance	Non nécessaire à la croissance
Rôle physiologique connu	Rôle physiologique difficile à déterminer
Actif durant toute la vie	Actif uniquement à certains moments du développement
Présent quel que soient les conditions de développement	Très dépendant des conditions de développement
Ubiquitaire	Non ubiquitaire
Molécules finales définies	Familles de molécules proches
Molécules généralement simples	Produits souvent chimiquement complexes

## 2. Intérêt de l'étude des composés d'origine naturelle

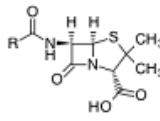
Les structures de plus de 200 000 métabolites secondaires ont été élucidées à ce jour, nous fournissant une complexité et une diversité énorme de squelettes carbonés uniques. Ces métabolites secondaires ont abouti à un arsenal de produits naturels aux propriétés spécifiques qui sont à l'origine de médicaments inestimables [252].

En 1990, environ 80% des médicaments étaient soit des produits naturels, soit des analogues hémi-synthétiques de ceux-ci [255]. L'essentiel de ces médicaments appartient aux familles de composés PK, NRP et PK-NRP comme de nombreux antibiotiques (ex. : les pénicillines, les céphalosporines, les tétracyclines, les macrolides comme l'érythromycine), des antifongiques (ex : l'amphotéricine B et la nystatine), des antiparasitaires (ex. : l'ivermectine), des antipaludéens (ex. : l'artémisinine), des hypocholestérolémiants (ex. : la lovastatine et ses nombreux analogues), des immunosuppresseurs (ex. : la cyclosporine, la rapamycine et le tacrolimus) et des médicaments anticancéreux (ex. : la doxorubicine et la bléomycine) (**figure 18**). Ces composés ont révolutionné la médecine et allongé l'espérance de vie dans une grande partie du monde de près de 40 ans au début du 20<sup>ème</sup> siècle à plus de 77 ans aujourd'hui. Aussi, malgré l'expansion de la chimie médicinale synthétique dans les années 1990, 50% des nouveaux médicaments restent encore aujourd'hui basés sur des produits naturels [256]. Quatre raisons expliquent ce succès : [257] :

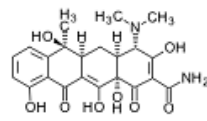
- la conception de produits actifs passe par l'élaboration de composés et de procédés ayant la complexité retrouvée dans les composés naturels,
- des siècles d'évolution ont permis d'optimiser la structure des produits naturels, conduisant souvent à des mécanismes d'adressage à la cible efficaces, une affinité sub-nanomolaire pour les cibles et une grande spécificité,
- de très nombreux produits naturels restent à découvrir par l'exploration de nouvelles niches écologiques et de nouveaux taxons microbiens,
- les progrès technologiques du séquençage, de la manipulation et de la reconstitution des voies de biosynthèses offrent également de nouvelles perspectives.



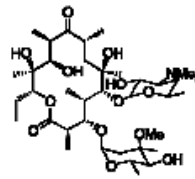
## Anti-infectieux



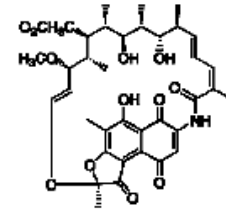
**Pénicilline (NRP)**  
*Penicillium chrysogenum*  
Antibiotique



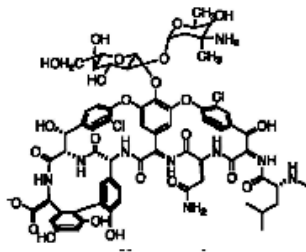
**Tétracycline (PK itératif)**  
*Streptomyces aureofaciens*  
Antibiotique



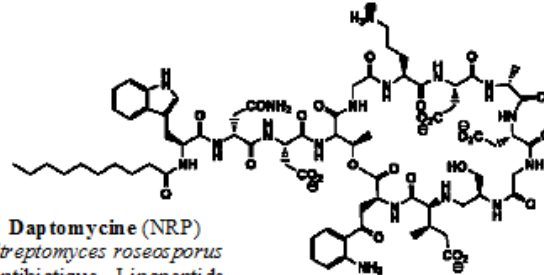
**Erythromycine A (PK)**  
*Saccharopolyspora erythraea*  
Antibiotique - Macrolide



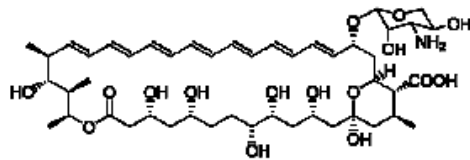
**Rifamycine S (PK)**  
*Amycolatopsis mediterranei*  
Antibiotique - Ansamycine



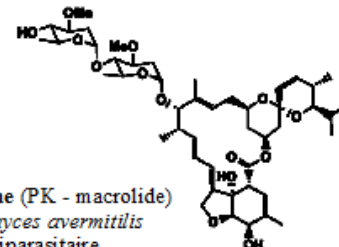
**Vancomycine (NRP)**  
*Amycolatopsis orientalis*  
Antibiotique - Glycopeptide



**Daptomycine (NRP)**  
*Streptomyces roseosporus*  
Antibiotique - Lipopeptide

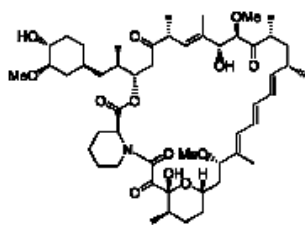


**Amphotéricine B (PK)**  
*Streptomyces nodosus*  
Antifongique

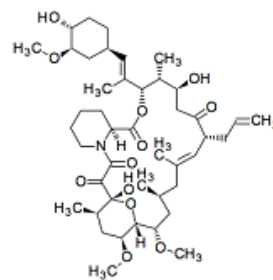


**Avermectine (PK - macrolide)**  
*Streptomyces avermitilis*  
Antiparasitaire

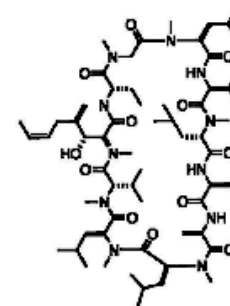
## Immunosuppresseurs



**Rapamycin - Sirolimus (PK-NRP)**  
*Streptomyces hygroscopicus*

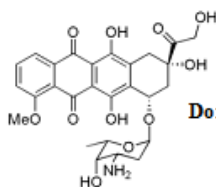


**Tacrolimus - FK 506 (PK)**  
*Streptomyces tsukubaensis*



**Cyclosporine A (NRP)**  
*Tolypocladium inflatum*

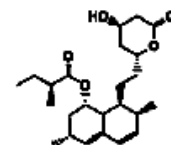
## Antitumoral



**Doxorubicine - Adriamycine (PK itératif)**  
*Streptomyces peucetius*  
Anthracycline

## Hypocholestérolémiant

**Lovastatine (PK itératif)**  
*Aspergillus terreus*



**Figure 18 - Exemples de métabolites secondaires de type PK, NRP et hybrides utilisés en médecine.**

Les structures, les micro-organismes producteurs et les activités pharmacologiques des composés sont indiqués.

## **B. La synthèse des composés PK et NRP**

La biosynthèse des PK a été observée chez les bactéries, les champignons, les plantes et les animaux, et celle des NRP principalement chez des micro-organismes tels que les bactéries et les champignons [257, 258].

Les composés de type PK, NRP et PK-NRP (ensemble de composés désigné ci-après NRPK) sont des oligomères synthétisés par des complexes multienzymatiques incluant des mégasynthases (PK synthases ou PKS, NRP synthases ou NRPS et les PK-NRP synthases hybrides ou PKS-NRPS) et des enzymes discrètes. Les gènes codant ces voies métaboliques secondaires sont organisés en *clusters*, permettant une régulation coordonnée de la synthèse des composés à la demande ou « à flux tendu » [259, 260]. Ces complexes multienzymatiques, souvent appelés modules, assurent (i) l'activation de monomères destinés à l'assemblage, (ii) l'assemblage des monomères en un oligomère linéaire par des étapes itératives de condensation chimique, à l'image de la synthèse des protéines par le ribosome, et (iii) la modification des monomères en cours ou après leur assemblage. Ces protéines agissent de façon coordonnée selon une organisation évoquant une chaîne d'assemblage d'automobiles, et très fréquemment dans un ordre colinéaire à l'organisation génique [25, 261]. Les enzymes discrètes modifient les monomères ou oligomères pendant ou après l'assemblage. Ainsi, la structure unique des composés NRPK repose sur le nombre de modules enzymatiques mis en jeu, l'ordre d'action des enzymes et la chimie catalysée par chaque enzyme dans et après la chaîne d'assemblage des monomères [25].

## 1. La logique chimique

L'assemblage des NRPK repose sur l'assemblage itératif de monomères d'acides et/ou d'acides aminés en un oligomère, qui subit en sus des modifications enzymatiques complexes.

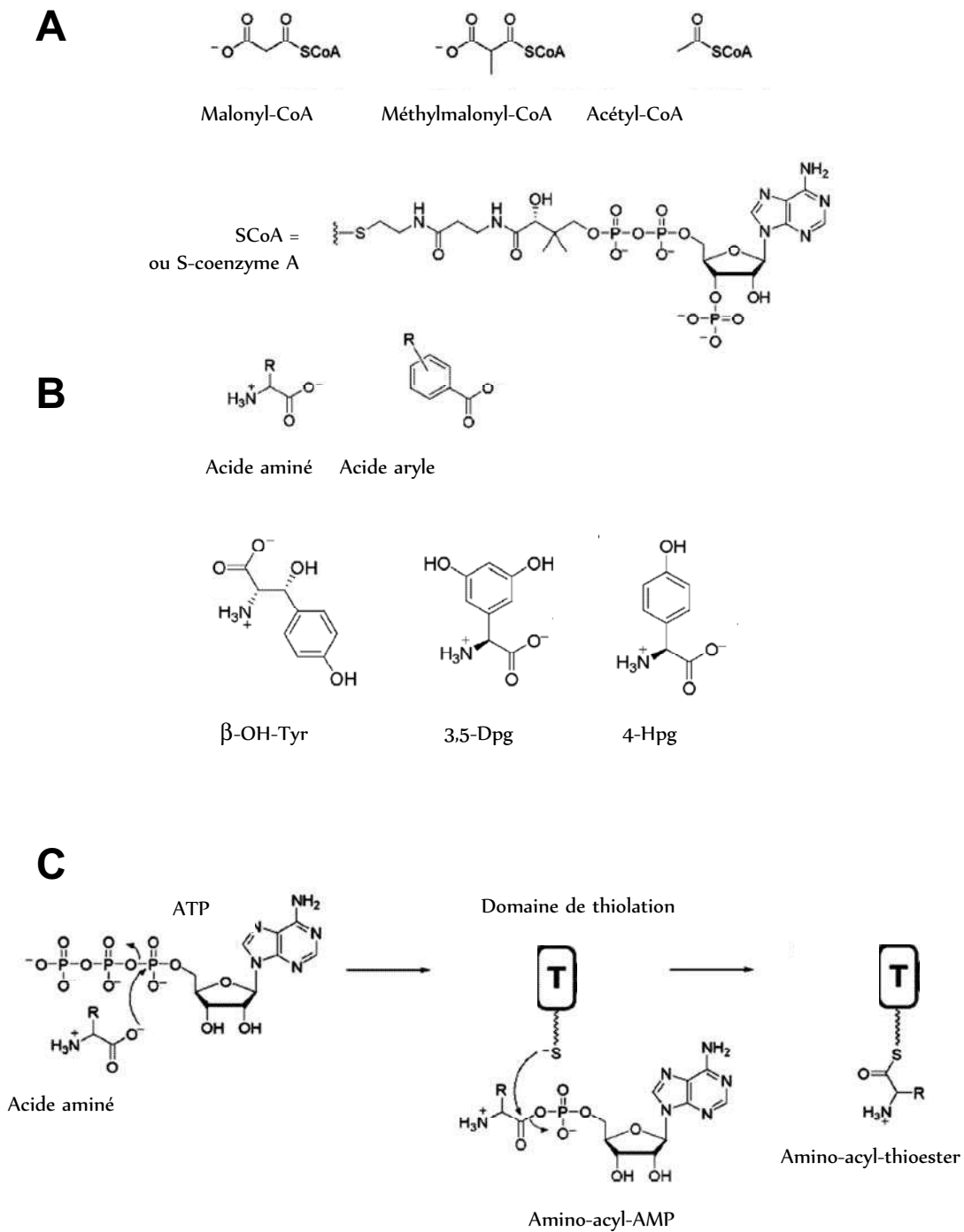
### a) L'activation des monomères

Les monomères qui composent les PK et les NRP sont issus du métabolisme primaire. D'autres monomères comme certains acides aminés non protéinogènes sont synthétisés à partir des métabolites primaires par des voies métaboliques secondaires fréquemment codées dans le *cluster* comprenant les gènes des mégasynthèses. L'utilisation de ces monomères repose sur leur activation sous forme de thioesters [261, 262].

Les composés PK sont constitués de monomères d'acyles portant un groupe carboxyle activé permettant la condensation chimique avec l'oligomère en croissance. L'activation repose sur une liaison au coenzyme A (CoA) par une liaison thioester (**Figures 19A**). Les unités acyles sont des monoacyles comme l'acétyl-, le propionyl- et le benzoyl-CoA, ou des variants, tels que le malonamyl-CoA ou le méthoxymalonyl-CoA [263]. Le malonyl-CoA et le méthylmalonyl-CoA sont les unités monomériques les plus fréquemment incorporées pendant l'élongation de la chaîne des PK.

Les NRP sont synthétisés à partir des 20 acides aminés protéinogènes, d'acides aminés non protéinogènes et d'acides aryles (**Figure 19B**) [264, 265]. Au moins 350 monomères de ce type sont dénombrés à ce jour. Ils sont à l'origine de l'essentiel de la diversité des composés NRP. Leur activation repose sur la synthèse d'anhydrides couplés à l'AMP (**Figure 19C**).

Les monomères activés des PK et des NRP sont ensuite chargés sur les domaines de thiolation T des synthèses PKS et NRPS pour former respectivement des acyl- et aminoacyl-thioesters conservant l'énergie d'activation (**figure 19C**) [266].



**Figure 19 - Utilisation de monomères thioesters par les PKS et les NRPS.**

D'après Fischbach et Walsh [25].

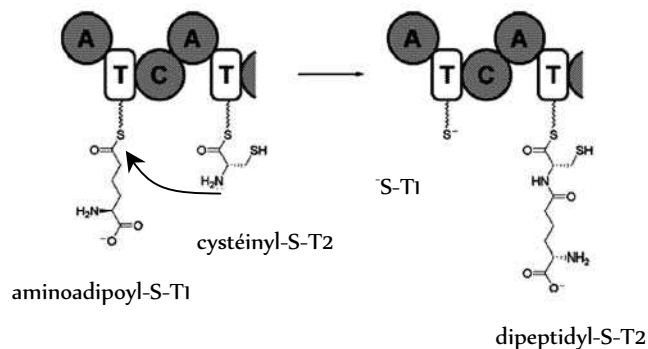
A. Acyl-CoA thioesters couramment utilisés par les PKS pour l'initiation et l'élongation de la chaîne.

B. Monomères d'amino- et d'aryl-acides génériques et exemples d'acides aminés non protéinogènes (β-hydroxytyrosine (β-OH-Tyr), 3,5-dihydroxyphénylglycine (3,5-Dpg), 4-hydroxyphénylglycine (4-Hpg)) utilisés par les NRPS pour l'initiation et l'élongation de la chaîne.

C. Formation d'amino-acyl-thioester par les NRPS à partir d'amino-acyl-AMP.

## b) L'élargissement des oligomères

L'élargissement de l'oligomère repose la formation d'une liaison covalente avec un monomère activé à chaque transfert de l'oligomère le long de la chaîne enzymatique d'assemblage [261, 262]. Ainsi, l'oligomère est transféré de la synthase n-1 sur l'acyl- ou l'aminoacyl-thioesters de la synthase n, qui catalyse la condensation de l'oligomère avec l'acyl- ou l'aminoacyl-thioesters (**Figure 20**). La condensation des acyl-thioesters forment les liaisons C-C des oligomères PK. La condensation des aminoacyl-thioesters forment les liaisons C-N (liaison peptidique) des oligomères NRP. Ce processus est répété tout au long de la chaîne d'assemblage selon un ordre guidé par l'agencement des synthases.



**Figure 20 - Transfert et condensation des aminoacyl-thioesters lors de la formation d'un NRP.** D'après Fischbach et Walsh [25].

L'aminoadipoyl-S-T1 est le donneur, et le cystéinyl-S-T2 l'accepteur pour la formation de la liaison peptidique, pour donner la protéine déacylée S-T1 et le dipeptidyl-S-T2. Domaines d'adénylation (A), de thiolation (T) et de condensation (C).

## c) La modification des oligomères

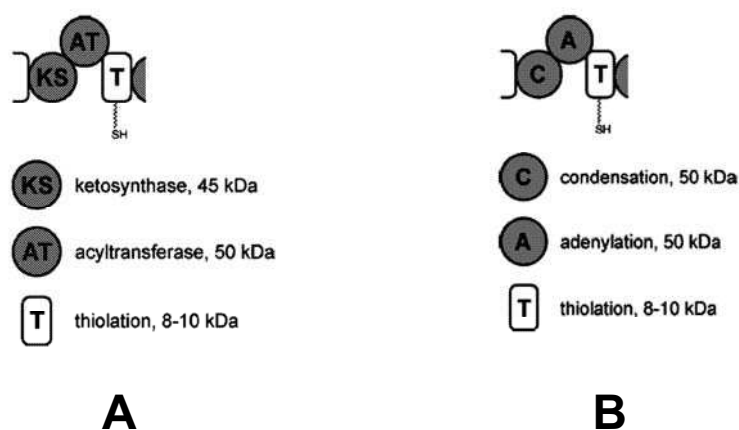
L'oligomère peut être modifié sur la chaîne d'assemblage multienzymatique au cours des différentes phases d'élargissement ou après celles-ci. Ces modifications sont assurées en cours d'élargissement par des domaines enzymatiques optionnels des mégasynthèses, qui peuvent assurer notamment l'épimérisation, la transméthylation des composés NRP ou la réduction et la déshydratation des composés PK. Des enzymes discrètes, dites de « finition », et fréquemment codées dans le *cluster* comprenant les gènes des mégasynthèses peuvent également modifier l'oligomère au cours de sa synthèse ou après son assemblage. Ces modifications nombreuses et complexes de l'oligomère sont à l'origine de la diversité des composés, notamment de type PK, et sont essentielles à leur activité biologique [25].

## 2. Les machineries enzymatiques

Les mégasynthèses PKS, NRPS ou PK-NRPS assurent l'essentiel de la synthèse du squelette carboné. Leur activité repose sur trois domaines de base [25, 261, 262] (**Figure 21**) :

- 1- le domaine Thiolation (T) de fixation des monomères,
- 2- le domaine KetoSynthase (KS) des PKS ou le domaine de Condensation (C) des NRPS,
- 3- le domaine AcylTransférerase (AT) des PKS ou le domaine d'Adénylation (A) des NRPS.

L'association de ces trois types de domaine sur une ou plusieurs chaînes peptidiques définit un module enzymatique PKS ou NRPS.

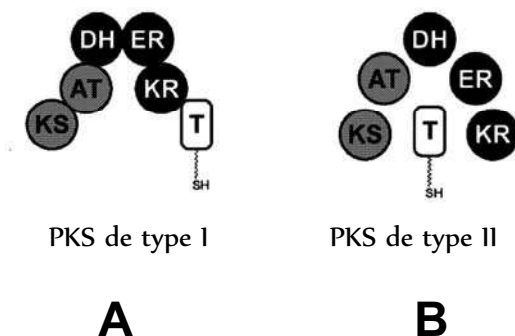


**Figure 21 - Domaines essentiels constituant les modules des PKS (A) et des NRPS (B).**  
D'après Fischbach et Walsh [25].

### a) Les chaînes d'assemblage des composés PK

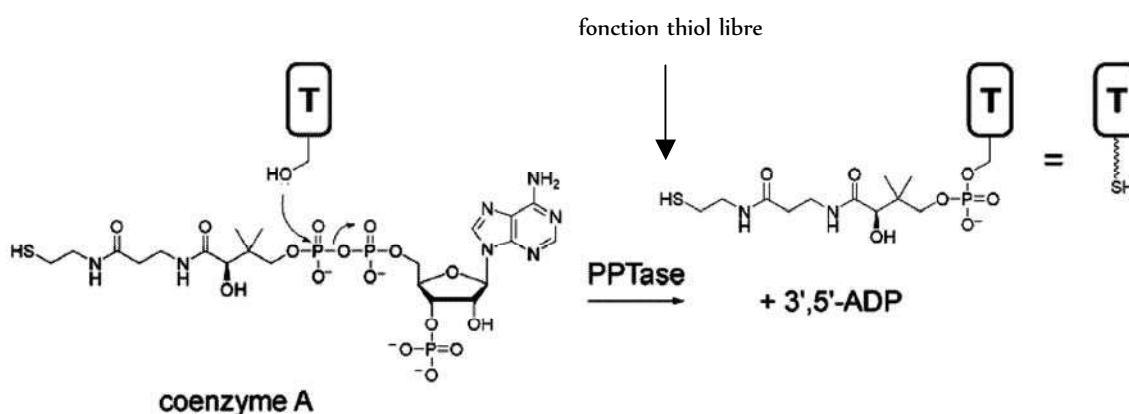
#### *(1) Organisation standard des PKS*

Trois types de PKS ont été identifiés. Dans les PKS de type I (modulaires), les domaines KS, AT et T sont fusionnés en *cis* pour former des modules, qui peuvent également être fusionnés en une chaîne peptidique pour former une mégasynthase multimodulaire. L'ordre typique des domaines sur le module enzymatique est KS-AT-T. Dans les systèmes PKS de type II (itératifs), les domaines existent sous forme de protéines distinctes interagissant en *trans* pour former des complexes fonctionnels multimériques (**Figure 22**) [25]. Enfin, les PKS de type III et apparentées se distinguent par l'absence du domaine T [267].



**Figure 22 - Comparaison des PKS de type I et de type II.** D'après Fischbach et Walsh [25].  
 A. Dans les systèmes de type I, les domaines protéiques sont connectés en *cis*.  
 B. Dans les systèmes de type II, les domaines protéiques interagissent en *trans*.

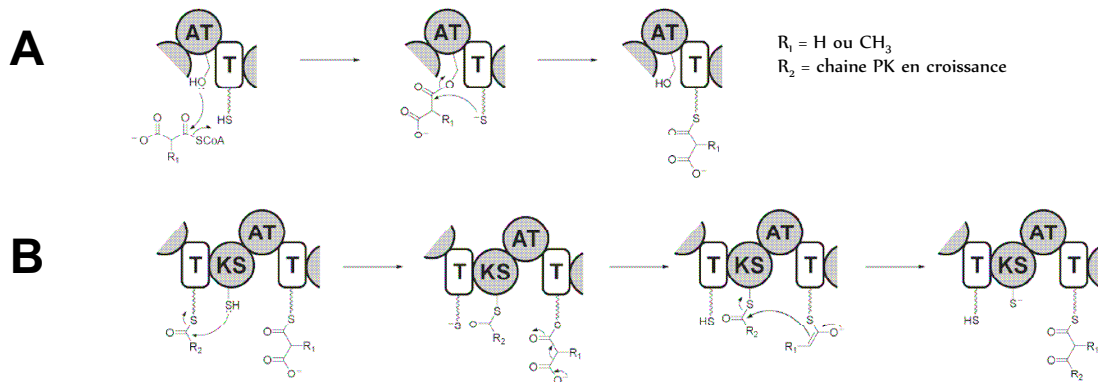
Les domaines T des PKS de type I et de type II sont souvent désignés sous le nom de protéines portant la chaîne acyle ou ACP pour « Acyl Carrier Protein ». La chaîne latérale d'une sérine spécifique du domaine T est modifiée post-traductionnellement par la fixation d'un bras flexible pantéthéinyle terminé par un groupe thiol libre (**Figure 23**). Cet amorçage des domaines apo-T en domaines holo-T fonctionnels est effectué par des PhosphoPantéthéinyl-Transférases (PPTases), souvent codées dans les *clusters* de gènes codant les mégasynthèses correspondantes [266, 268]. A partir d'un acyl-CoA activé, le site holo-T peut alors fixer, grâce à la fonction thiol libre un monomère acyle pour former un acyl-thioester prêt à être assemblé.



**Figure 23 - Modification post-traductionnelle des domaines de Thiolation T par les PhosphoPantéthéinyl-Transférases (PPTases).** D'après Fischbach et Walsh [25].

Les membres de cette classe d'enzymes catalysent le transfert d'un groupement PhosphoPantéthéinyl à partir d'un Coenzyme A vers une sérine conservée du domaine T. Cet « amorçage » des domaines T des chaînes d'assemblage PKS et NRPS est identique.

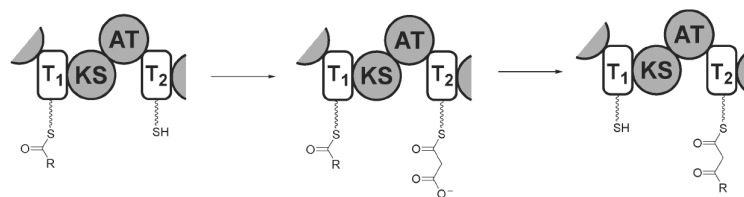
Le domaine AT sélectionne un acyl-CoA spécifique comme substrat. C'est le gardien de la spécificité des monomères d'acyle incorporés. Il assure la fixation du groupement acyle à l'extrémité thiolate du bras pantéthéinyle du domaine holo-T pour former un acyl-thioester (acyl-S-T) activé et prêt à être assemblé (**figure 24A**). Le domaine KS est le catalyseur permettant le transfert et la condensation de l'oligomère en croissance avec les monomères selon un mécanisme en 3 phases, détaillé dans la **figure 24B**.



**Figure 24 - Activités des domaines AT et KS des PKS.** D'après Fischbach et Walsh [25].

A. Le domaine AT sélectionne le monomère acyl-CoA à incorporer et catalyse sa transthiolation au domaine T en aval via un intermédiaire acyl-*O*-enzyme.  
 B. Le domaine KS catalyse la formation d'une liaison C-C entre l'acyl-thioester en amont (à gauche), transitoirement lié à KS et l'accepteur acyl carbanion en aval (à droite) résultant de la décarboxylation du acyl-S-T ajouté.

Au cours d'un assemblage utilisant un système PKS de type I, la chaîne acyle en croissance est transloquée du domaine  $T_{n-1}$  vers le domaine  $T_n$  à chaque action d'un domaine KS (**Figure 25A**). Le plus souvent, les domaines KS et T sont utilisés une seule fois. Les systèmes PKS de type II itératif utilisent un seul domaine T, AT, et KS pour toutes les étapes d'élongation. La chaîne s'accroît à chaque cycle, mais reste attachée au même domaine T, et les domaines AT et KS agissent de manière itérative (**figure 25B**). Dans ces systèmes itératifs, le domaine KS est un hétérodimère d'une sous-unité active ( $KS\alpha$ ) et d'une sous-unité inactive CLF (pour « Chain Length Factor »), qui forme un tunnel amphipathique dont la taille contrôle le nombre d'itération réalisé [269].



**Figure 25 - Elongation de la chaîne dans les systèmes PKS de type I.**

D'après Fischbach et Walsh [25].

Chaque condensation résulte en une addition de deux ou trois carbones sur la chaîne naissante, qui est simultanément transférée à partir d'un domaine T au suivant dans les PKS de type I.



### (2) Les modules d'initiation et de terminaison

Les modules enzymatiques d'initiation et de terminaison fonctionnent différemment des modules d'élongation décrits ci-dessus [25, 261, 262].

Un acyl-CoA est sélectionné par le domaine AT d'initiation et chargé sur le premier domaine T. Le domaine KS est ainsi généralement absent du module d'initiation de la chaîne d'assemblage ou il y est non fonctionnel.

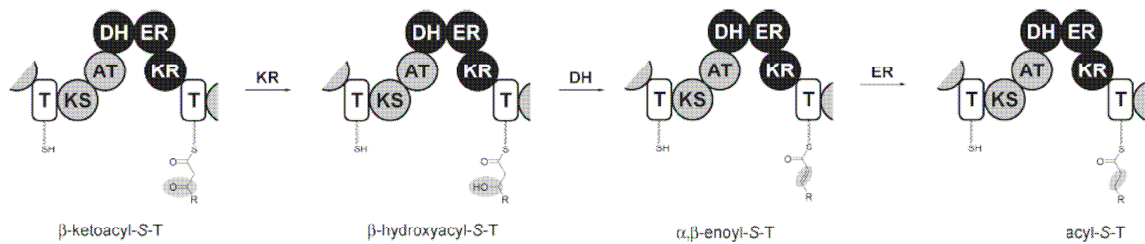
Le dernier module d'une chaîne d'assemblage PKS entraîne la rupture de la liaison thioester entre le domaine T final et l'oligomère terminé. Ce processus d'hydrolyse libérant le composé est généralement catalysé par un domaine thioestérase (TE) de 35 kDa. Celle-ci génère un acide libre à l'extrémité de l'oligomère ou une lactone cyclique par condensation avec une fonction hydroxyle de la chaîne PK [270].

### (3) Les domaines optionnels et enzymes de finition

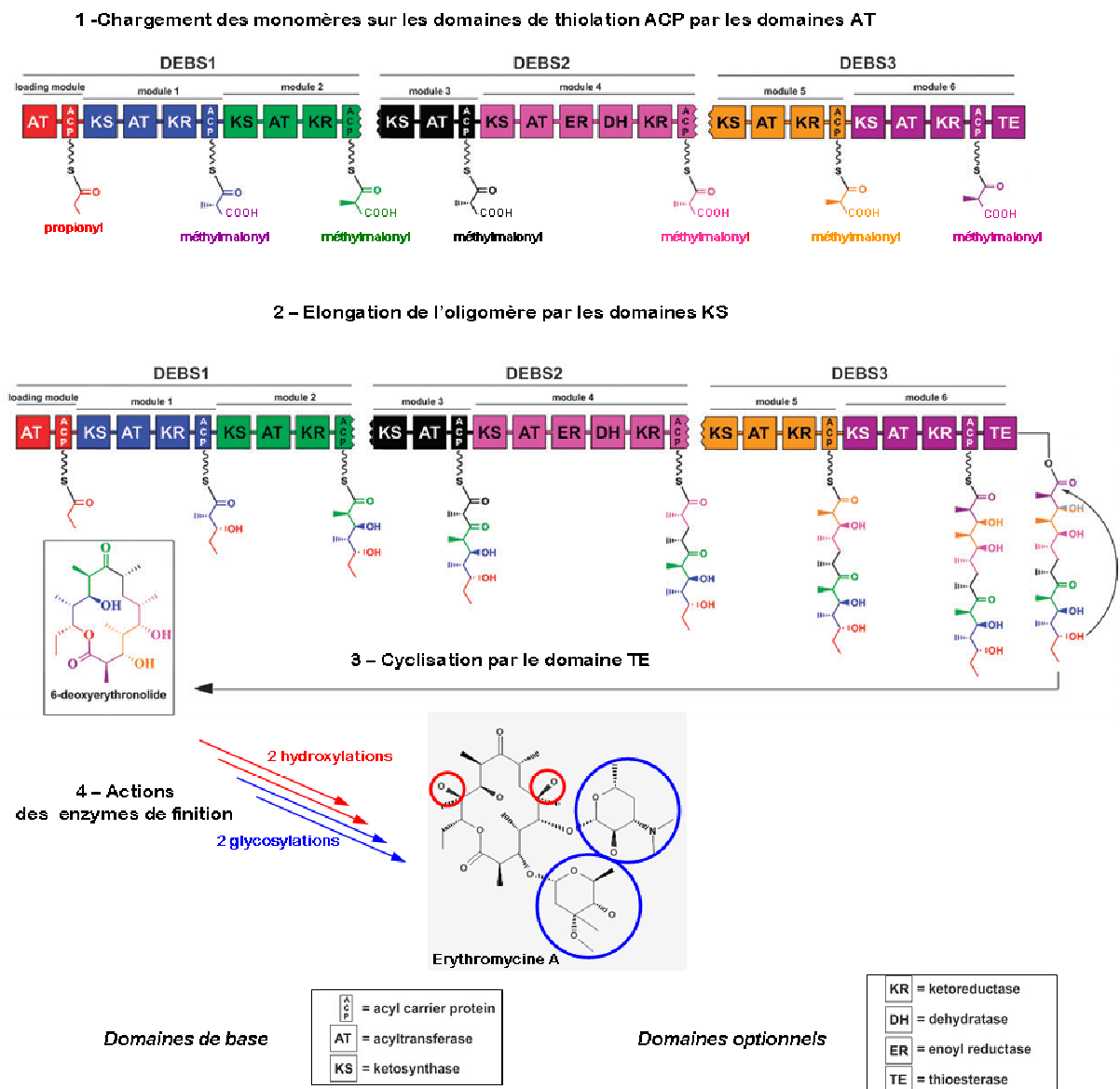
Les domaines facultatifs des PKS de type I et II sont les domaines catalytiques kétéroréductases (KR), déshydratases (DH) et enoylréductases (ER) [25, 271]. Ils opèrent suivant la séquence KR > DH > ER ; le  $\beta$ -kétolacyl-S-T est d'abord réduit en  $\beta$ -hydroxyacyl-S-T par l'action KR, puis ce dernier est déshydraté pour former un  $\alpha,\beta$ -enoyl-S-T par l'action DH, et enfin ce conjugué oléfine est saturé pour donner un acyl-S-T saturé suite à l'action ER (**Figure 26**). Ces trois domaines enzymatiques catalysent au total la réduction du groupement cétone  $\beta$ -C=O en méthyle  $\beta$ -CH<sub>2</sub>. Cependant, cette réduction est rarement complète et de nombreux modules préservent les fractions intermédiaires  $\beta$ -hydroxyacyle ou  $\alpha,\beta$ -enoyle en raison de l'absence ou de la non-fonctionnalité de domaines ER, DH et/ou KR. Par ailleurs, des modules C-méthyltransférases ainsi que des modules aminotransférases peuvent convertir un  $\beta$ -kétolacyl-S-T en  $\beta$ -aminoacyl-S-T [272, 273].

Enfin des enzymes de finition peuvent agir en *trans* sur les intermédiaires acyl-S-T en croissance, afin de catalyser des réactions de déshydratation, de décarboxylation, de réduction ou d'addition de carbanions [25].

L'organisation standard des chaînes d'assemblage de PK de type I peut être illustrée par celle du 6-dEB (précurseur inactif de l'antibiotique érythromycine A), système modèle et historique de PKS modulaires (**figure 27**) [274, 275].



**Figure 26 - Transformation du carbone  $\beta$  dans les PKS.** D'après Fischbach et Walsh [25]. Le produit de condensation  $\beta$ -kétacyl-S-T est réduit par un domaine kétéductase (KR) pour former un  $\beta$ -hydroxyacyl-S-T, qui est déshydraté par un domaine déshydratase (DH) générant un  $\alpha,\beta$ -enoyl-S-T (conjugué oléfine), qui est enfin réduit par un domaine enoylréductase (ER) pour donner un acyl-S-T saturé.



**Figure 27 : Organisation des PKS de type I ou modulaires synthétisant le 6-dEB.** D'après Meier et Burkart [261]. La première ligne indique les protéines multienzymatiques (DEBS : 6-dEB synthèses) ; la deuxième ligne indique les modules constitués d'un nombre de domaines enzymatiques variable sur les différentes protéines (« Loading module » : module d'initiation ou de chargement ne possédant pas de domaine KS). Le domaine TE cyclise le PK en formant une lactone par condensation de la fonction acide libérée de la liaison thioester du domaine ACP final avec la dernière fonction hydroxyle de la chaîne PK. Des enzymes de finition catalysent 2 hydroxylations et 2 glycosylations pour générer le composé final actif, l'érythromycine A.

## b) Les chaînes d'assemblage des composés NRP

### *(1) Organisation standard des NRPS*

Trois types de NRPS sont décrits : les NRPS de type A (ou type linéaire) dont les modules sont utilisés une seule fois, les NRPS de type B (ou type itératif), dont tous les modules sont utilisés plusieurs fois et le type C (ou type non linéaire), dont certains domaines sont absents et d'autres fonctionnent plus d'une fois [276].

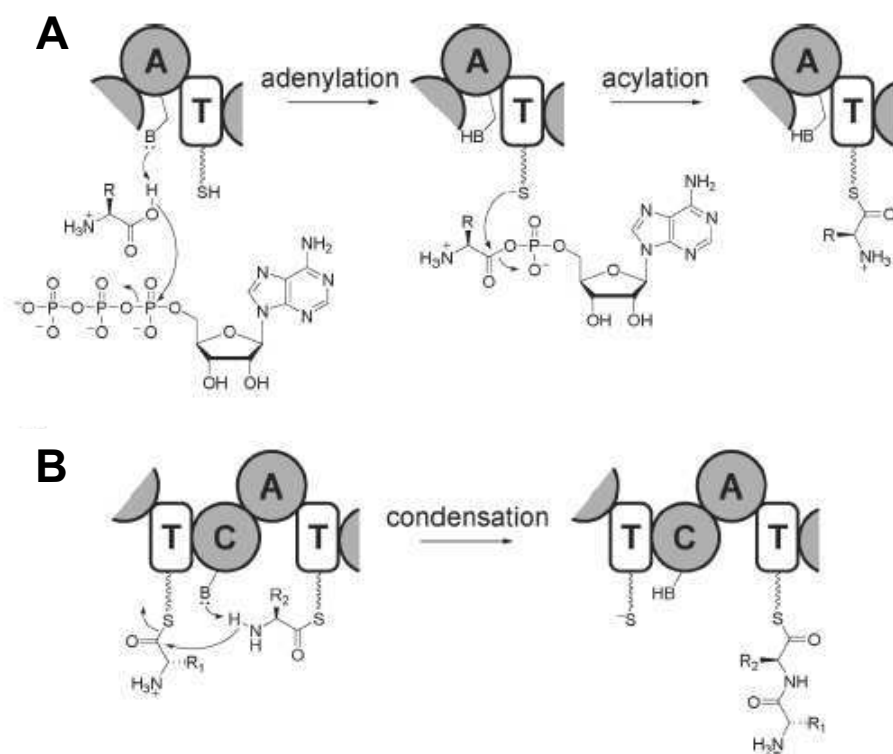
Les NRPS suivent les préceptes généraux de la logique des PKS de type I. Le mécanisme de biosynthèse est composé de chaînes d'assemblage enzymatiques multimodulaires qui contiennent un module pour chaque monomère d'acide aminé incorporé. Par analogie avec les trois domaines d'un module PKS (KS-AT-T), il existe trois domaines de base dans un module minimal de NRPS (**Figure 28**) [25, 261, 262] :

- 1- le domaine T (ou PCP pour « peptidyl carrier protein ») supporte la chaîne peptidique en croissance attachée à un groupement phosphopantéthéinyle ;
- 2- le domaine d'adénylation A est responsable de la sélection du monomère ;
- 3- le domaine de condensation C assure l'élongation de la chaîne peptidique.

L'ordre typique des domaines sur le module multienzymatique NRPS est C-A-T.

### *(2) Les modules d'initiation et de terminaison*

Le module d'initiation des lignes d'assemblage NRPS est généralement un module à deux domaines « A-T<sub>1</sub> » qui sélectionne le premier acide aminé et l'installe de façon covalente sur T<sub>1</sub> [25, 261, 262]. Le premier module peut présenter trois domaines C-A-T dans les chaînes d'assemblage synthétisant des NRP N-acylés ou lipopeptides, comme la daptomycine (**figure 18**). Ce type de domaine C spécifique, appelé « Starter C » réalise le processus dit de lipoinitiation : il fixe un acide carboxylique  $\beta$ -hydroxylé (typiquement un acide gras  $\beta$ -hydroxylé) au premier acide aminé porté par le premier domaine T [277, 278].



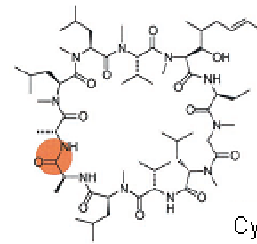
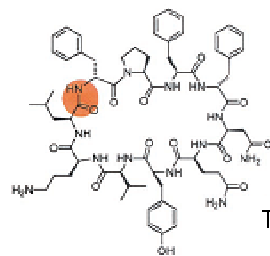
**Figure 28 - Fonctionnement du domaine T et activités des domaines A et C des NRPS.** D'après Fischbach et Walsh [25].

Le fonctionnement des domaines T des chaînes d'assemblage NRPS est équivalent à ceux des lignes d'assemblage des PKS. Chaque domaine apo-T doit être amorcé par un groupement phosphopantéthéine. Lors de l'assemblage du NRP, les résidus aminoacyl et les chaînes peptidiques en croissance sont liés par un pont thioester au thiolate terminal du groupement prosthétique phosphopantéthéine.

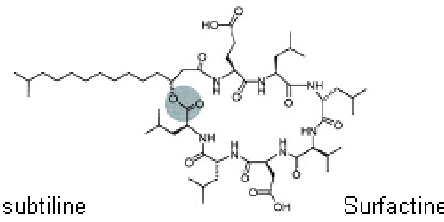
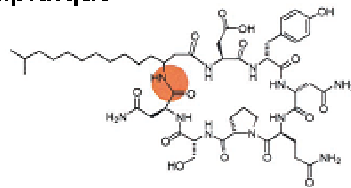
- A. Le domaine A, par analogie avec le domaine AT des PKS sélectionne l'acide aminé à incorporer, active son groupement carboxylate avec un ATP pour créer un aminoacyl-AMP (adénylation), puis fixe le résidu aminoacyl sur le thiolate du domaine T adjacent en aval (acylation).
- B. Le domaine C, par analogie avec le domaine KS des PKS catalyse l'élongation de la chaîne peptidique, formant une liaison N-C entre le peptidyl-S-T<sub>n-1</sub> amont (à gauche) à l'aminoacyl-S-T<sub>n</sub> aval (à droite). Cet allongement de la chaîne se produit de façon concomitante à la translocation de la chaîne en croissance vers le T<sub>n</sub> et préserve le radical aminé libre du résidu N-terminal.

Comme dans les PKS de type I, les modules de terminaison portent un domaine thioesterase (TE) et ont généralement une organisation C-A-T-TE. Le domaine TE peut être hydrolytique ou cyclisant comme observé dans les chaînes d'assemblage PKS [279]. La cyclisation est souvent un caractère fondamental car elle rigidifie la conformation du composé, permet des interactions spécifiques avec les cibles biologiques et dans le cas des NRP, confère une résistance aux protéases. La macrocyclisation des NRP peut se réaliser suivant quatre processus différents (**figure 29**) [280]. La libération de l'oligomère sous forme linéaire ou cyclique peut aussi être médiée par une réductase [281].

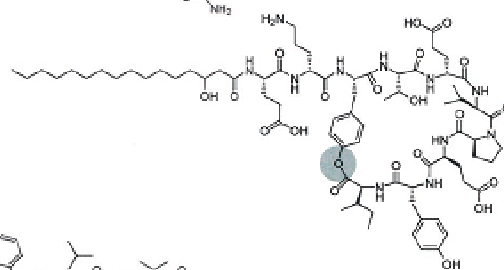
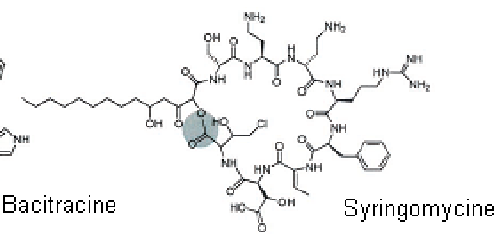
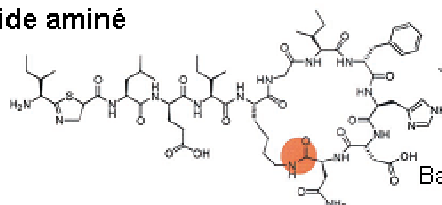
**Cyclisation en tête à queue**



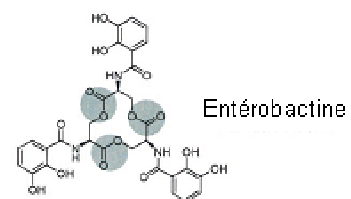
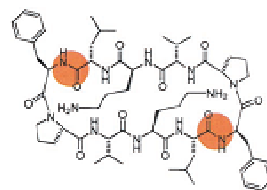
**Cyclisation avec une chaîne lipidique branchée**



**Cyclisation avec une chaîne latérale d'acide aminé**



**Oligomérisation**



● Macrolactame    ● Macrolactone

**Figure 29 - Les quatre processus de macrocyclisation des NRP.**  
D'après Sieber et Marahiel [280].

### (3) Les domaines optionnels

Les composés NRP peuvent subir en cours de synthèse des modifications par des domaines enzymatiques optionnels des enzymes NRPS.

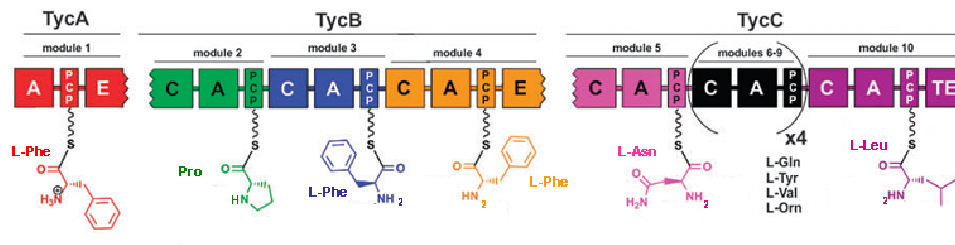
Une caractéristique des NRP est l'abondance de résidus acides aminés de la série D. Ces acides aminés sont générés par un système d'épimérisation des énantiomères usuels L en énantiomères D [277]. Ce système est composé d'un domaine épimérase E associé aux domaines standards dans l'ordre C-A-T-E qui modifie le dernier acide aminé incorporé [282, 283]. En aval du module C-A-T-E sur la chaîne d'assemblage, un domaine de condensation C spécifique, appelé  $^D C_L$  (donneur d'énantiomère D et accepteur d'énantiomère L) lie un acide aminé L à la chaîne peptidique en croissance terminée par l'acide aminé D modifié, contrairement aux domaines de condensation usuels  $^L C_L$ , accepteurs et donneurs d'énantiomères L. Certains domaines C couplent les deux fonctions de condensation et d'épimérisation et sont appelés domaines « dual C/E » [284]. L'épimérisation des acides aminés est parfois réalisée par une racémase discrète agissant en *trans* [285].

D'autres variants du domaine C appelés domaines Cy (pour cyclodéshydratation) permettent l'ajout et la déshydratation de résidus Cys, Ser ou Thr pour former des hétérocycles à cinq atomes (cycle thiazoline (Cys) ou oxazoline (Ser ou Thr)) [277, 286]. Ces hétérocycles peuvent être ensuite modifiés par d'autres enzymes ou domaines. Le domaine oxydase (Ox) est impliqué dans l'oxydation des cycles thiazolines facilement hydrolysables, en hétérocycles thiazoles stables [287, 288]. De même, des réductases peuvent transformer les cycles thiazolines en cycle tétrahydro-thiazolines [289].

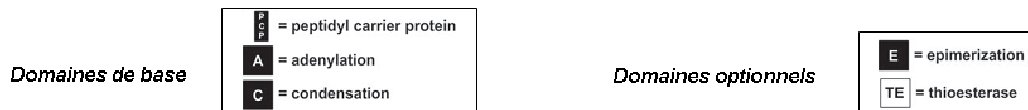
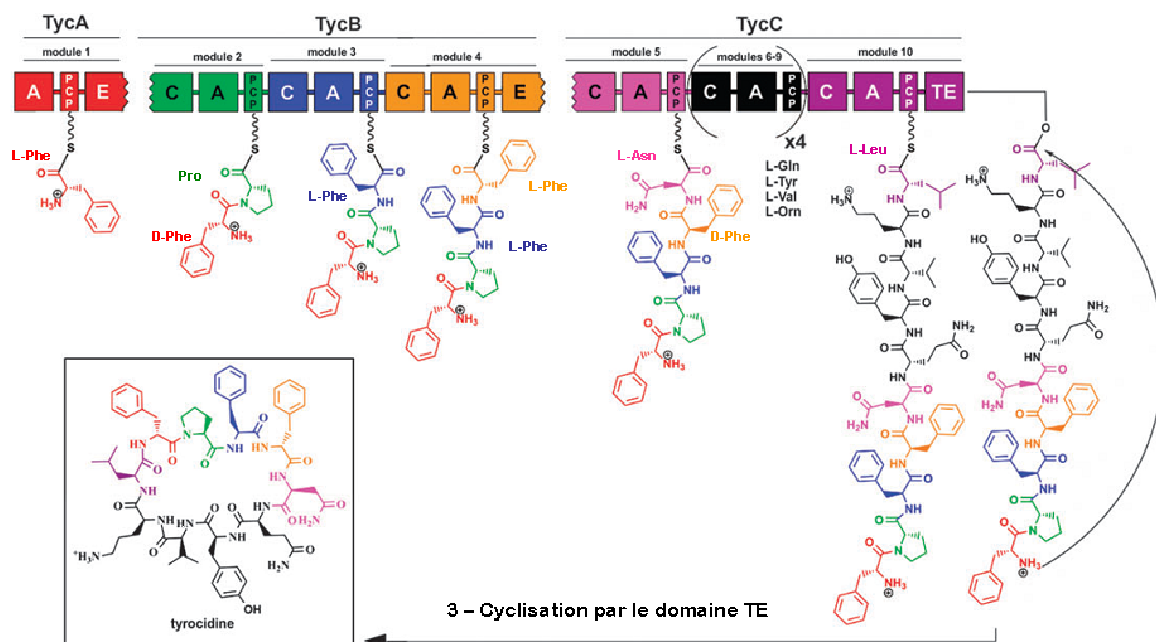
De nombreux NRP contiennent des résidus d'acides aminés méthylés suite à l'action de domaines N-méthyltransférases. Plus rarement des domaines C-méthyltransférases peuvent agir sur des hétérocycles thiazolines [290, 291].

L'organisation des chaînes d'assemblage de NRP peut être illustrée par celle de la tyrocidine, un NRP cyclique (**figure 30**).

1 - Chargement des monomères sur les domaines de thiolation PCP par les domaines A



2 - Elongation de l'oligomère par les domaines C



**Figure 30 - Organisation des NRPS linéaires synthétisant la tyrocidine.**

D'après Meier et Burkart [261].

La première ligne indique les protéines multienzymatiques (mégasynthèses TycA, TycB et TycC) ; la deuxième ligne indique les modules constitués d'un nombre de domaines enzymatiques variable sur les différentes protéines : module 1 d'initiation ne possédant pas de domaine C ; modules 1 et 4 possédant un domaine E ; module 10 de terminaison possédant un domaine TE qui cyclise le NRP en formant un macrolactame par condensation de la fonction acide libérée de la liaison thioester du domaine PCP final avec la fonction amine N-terminale de la chaîne NRP.

### c) Les chaînes d'assemblages hybrides PKS-NRPS

Les chaînes d'assemblage hybrides contiennent des modules enzymatiques de type PKS et NRPS. Ces modules PKS et NRPS peuvent aussi coexister sur une même protéine et sont alors appelés PKS-NRPS [292-295]. Le fonctionnement de ces chaînes d'assemblage repose probablement sur les principes évoqués précédemment. Des investigations complémentaires sont encore nécessaires, notamment pour expliquer le transfert de l'oligomère et sa condensation lors des interfaces PKS/NRPS et NRPS/PKS.

### d) Les modifications enzymatiques après la chaîne d'assemblage

Les composés NRPK libérés du dernier module enzymatique de la chaîne d'assemblage subissent souvent des modifications supplémentaires jouant un rôle important dans l'établissement de la bioactivité du composé. Ces modifications sont catalysées par des enzymes de finition codées dans le *cluster* de gènes de biosynthèse du composé. Ces enzymes catalysent typiquement des réactions d'oxydation, de O-/C-/N-glycosylation, de N-/O-acylation, d'alkylation (méthylation généralement) et d'halogénéation [265, 296-298].

### e) Les enzymes de réparation des chaînes d'assemblage

Les chaînes de production des composés NRPK peuvent faire des erreurs, notamment incorporer des monomères inadaptés sur les domaines T, bloquant la chaîne d'assemblage. Les *clusters* de gènes de biosynthèse des NRPK codent parfois des enzymes de réparation discrètes appelées thioestérases TE de type II qui élimineraient les monomères inadaptés par hydrolyse, régénérant ainsi les domaines T bloqués [299, 300].

### f) Les domaines manquants et autres violations de colinéarité

Dans le petit échantillon des lignes d'assemblage étudiées à ce jour, des variantes ont été détectées dans lesquelles des domaines normalement présents en *cis* sont manquants [301]. Par exemple, le complexe multienzymatique produisant la yersiniabactine contient trois modules NRPS mais un seul domaine A. Il a été démontré que ce domaine active la L-cystéine et la transmet en *cis* à un domaine T et en *trans* pour les deux autres domaines T [302]. De nombreux gènes de PKS de type I ont été séquencés auxquels il manque un, plusieurs ou l'ensemble des domaines AT [303-306]. Des domaines AT agissent alors en *trans* pour fournir des unités malonyle ou méthylmalonyle aux domaines T dans les modules sans modules AT.



Dans certaines voies de biosynthèse, des modules semblent être ignorés, probablement en raison d'une non-fonctionnalité enzymatique ou architecturale de ces domaines : ce phénomène est appelé le « domain skipping » [307, 308]. Par ailleurs, certaines chaînes de PK en croissance peuvent faire une pause et être allongées à deux reprises sur le même module, à la manière d'un bégaiement, ce processus est dénommé le « domain stuttering » [309-311].

#### g) Les fragments de chaîne d'assemblage

Des études bioinformatiques ont révélé des gènes codant pour des fragments de chaînes d'assemblage : des didomaines A-T ou des domaines A et T séparés (A/T) mais adjacents se retrouvent en l'absence de tout autre composant de type NRPS. Ces domaines A activent un L-acide aminé spécifique et l'installent sur le domaine T. Ce groupe aminoacyl lié par un pont thioester au domaine T est alors le substrat pour une enzyme partenaire, qui induit une transformation chimique au carbone  $\beta$  ou  $\gamma$  de l'aminoacyl-S-T. L'acide aminé modifié peut alors être libéré par une thioestérase. Cette stratégie permet de piéger une fraction du pool d'acides aminés protéinogènes et de les modifier en une forme non protéinogène utilisable ensuite par différentes voies métaboliques secondaires [312]. Cette logique est aussi connue dans des systèmes NRPS [313].

### **3. La taille des chaînes d'assemblage**

Les chaînes d'assemblage des PKS de type I et des NRPS se composent parfois de composants protéiques pouvant atteindre plusieurs mégadaltons. Ces chaînes d'assemblage sont comparables à un chapelet dans lequel chaque perle est un domaine replié de manière autonome [25]. La chaîne d'assemblage du 6-dEB comporte sept modules répartis sur trois protéines d'environ 200 kDa chacune (**figure 27**) [275]. La synthétase de la cyclosporine possède 41 domaines répartis sur 11 modules, regroupés dans une seule protéine de 15 000 résidus correspondant à une masse de 1,5 MDa [314], alors que le peptide cyclique qu'elle synthétise n'est long que de 11 acides aminés.

De si grandes protéines sont vraisemblablement sensibles aux problèmes de repliement, vulnérables à la protéolyse, et enclin à l'inactivation par mutation [25]. Le modèle le plus commun d'organisation des modules sur plusieurs sous-unités protéiques peut aider à éviter ces problèmes et pourrait également faciliter l'évolution des composants de ces chaînes d'assemblage par des additions, soustractions ou substitutions de domaines. Cependant, le fractionnement d'une chaîne d'assemblage en de multiples protéines crée d'autres problèmes comme la reconnaissance inter-protéines lors du transfert des oligomères en croissance entre les modules enzymatiques.

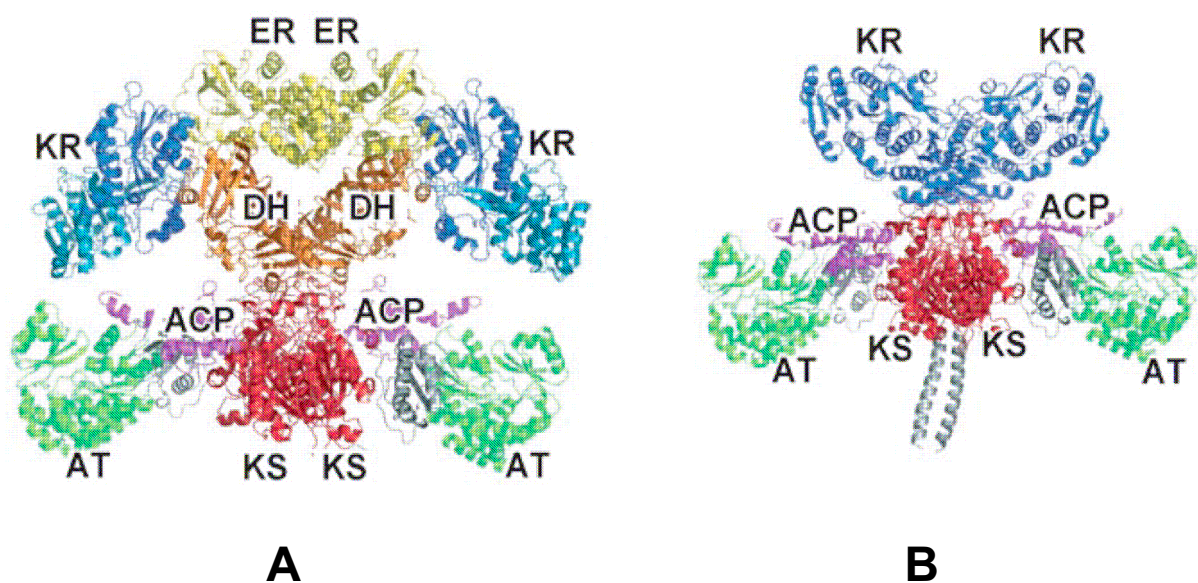
#### **4. Structure des chaînes d'assemblage PKS et NRPS**

La biosynthèse catalysée par les chaînes d'assemblage dépend de l'action successive de plusieurs dizaines d'enzymes différentes. De nombreuses interactions protéine-protéine doivent opérer de manière précise et coordonnée afin d'optimiser le processus d'assemblage. Les détails moléculaires de ces interactions commencent à être déchiffrés.

##### a) L'architecture modulaire

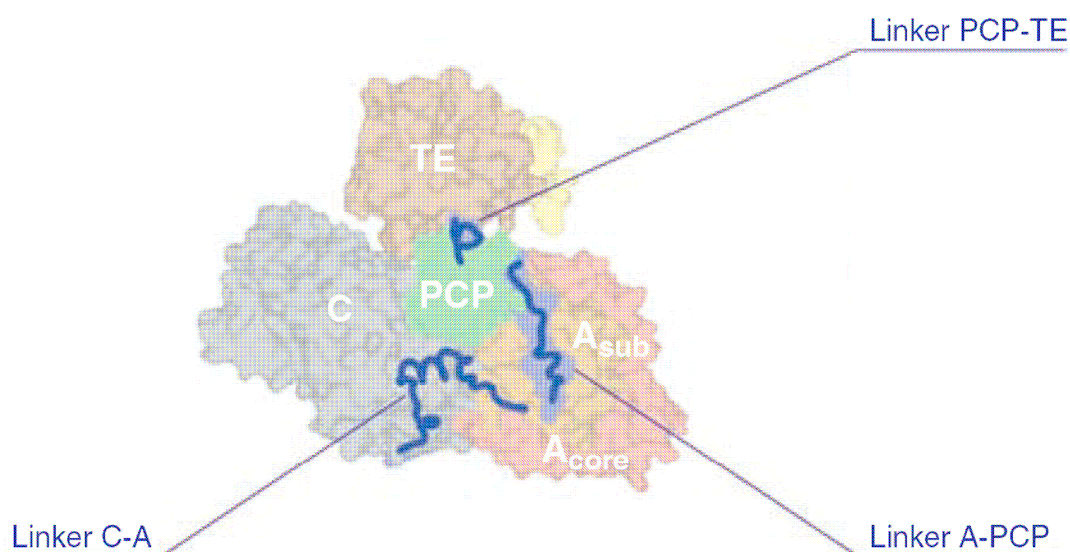
La logique d'assemblage des NRPK est basée sur la liaison du substrat à un petit domaine protéique T porteur. Cette microcompartimentalisation présente l'avantage de séquestrer les substrats et les intermédiaires réactionnels loin de la concurrence d'autres processus cellulaires, comme les autres lignes d'assemblage de NRPK [315]. Le bras phosphopantéthéine des domaines T ou les domaines T eux-mêmes sont des éléments mobiles oscillants, qui jouent un rôle essentiel dans l'optimisation des interactions avec les domaines catalytiques des modules de PKS et de NRPS [316, 317]. Cependant, il n'existe pas d'homologies significatives entre les domaines T des PKS et des NRPS et certains aspects de leur mobilité reposeraient sur des mécanismes différents. Un ancrage temporaire du domaine T des PKS sur les domaines catalytiques partenaires serait facilité par des interactions électrostatiques, grâce à des motifs de reconnaissance du domaine T chargés négativement [318]. Chez les NRPS, un changement de conformation du domaine T induit par le substrat serait utilisé pour favoriser les interactions avec les domaines partenaires [316].

Certains domaines catalytiques des PKS ont été cristallisés comme le didomaine KS-AT par exemple. La structure d'un module entier de PKS de type I peut être déduite de celles de synthèses d'acides gras (FAS), des systèmes multienzymatiques itératifs étroitement apparentés aux PKS [319, 320]. Cette analyse a révélé une grande interface d'homodimérisation de la FAS (**figure 31**) [321]. L'homodimérisation des PKS est aussi supportée par des études de cross-linking et de mutation-complémentation. Par ailleurs, au sein d'un module, des séquences de liaison appelées « linkers » inter-domaines régissent la position relative des différents domaines [317].



**Figure 31 - Modèles structuraux des modules de PKS de type I.** D'après Khosla *et al*[321].  
 A. Modèle structural du module 4 de la DEBS, qui incorpore une boucle réductrice complète (DH-ER-KR).  
 B. Modèle structural des modules 3 et 5 de la DEBS, qui incluent un seul domaine facultatif KR.

La structure cristalline d'un module de terminaison NRPS (C-A-T-TE) a été récemment rapportée [322]. Les domaines C et A s'associent étroitement pour former une plate-forme catalytique, avec leurs sites actifs sur le même côté de la plate-forme. Le domaine porteur T semble mobile et pourrait présenter son bras phosphopantéthéine porteur du substrat aux domaines A et C. Contrairement aux PKS, les NRPS sembleraient fonctionner sous forme monomérique ou dimérique [323, 324]. Par ailleurs, les « linkers » inter-domaines des NRPS (**figure 32**) sont courts (<30 résidus) et leur flexibilité est régulée par des ponts peptidiques Ala-Pro apportant une certaine rigidité. Ces caractéristiques individualisent les domaines, tout en permettant les interactions avec les domaines partenaires [322].



**Figure 32 - Structure cristalline du module de la dernière NRPS de la surfactine.**

D'après Tanovic *et al* [322].

Domaines de condensation C, d'adénylation A, de thiolation PCP et thioestérase TE. Linker : séquences de liaison inter-domaines. A<sub>core</sub> : sous-domaine principal N-terminal de A, A<sub>sub</sub> : sous-domaine C-terminal de A.

#### b) Reconnaissance entre les sous-unités protéiques

Les modules PKS et NRPS peuvent être répartis sur plusieurs protéines. Cette organisation impose des transferts inter-protéiques entre les domaines T et KS des PKS, et les domaines T et C des NRPS. Chaque protéine doit donc reconnaître spécifiquement la protéine partenaire appropriée.

##### (1) Reconnaissance entre les sous-unités PKS

Dans les PKS, la région N-terminale réalise des interactions spécifiques stables avec la région C-terminale de la sous-unité PKS partenaire. Ces régions, appelées domaines d'ancrage, sont portables, c'est-à-dire qu'elles fonctionnent indépendamment de leur sous-unité d'origine [317]. La stabilité des interactions entre ces domaines d'ancrage repose sur la formation d'un complexe de 4 hélices  $\alpha$  entrelacées, présentant des contacts hydrophobes et des interactions électrostatiques critiques [325]. Ces domaines d'ancrage présentent deux éléments de dimérisation (une interaction de 4 hélices  $\alpha$  entrelacées et de 2 hélices  $\alpha$  enroulées à l'extrémité des domaines T et KS respectivement) qui interviennent aussi dans l'association homodimérique des PKS [325]. Par ailleurs, des résidus de surface des domaines T et KS sont aussi impliqués dans la discrimination des partenaires protéiques [326].

### (2) Reconnaissance entre les sous-unités NRPS

Dans les NRPS, les sites d'interaction entre protéines sont nommés domaines favorisant la communication ou domaines COM et sont portables. Ils présentent peu d'homologie de séquence avec les domaines d'ancrage des PKS et ne semblent pas être impliqués dans la dimérisation des modules NRPS. Les domaines COM C-terminaux incorporent une forte proportion de résidus acides alors que les domaines COM N-terminaux contiennent surtout des résidus polaires, suggérant que la spécificité des domaines COM est basée sur des interactions électrostatiques et/ou polaires entre des résidus clés [317, 327].

### (3) Reconnaissance entre les sous-unités PKS et NRPS

Les analyses de séquences prédisent que les éléments d'ancrage putatifs des systèmes mixtes ne présentent pas d'homologie notable avec ceux observés chez les systèmes PKS ou NRPS purs (**figure 33**) [328].



**Figure 33 - Domaines d'ancrage entre les sous-unités protéiques des systèmes de biosynthèse mixtes PKS-NRPS . D'après Weissman et Muller [317].**

Les analyses de séquences prédisent que l'ancrage s'opère par un mécanisme commun aux jonctions des protéines PKS-NRPs et NRPS-NRPS dans les systèmes mixtes, alors qu'une architecture de reconnaissance alternative s'effectue aux interfaces PKS-PKS. Les éléments d'ancrage aux intersections NRPS-PKS ne montrent pas d'homologie apparente avec les autres domaines d'ancrage et semblent très variables.

### c) Autres interactions

Les enzymes PPTases et TE de type II semblent reconnaître les domaines porteurs T au niveau de boucles et d'une hélice adjacentes au site actif [329, 330]. Dans les systèmes PKS sans domaine AT covalent, il a été proposé que les séquences en aval des domaines KS soient susceptibles d'être des reliques de domaines AT. Celles-ci serviraient de domaines d'ancrage aux enzymes AT agissant en *trans* afin de rétablir leur position normale dans les complexes enzymatiques [293, 303].

## **C. Fonctions naturelles des PK et des NRP**

Les organismes produisant des composés NRPK investissent une énergie parfois énorme dans une machinerie enzymatique qui fabrique un produit unique. L'importance des fonctions exercées par ces NRPK doit justifier le coût énergétique représenté par ces chaînes d'assemblage [25]. Si la diversité structurale de ces composés suggère des mécanismes d'action différents, il semble que la finalité de ces composés est essentiellement de faciliter la colonisation de certaines niches écologiques.

### **1. La colonisation**

#### a) Armes de destruction massive ou molécules de signalisation ?

La grande majorité des antibiotiques utilisés pour traiter les infections sont des composés de type NRPK d'origine environnementale. Leur rôle naturel a été généralement déduit de leur activité thérapeutique. Dès 1940, Waksman et Woodruff ont proposé que ces composés inhibent la multiplication ou tuent des organismes concurrents dans l'habitat naturel de leur organisme producteur [331]. Toutefois, les molécules sélectionnées par l'industrie pharmaceutique pour leurs propriétés antibiotiques à des concentrations thérapeutiques auraient cependant des fonctions distinctes aux concentrations probablement inférieures rencontrées dans la nature [332].

Les antibiotiques PK tels que l'érythromycine, la tétracycline et la rifampicine induisent des modifications transcriptionnelles spécifiques chez les bactéries de façon dose-dépendante et indépendante des voies de réponse bactérienne au stress [332, 333]. La tétracycline induit par exemple l'expression du SST3 et la cytotoxicité de *Pseudomonas aeruginosa* [334]. Par ailleurs, les antibiotiques déclenchent une réponse spécifique, qui peut moduler les interactions au sein des communautés microbiennes [334, 335]. Ils peuvent donc avoir un rôle de molécules de signalisation en plus de celui d'inhibiteurs. Cette hypothèse est soutenue par la découverte d'une activité antibiotique pour des molécules bactériennes de signalisation intra- et interspécifiques comme les homosérine lactones ou les antibiotiques qui sont impliqués dans le « quorum-sensing » [336-338]. Par exemple, une homosérine lactone de *P. aeruginosa* inhibe les bactéries à Gram positif [337]. Le double rôle de cette molécule du « quorum-sensing » pourrait rendre compte de la colonisation successive de *Staphylococcus aureus* puis de *P. aeruginosa* chez les patients atteints de mucoviscidose [335]. Les composés NRPK pourraient ainsi être des molécules de signalisation ayant un effet synergique ou antagoniste sur les organismes présents dans la niche écologique de la bactérie productrice.

### b) La protection contre les prédateurs eucaryotes

La prédation représente un facteur de mortalité majeur des bactéries dans leur habitat naturel. Pour se protéger les bactéries ont développé des stratégies de défense impliquant les NRPK. La production de la séravettine et de la surfactine par *Serratia marcescens* et certains *Bacillus* spp. respectivement, induit un comportement d'évitement chez *Caenorhabditis elegans* en agissant sur ses neurones chimiosensoriels [339]. La production du massétolide et de la viscosine chez *Pseudomonas fluorescens* est induite par *Naegleria americana* et lysent cette amibe prédatrice [340]. La production des toxines peut aussi être induite après ingestion de la bactérie par le prédateur, comme pour le composé PK violacéine de *Chromobacterium violaceum* qui cible des protozoaires non flagellés prédateurs [341].

### c) La mobilité, l'adhérence et la formation de biofilm

Les lipopeptides (composés NRP portant un acide gras) biosurfactants peuvent changer la viscosité des surfaces et influencer la mobilité et la différenciation multicellulaire bactériennes [342, 343].

Sur les plantes, les lipopeptides peuvent agir comme des agents de mouillabilité de la cuticule hydrophobe des feuilles, en réduisant la tension de surface, ce qui favorise la solubilisation et la diffusion des substrats nécessaires à la croissance microbienne, et la mobilité vers des niches plus riches en nutriments [344-346]. Il est proposé que les lipopeptides soient impliqués dans la coordination de leur mouvement au cours du « swarming »<sup>9</sup> [347].

Les lipopeptides jouent un rôle important dans l'attachement aux surfaces et la formation de biofilms chez *Bacillus* spp. et *Pseudomonas* spp. [342]. Selon leur structure, les lipopeptides peuvent être orientés avec la partie hydrophile face à la surface cellulaire bactérienne, et exposer leur partie hydrophobe vers l'extérieur, et donc faciliter l'attachement aux surfaces hydrophobes ; quand l'orientation est inversée, ces composés favorisent l'adhésion aux surfaces hydrophiles [348]. Les biofilms sont des regroupements structurés de cellules bactériennes adhérant entre elles et incluses dans une matrice extracellulaire de polymères qui adhère aux surfaces biologiques et non biologiques [349]. Certains lipopeptides fonctionnent comme des molécules morphorégulatrices dans ce comportement multicellulaire bactérien. La surfactine de *Bacillus subtilis* agit comme une molécule signal dans le déclenchement de la formation de la matrice des biofilms [350, 351].

---

<sup>9</sup> Déplacement bactérien en groupe sur une surface solide humide, précédant la formation de biofilm.

Les lipopeptides peuvent aussi influencer défavorablement l'attachement aux surfaces et la formation de biofilm par d'autres micro-organismes. La surfactine réduit fortement l'adhérence de pathogènes d'origine alimentaire tels que *Listeria monocytogenes* et *Enterobacter sakazakii* [352], inhibe la formation de biofilm chez *Salmonella* Typhimurium [353], altère la formation de structures aériennes chez *Streptomyces coelicolor* et affecte ainsi sa capacité à coloniser une niche écologique [354].

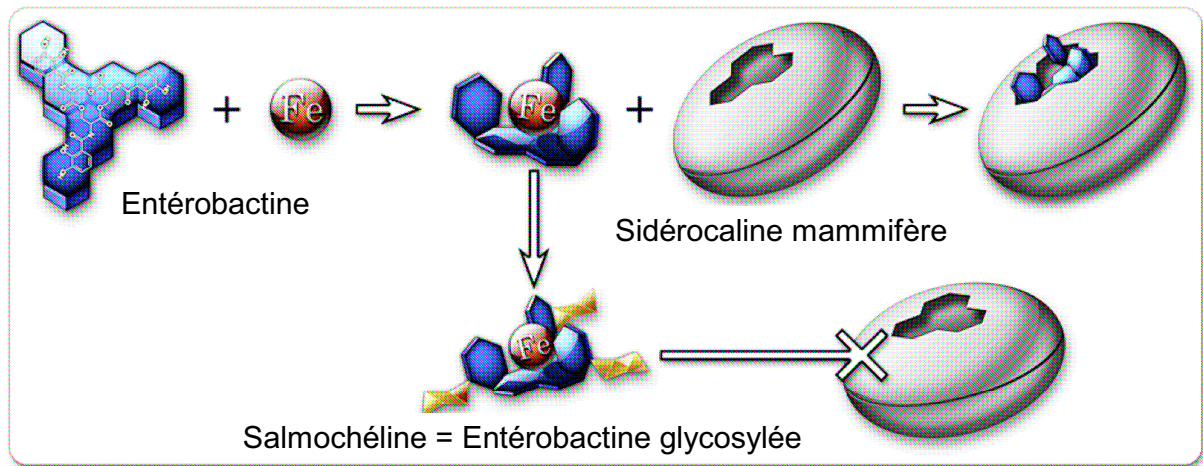
#### d) La chélation du fer et des autres ions métalliques

Le fer est indispensable au développement des organismes. Les sidérophores, molécules permettant la capture du fer, sont fréquemment des composés de type NRPK sécrétés par les bactéries, les champignons et les plantes [355]. Ils sont produits et sécrétés dans les environnements pauvres en fer libre. Chez les vertébrés, le fer est en grande partie lié à des macromolécules. Lorsque les bactéries sont au contact de ces derniers, les sidérophores sécrétés par les bactéries entrent en compétition avec les macromolécules eucaryotes pour la chélation du fer et sont ensuite absorbés à l'aide de récepteurs spécifiques par les bactéries [356]. De nombreux organismes sécrètent plus d'un sidérophore. Dans différentes conditions environnementales et en fonction de l'état ionisé du fer, la sécrétion de l'un ou l'autre des sidérophores semble prédominer. Certains sidérophores sont utilisés comme molécules de stockage pour le fer, restituant lentement le fer une fois entrés dans la cellule.

Toutes les souches de *E. coli* produisent un sidérophore primaire de type NRP, l'entérobactine (**figures 34 et 35**). Les mammifères ont développé une parade pour bloquer l'acquisition du fer par ce sidérophore, dans le cadre de leur réponse immunitaire innée. La lipocaline 2, ou sidéocaline, sécrétée par les macrophages et les hépatocytes lors d'une infection séquestre l'entérobactine, empêchant *E. coli* d'acquérir le fer (**figure 35**) [357]. Certaines souches pathogènes de *E. coli* sécrètent des sidérophores secondaires, considérés comme des facteurs de virulence comme la yersiniabactine et la salmochéline (**figures 34 et 35**). Cette dernière est une entérobactine glycosylée échappant à la séquestration par la lipocaline [21, 358].

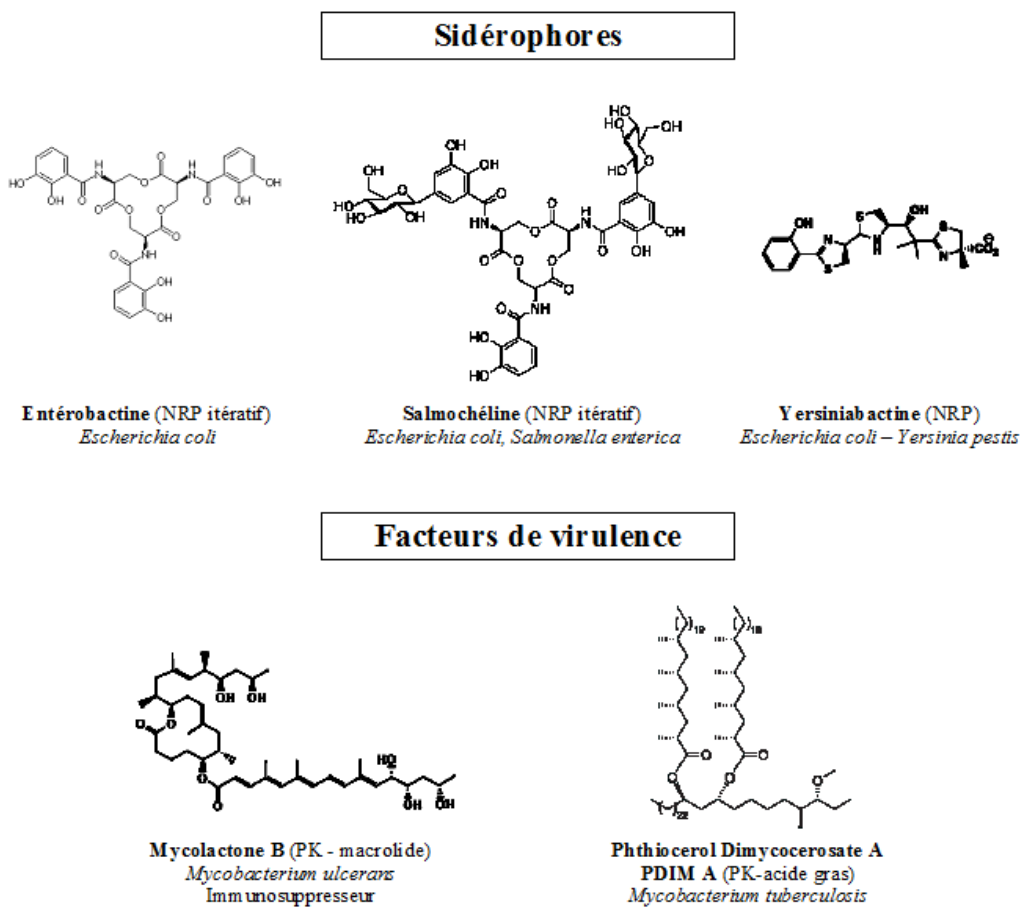
De nombreux lipopeptides bactériens ont un potentiel de chélation des ions métalliques ( $\text{Ca}^{++}$ ,  $\text{Na}^+$ ,  $\text{K}^+$  et les métaux lourds). Certains protégeraient les bactéries des effets toxiques d'ions métalliques en les chélatant. D'autres favorisent leur absorption et leur utilisation par la bactérie à l'image des sidérophores et du fer [343].





**Figure 34 - Acquisition du fer par *E. coli* chez les mammifères.** D'après Wiles *et al* [21].

L'hôte et le pathogène rivalisent pour l'acquisition du fer et ont développé des stratégies multiples pour surpasser l'autre. Le sidérophore entérobactine des souches de *E. coli* séquestre le fer avec une forte affinité, alors que la sidérocaltine mammifère lie l'entérobactine et empêche son absorption par les bactéries. Des ExPEC portant le cluster de gènes *iro* peut modifier l'entérobactine par glycosylation, créant la salmochéline qui lie efficacement le fer mais n'est pas reconnue par la sidérocaltine.



**Figure 35 - Exemples de composés NRPK impliqués dans la colonisation et la survie bactériennes.**

## 2. La symbiose

De nombreuses souches de *Bacillus* spp. et de *Pseudomonas* spp. protègent les plantes contre des infections bactériennes par la production d'antibiotiques. La souche de *B. subtilis* 6051 colonise les racines d'*Arabidopsis thaliana* et forme un biofilm stable et étendu. Ce biofilm associé à la sécrétion de la surfactine protège la plante contre le pathogène *Pseudomonas syringae* pv *tomato* [359]. La production du lipopeptide viscosinamide par *Pseudomonas fluorescens* inhibe l'infection des racines de la betterave à sucre par l'oomycète<sup>10</sup> *Pythium ultimum*, en inhibant la croissance de son mycélium et la motilité de ses zoospores [360].

Certains lipopeptides ont aussi un effet bénéfique sur les plantes en stimulant leur système immunitaire. De nombreux lipopeptides produits par des souches de *Pseudomonas* spp. et *Bacillus* spp. non pathogènes déclenchent des réponses de défense chez les plantes contre les champignons et les oomycètes [344, 361].

L'une des découvertes symbiotiques les plus surprenantes a été faite sur le champignon *Rhizopus microsporus*, l'agent causal de la fonte des semis du riz. Un facteur de virulence du champignon, la rhizoxine est un puissant PK cytotoxique qui inhibe l'assemblage des microtubules. Les gènes produisant la toxine appartiennent à une bactérie symbiotique intracellulaire du genre *Bulkhoderia*, localisée dans le cytosol fongique et indispensable à la formation des sporanges et spores fongiques [362, 363]. En inhibant la mitose, la rhizoxine détruit et tue les jeunes plants de riz, dont les tissus en décomposition profitent alors à l'hôte fongique. Il s'agit d'un mutualisme intime ayant une fonction métabolique spécifique qui influe sur le mode de vie du champignon. Deux organismes en symbiose peuvent ainsi être bénéficiaires de la destruction d'un troisième.

## 3. La virulence

Deux excellents exemples d'implication des NRPK dans la virulence bactérienne ont été mis en évidence dans le genre *Mycobacterium*.

*M. ulcerans* sécrète la mycolactone (**figure 35**), un PK de type macrolide, qui est un facteur majeur de virulence. Cette bactérie est l'agent causal de l'ulcère de Buruli, une maladie de peau invalidante sévissant en Australie et en Afrique de l'Ouest. Elle se caractérise par la persistance de lésions nécrotiques de la peau et du tissu graisseux sous-jacent, l'absence de réaction inflammatoire aiguë et de douleur. La mycolactone reproduit la plupart des signes de

---

<sup>10</sup> Protistes filamenteux aquatiques, exclus phylogénétiquement du règne des champignons.

l'infection lorsqu'elle est injectée par voie intradermique à des cobayes. La mycolactone est une cyclomoduline, qui induit un arrêt du cycle cellulaire en phase G<sub>0</sub>/G<sub>1</sub>, l'inhibition de la synthèse des protéines et éventuellement la mort cellulaire [364]. Elle présente des propriétés immunosuppressives en inhibant la voie NF-κB des monocytes et des lymphocytes T, la production de cytokines, et la maturation et la migration des cellules dendritiques [365, 366]. Ces résultats suggèrent que la production de ce PK par *M. ulcerans* peut limiter à la fois l'initiation des réponses immunitaires et le recrutement de cellules inflammatoires au site de l'infection. En outre, l'anergie profonde et systémique des lymphocytes T aux antigènes mycobactériens persiste de nombreuses années après un traitement curatif de l'ulcère de Buruli [367]. Des isolats de *M. ulcerans* producteurs de mycolactone ont été retrouvés chez plusieurs insectes aquatiques (punaises et araignées d'eau, moustiques) qui semblent être le réservoir de ce pathogène [368, 369]. La mycolactone est essentielle à la colonisation initiale et à long terme de la punaise d'eau par *M. ulcerans* [370] et est produite par certaines mycobactéries aquatiques proches de *M. ulcerans* responsables de maladies chez les poissons et les grenouilles [371].

Tous les membres du complexe *M. tuberculosis* et quelques autres espèces de mycobactéries dont la plupart sont pathogènes pour l'homme et les animaux produisent un groupe de lipides, les phthiocérol dimycocérosates (PDIM) (**figure 35**) [372]. Ces composés sont produits par l'action combinée de PKS et de synthases d'acide gras (FAS), et se localisent à la surface des bacilles où ils constituent une partie de la pseudo-membrane externe de la bactérie [373]. Des modèles murins d'infection ont révélé la nécessité des PDIM pour la multiplication de *M. tuberculosis* au cours de la phase aiguë de l'infection [374, 375]. Les PDIM participent à l'invasion des macrophages en favorisant la phagocytose et en prévenant l'acidification phagosomale, par l'induction de changements dans l'organisation des lipidiques de la membrane cytoplasmique des cellules hôtes [376]. Ils contribuent aussi à la croissance initiale des bacilles tuberculeux en les protégeant du monoxyde d'azote. Enfin, les PDIM jouent un rôle dans la modulation des réponses immunitaires précoces des macrophages murins en altérant la production de cytokines pro-inflammatoires [377].

Les lipopeptides sont aussi impliqués dans le pouvoir pathogène de *Pseudomonas* chez la plante. La production de syringomycine et de syringopeptine par *P. syringae* pv *syringae* augmente substantiellement la sévérité de la maladie [378]. Le viscosine de *P. fluorescens*, agent de la pourriture du brocoli, joue un rôle indirect dans la virulence en facilitant la colonisation du tissu de la plante hôte [346].

## Chapitre II

# Résultats expérimentaux

”Be naïve and try the experiment even if its chances of success *are very small.*”

Feodor F. K. LYNEN (1911-1979)  
Prix Nobel de physiologie/médecine en 1964



# I. Article 1 – Structure-fonction de la peptidase ClbP

**ClbP is the prototype of a periplasmic peptidase subgroup involved in the biosynthesis of nonribosomal peptides.**

Damien Dubois, Olivier Baron, Antony Cougnoux, Julien Delmas, Nathalie Pradel, Michèle Boury, Bernadette Bouchon, Marie-Agnès Bringer, Jean-Philippe Nougayrède, Eric Oswald et Richard Bonnet

Soumis à “*Journal of Biological Chemistry*”

Le gène *clbP* est l'un des 16 gènes de l'îlot *pks* nécessaires à son activité [24]. L'analyse *in silico* de la protéine correspondante suggère qu'il code une protéine périplasmique fixée à la membrane interne et constituée de trois parties : (i) une séquence signal clivable N-terminale dépendant de la voie de translocation Sec (ALA-QE 30-31), (ii) un domaine principal périplasmique (positions 31 à 375), pouvant appartenir à la famille des enzymes S12 de la classification MEROPS, et (iii) trois hélices transmembranaires en C-terminal (résidus 390-412, 433-455, 465-485). Ce type de protéine n'avait pas été identifié jusqu'alors comme intervenant dans la biosynthèse de PK ou de NRP.

Nous avons déterminé la structure cristallographique du domaine médian de ClbP par diffraction aux rayons X avec une résolution de 2,4 Å. La structure de ce domaine, appelé ClbPpep est similaire à celle des enzymes S12, notamment les  $\beta$ -lactamases de classe C et surtout la peptidase Pab87 [379, 380]. La comparaison de la structure de ClbP avec ces dernières a permis d'identifier la poche active. Elle forme une cavité ouverte sur le solvant, profonde et exceptionnellement large, grâce au positionnement atypique d'une boucle (résidus 295 à 315) et de deux brins (b9 et b10, résidus 310 à 311 et 316 à 318 respectivement) à une extrémité de la poche catalytique et à l'absence de deux hélices (h9 et h10) à l'autre extrémité de la poche. Cette caractéristique de ClbP suggère une adaptation de la poche active de ClbP à un substrat atypique.

Au fond de cette poche, nous avons observé des résidus correspondant aux motifs critiques du site actif des enzymes de la famille S12, à savoir les motifs <sup>95</sup>SxxK et <sup>186</sup>Yxx, qui sont connus pour catalyser le processus enzymatique. Le rôle de ces résidus a été exploré par mutagenèse dirigée et analyse des effets cytopathogènes engendrés par une souche de *E. coli* K12 contenant l'îlot *pks* défectif pour *clbP* et transcomplémentée par les mutants de ClbP.

Les résidus  $^{95}\text{S}$ ,  $^{98}\text{K}$  et  $^{186}\text{Y}$  des motifs décrits ci-dessus se sont révélés essentiels, montrant ainsi que ClbP agit selon un processus catalytique basé sur le mécanisme généralement accepté pour des enzymes à sérine active S12. En revanche, les mutations des autres résidus pointant vers le site actif n'affectaient pas sensiblement l'activité cytopathogène de l'îlot *pks*.

Enfin, l'analyse de la nature et du positionnement des résidus du site actif de ClbP suggéraient une activité peptidasique. Nous avons confirmé l'activité peptidase de ClbPep à l'aide de substrats chromogéniques des aminopeptidases, de même que sa capacité à lier un composé de type NRP tel que l'imipénème. Par contre, ClbP ne présentait pas d'activité  $\beta$ -lactamase notable et était incapable de conférer une résistance aux  $\beta$ -lactamines antibiotiques, contrairement aux  $\beta$ -lactamases.

Par ailleurs, nous avons montré par fractionnement cellulaire que ClbP était associée à la membrane cytoplasmique bactérienne. En utilisant une protéine de fusion peptide signal de ClbP-GFP et des *E. coli* mutants isogéniques du système Sec, nous avons aussi confirmé que la séquence N-terminale était une séquence d'adressage au périplasma dépendante de la voie de translocation SRP/Sec/YidC. Ces données suggèrent une localisation périplasmique du domaine catalytique ClbPep, qui reste solidaire de la membrane *via* les 3 hélices transmembranaires putatives C-terminales.

L'importance de ClbP pour l'activité de l'îlot *pks* suggérait que ClbP ait un rôle critique, potentiellement conservé dans d'autres îlots génomiques de même nature. Nous avons donc recherché des orthologues de ClbPep dans les banques de données. Les orthologues ont été analysés sur le plan phylogénétique, topologique et génomique. L'arbre phylogénétique construit à partir des séquences peptidiques a révélé trois branches majeures. La première portait ClbP et des homologues appelés ClbP-like qui présentaient une topologie similaire. Ces protéines étaient dans 31% (n=19) des cas codées par des clusters de gènes codant des NRPS ou PKS-NRPS. Elles étaient principalement (59/61) issues de bactéries appartenant au phylum bactérien des Firmicutes (*Bacillus* et *Clostridium*). La deuxième branche phylogénétique incluait les protéines « FmtA-like » du genre *Staphylococcus*. Ces protéines de fonction inconnue présentent une topologie similaire à celle de ClbP, mais leurs gènes n'étaient jamais observés à proximité de gènes codant des mégasynthèses. La troisième branche phylogénétique s'écartait distinctement des deux autres et regroupait des membres de la famille S12 inclus dans l'analyse pour représenter les enzymes de cette famille dont les activités biochimiques et/ou biologiques sont connues. Ces données suggéraient que ClbP et les protéines ClbP-like d'une part, et les protéines FmtA-like d'autre part constituent des sous-groupes distincts de la famille S12.

Afin de tester la proximité des protéines ClbP, ClbP-like et FmtA-like, nous avons analysé la capacité de protéines ClbP-like représentatives (provenant de différents phyla, de différents genres bactériens et associées génétiquement ou non à des îlots codant des mégasynthèses) et d'une FmtA-like à xéno-complémenter l'activité cytopathique d'une souche de *E. coli pks*-positive déficiente pour ClbP. Une xéno-complémentation a été observée pour toutes les protéines ClbP-like testées mais pas pour FmtA-like. Ces données soutiennent la promiscuité fonctionnelle des protéines ClbP et ClbP-like. Elles révèlent aussi leur tolérance au substrat car elles sont interchangeable alors qu'elles peuvent être issues d'îlots génomiques codant des NRPS ou PKS-NRPS sans homologie significative.

Les protéines ClbP-like présentent peu de résidus à la fois conservés et spécifiques. La plupart des résidus conservés chez ClbP et les protéines ClbP-like sont observés dans les protéines FmtA-like et ne sont donc pas liés à leur spécificité fonctionnelle. Nous avons donc investigué les propriétés physico-chimiques de surface de ClbP et des autres enzymes S12 de structure connue afin d'identifier une caractéristique pouvant expliquer la spécificité des ClbP-like. La poche catalytique de ClbP a présenté un potentiel de surface négatif, alors que celui des autres enzymes S12 un potentiel de surface positif. Des résultats comparables ont été observés pour les protéines ClbP-like (surface électro négative) et FmtA-like (surface électro positive) à partir de structures modélisées par homologie. Le potentiel électrostatique de surface de la poche catalytique des ClbP-like est ainsi corrélé à la capacité de restaurer la bioactivité de l'îlot *pks* déficient pour ClbP, et semble donc être un déterminant majeur de leur activité. Ce résultat était soutenu par le calcul du potentiel de surface des ligands connus ou putatifs de ces enzymes, qui était complémentaire de celui des enzymes ; le potentiel électrostatique de surface du composé PK-NRP Zwittermicine A (possédant une ClbP-like dans son îlot de biosynthèse [381] était positif alors que celui des substrats des  $\beta$ -lactamases ou des carboxypeptidases était négatif ( $\beta$ -lactamines et précurseurs du peptidoglycane).

Nous avons proposé finalement de désigner cette nouvelle sous-famille de peptidases à sérine active, probablement impliquée dans la maturation finale de NRP bioactifs chargés positivement, les peptidases associées aux NRP.





**ClbP IS THE PROTOTYPE OF A PERIPLASMIC PEPTIDASE SUBGROUP INVOLVED IN THE BIOSYNTHESIS OF NONRIBOSOMAL PEPTIDES**

**Damien Dubois<sup>1,2</sup>, Olivier Baron<sup>3,4,5\*</sup>, Antony Cougnoux<sup>2</sup>, Julien Delmas<sup>1,2</sup>, Nathalie Pradel<sup>2</sup>, Michèle Boury<sup>3</sup>, Bernadette Bouchon<sup>6</sup>, Marie-Agnès Bringer<sup>2</sup>, Jean-Philippe Nougayrède<sup>3,5</sup>, Eric Oswald<sup>3,4,5</sup> and Richard Bonnet<sup>1,2</sup>**

From CHU Clermont-Ferrand, Laboratoire de bactériologie clinique<sup>1</sup>, Clermont Université, JE 2526, USC-INRA 2018<sup>2</sup>, Inserm U 990<sup>6</sup>, Clermont-Ferrand, France, and INRA, UMR1225<sup>3</sup>, CHU Toulouse, Laboratoire de Bactériologie-Hygiène<sup>4</sup>, Université de Toulouse, ENVT<sup>5</sup>, Toulouse, France

Running head: Structure and functionality of ClbP peptidase

Corresponding authors: Richard Bonnet; [rbonnet@chu-clermontferrand.fr](mailto:rbonnet@chu-clermontferrand.fr); Phone (+33) 4 73 75 49 20; Fax (+33) 4 73 75 49 22 or Eric Oswald; [e.oswald@envt.fr](mailto:e.oswald@envt.fr); Phone (+33) 5 61 19 39 91; Fax (+33) 5 61 19 39 75

<sup>†</sup> in memoriam

**The *pks* genomic island of *Escherichia coli* encodes polyketide (PK) and nonribosomal peptide (NRP) synthases that allow assembly of a putative hybrid PK-NRP compound named colibactin that induces DNA double-strand breaks in eukaryotic cells. The *pks*-encoded machinery harbors an atypical essential protein, ClbP. Here we report that ClbP harbors a catalytic and periplasmic N-terminal part, and a C-terminal part associated to the cytoplasmic membrane. ClbP crystal structure and mutagenesis experiments revealed a serine-active site and original structural features compatible with peptidase activity confirmed by biochemical assays. Ten ClbP homologs were identified *in silico* in NRP genomic islands of close and distant-related bacterial species. All tested ClbP homologs were able to complement a *clbP*-deficient *E. coli* mutant. ClbP is therefore a prototype of a new subfamily of periplasmic peptidases, which are involved for the maturation of NRP compounds. Such peptidases will be powerful tools for the manipulation of NRP biosynthetic pathways.**

Polyketides (PK), nonribosomal peptide (NRP) and hybrid NRP-PK natural products have played a prominent role in and continue to inspire organic chemistry and medicines (1). Their biosynthesis occurs on multimodular enzymatic assembly lines, so-called megasynthases, and are typically released by thioesterases (2). The

conformations that establish biological activity are attained by diverse tailoring and editing enzymes such as acylases, alkylases, glycosylases, and oxidoreductases during chain growth and after thioesterase-mediated release. The dedicated tailoring enzymes are encoded by genes clustered with the assembly line genes for coordinated regulation.

Recently, Nougayrède *et al.* identified in *Escherichia coli* strains and other *Enterobacteriaceae* a large (54 kb) genomic island named *pks*, which encodes a typical cluster of NRP synthases (NRPS), PK synthases (PKS) and hybrid NRPS/PKS (3,4). The *pks* island was detected in *E. coli* strains involved in extraintestinal infections such as urinary tract infection and septicemia, but also in commensal strains from healthy people (3,5,6). The genetic and functional analyses of the *pks* island indicate that it encodes for synthesis of a PK-NRP hybrid compound, named colibactin. *E. coli* expressing this gene cluster causes DNA double-strand breaks in human eukaryotic cells upon infection (3). This DNA damage induces the phosphorylation of the H2AX histone ( $\gamma$ H2AX) and activates the ATM-Chk2 signaling pathway that leads to transient G<sub>2</sub>/M cell cycle-arrest and cell swelling (megalocytosis). Exposed cells exhibit signs of incomplete DNA repair leading to chromosomal instability, increase in gene mutation frequency and anchorage-independent colony formation, demonstrating the mutagenic and transforming potentials of colibactin (7).

In addition to eight NRP and PK megasynthases, the *pks* island encodes nine accessory, tailoring and editing enzymes (3). Eight of these proteins, including the protein ClbP, are required to induce the DNA damage in infected cells. No protein similar to ClbP has been reported so far as an accessory or tailoring enzyme. Here, we determined the crystal structure of ClbP main domain, its biochemical activity, its cellular localization, and we characterized distant ClbP homologous proteins that form a new subgroup of serine-reactive peptidases associated with NRP assembly lines.

### Experimental procedures

**Strains and plasmids.** The bacterial strains and plasmids used in this study are listed in Suppl. Tables 3 and 4.

**DNA cloning.** The DNA sequences of *clbP* were amplified from the genome of *E. coli* strain IHE3034 (3). The DNA fragments encoding ClbP, the signal sequence of ClbP plus the first six amino acids of the mature part (ClbPss), and the peptidase domain of ClbP (designated ClbPpep) were cloned into pASK-IBA33plus, p8761, pET9a resulting in plasmids p33ClbP (ClbP + C-terminal in-frame *RGS6His* tag), p61ssClbP (ClbPss + C-terminal in-frame GFPmut2) and pClbPpep, respectively. In other experiments, ClbP-6His was subcloned from pMB702 into pBRSK vector to obtain pOB901. Plasmids were subsequently electroporated into competent *E. coli* strains: p33ClbP into TG1 and DH10B/pBAC*pksΔclbP* strains for cellular localization of ClbP and HeLa cells infection experiments, respectively; pOB901 into DH10B/pBAC*pksΔclbP* strain for HeLa cellular challenges; p61ssClbP into wild-type and *secY*, *yidC*, *SRP*, *mreB* and *ftsZ* mutant *E. coli* K-12 strains, for localization and topology analysis; and pClbPpep into BL21(DE3) strain for ClbPpep and derivative overexpression.

The DNA fragments encoding ClbP-like and FmtA-like proteins were cloned into pASK-IBA33plus (C-terminal in-frame *RGS6His* tag). Plasmids were subsequently electroporated into competent *E. coli* strain DH10B/pBAC*pksΔclbP*. Primers used for

cloning are listed in Suppl. Table 5. For more details, see Suppl. Data.

**Site-directed mutagenesis** experiments were performed using Quick-Change™ Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA). For more details, see Suppl. Data.

**Translocation pathway and cellular fractionation.** Cell induction assays of translocation pathway were performed as previously described (8). Only conditional *secY* and *yidC* mutants were available; the cold-sensitive *secY-cs* and *yidC* conditional mutants were used in this study. Cellular fractionation experiments were performed as previously described (8). Western-blotting was performed with polyclonal rabbit anti-GFP (Invitrogen) at dilution of 1:10,000 and monoclonal anti-RGS6His according to the manufacturer's instructions (Qiagen, Courtaboeuf, France). For more details, see Suppl. Data.

**ClbPpep and ClbPpepY186E production and purification.** The overexpression of ClbPpep and its mutant Y186E was performed in *E. coli* BL21(DE3) from pClbPpep and pClbPpepY186E respectively. Purification was carried out by ion-exchange chromatography onto a HiTrap™ Q Sepharose™ High Performance column (Amersham Pharmacia Biotech) equilibrated with 20 mM Tris-HCl (pH 7.0) and eluted with a linear NaCl gradient (0 to 500 mM). The ClbPpep-containing elution peak was then purified by gel filtration chromatography onto a Superose™ 12 column (Amersham Pharmacia Biotech) in 100 mM NaCl, 5 mM phosphate buffer (pH 8.0). The ClbPpep-containing elution peak was concentrated by ultrafiltration to 10 mg/ml for crystallization. The enzyme was more than 95% homogeneous as judged by Coomassie blue staining after SDS-PAGE. For more details, see Suppl. Data.

**Crystallization and structure determination.** ClbP crystals were grown in hanging drops over a well solution of 0.8 M monosodium dipotassium phosphate buffer (pH 7.0). Data were collected using a Q315r ADSC-CCD detector on ESRF beamline 14-4 at the European Synchrotron Radiation Facility (Grenoble, France). Reflexions were indexed,

integrated, and scaled using CCP4 package (9). The initial model was obtained by molecular replacement with the program PHASER (10) and the X-ray structures 1E15, 1RGY and 1FR1 as a search model. The structure was automatically and manually refined with REFMAC5 and COOT programs, respectively (11,12). Cross-validation was used throughout and 5 % of the data were used for the  $R_{\text{free}}$  calculation. The stereochemical quality of the models was monitored with the PROCHECK program (13). Ramachandran plots were calculated by RAMPAGE (14). Processing and crystallographic refinement statistics for ClbPpep crystal structure are listed in Suppl. Table 1. The coordinates and structural factors of ClbPpep have been deposited in the RCSB Protein Data Bank, entry 3O3V. For more details, see Suppl. Data.

**Enzyme assays.** D-aminopeptidase activity was detected by monitoring the formation of *p*-nitroaniline from Gly-, D-Ala, L-Ala, and L-Ala-L-Ala-*p*-nitroanilide (Bachem, Weil am Rhein, Germany) at 405 nm in a 100 mM Tris, 75 mM NaCl buffer (pH 7.5) at 25°C.  $\beta$ -lactamase activity was tested by monitoring the hydrolysis of nitrocefin (Calbiochem, Darmstadt, Germany) at 482 nm in a 100 mM Tris, 75 mM NaCl buffer (pH 7.5) at 25°C.

**Sample preparation for mass spectrometry analysis.** ClbPpep (20  $\mu$ M) was incubated at 20°C for 24 h in a 100 mM Tris, 75 mM NaCl buffer (pH 7.5) alone or with 20 mM imipenem (Sigma-Aldrich), in a 1 mL final volume. The reaction mixtures were desalted by dialysis against a 5 mM ammonium bicarbonate buffer before analysis by matrix-assisted laser desorption/ionisation–time of flight-mass spectrometry. For more details, see Suppl. Data.

**Cell culture, bacterial infections,  $\gamma$ H2AX staining and cell cycle analysis.** Cell assays were performed as previously described with minor modifications (3). ~50% confluent HeLa cell cultures were infected at a multiplicity of infection (MOI) of 100. For more details, see Suppl. Data.

**Bioinformatics.** The search for ClbP homologous proteins and analysis of their topology were performed using BLAST (15) and InterProScan server

(<http://www.ebi.ac.uk/Tools/InterProScan/>).

Multiple sequence alignment of peptidase domain and phylogenetic tree were performed using COBALT and TOPALi v2.5 softwares (16,17). Analysis of the genetic context of ClbP- and FmtA-like-encoding genes (more or less 30 kb of both sides of the genes) was performed manually from NCBI genome database. The analysis of PK and NRP megasynthase domains and their substrate specificity was performed using the SEARCH NRPS-PKS program (<http://www.nii.res.in/searchall.html>).

Protonation, energy minimization, charge assignment, computation of the electrostatic surface potentials and visualization were performed with VEGA-ZZ (18) using the semi-empirical method AM1 with the quantum chemistry program MOPAC2009 (19) (<http://openmopac.net>). For Molecular modeling and Motif consensus elaboration, see Suppl. Data.

## RESULTS

***In silico* analysis of ClbP.** The *clbP* open reading frame was predicted to encode a 504 a.a. periplasmic inner membrane-anchored protein comprising three domains: (i) a C-terminal domain of three transmembrane helices (residues 390-412, 433-455, 465-485), (ii) an N-terminal domain supposed to be a cleavable Sec-dependent signal sequence (ALA-QE 30-31), and (iii) a large periplasmic domain (positions 31 to 375). According to primary structure comparisons, the last domain exhibits the two conserved motifs ( $^{95}\text{SMS}^{98}\text{K}$  versus SxxK and  $^{186}\text{YAS}$  versus YxN) that characterize the catalytic center of the MEROPS S12 enzyme family (20), and is therefore designated hereafter ClbPpep. The S12 family is a heterogeneous group of active-site serine enzymes, of which the substrates are peptides or closely related compounds such as  $\beta$ -lactams. The activity and biological function of a few of these enzymes have been described such as class C  $\beta$ -lactamases, R61 D-Ala-D-Ala carboxypeptidase B, DmpB aminopeptidase, Pab87 peptidase, D-amino acid amidase and EstB esterase (21-26). However, the activity and the function of

numerous enzymes of the MEROPS S12 family remain unassigned, notably the FmtA-like proteins (27), which have been previously reported to be related to ClbP (3).

**Structure determination of the ClbP peptidase domain.** In an effort to further understand the function of ClbP, the structure of the ClbPpep peptidase domain was determined by X-ray diffraction to investigate its activity. The structure was refined against diffraction data extending to 2.4 Å resolution. The electron density map (Suppl. Fig. 1) showed well-defined density throughout most of the structure, with the exception of the six N-terminal residues. Diffuse density indicated disorder in the loop harboring residues 241 to 249. The stereochemical parameters of the model were satisfactory (Suppl. Table 1); a Ramachandran plot showed no residues in disallowed regions of  $\phi/\psi$  space. The final model included three ClbPpep monomers and 277 water molecules.

Since the three molecules of ClbP were very similar, we thereafter confine our description of the structure to the monomer B, which yielded a slightly clearer density. The 3D structure of ClbPpep consists of two structural domains, all- $\alpha$  and  $\alpha/\beta$  (Suppl. Fig. 2A). The  $\alpha/\beta$  region is formed by residues 41–112 and 206–375 and the domain folds as a seven stranded anti-parallel  $\beta$ -sheet (b1, b2, b9, b10, b11, b12, b13) with six  $\alpha$ -helices and three  $\beta$ -strands (b3, b4 and b7) packed on both faces of the sheet. The all-helical region (residues 113–205) contains four helices and loops. The conserved motifs <sup>95</sup>SMSK and <sup>186</sup>YAS, which correspond to S12 enzyme active site, are located in a large groove, between the two structural domains near the N-terminus of the first helix of the all- $\alpha$  domain.

The DALI program was used for a quantitative and exhaustive comparison of ClbPpep structure and those deposited in RCSB Protein Data Bank (28). As expected, the most pronounced homology was obtained with the structure of S12 family members (Suppl. Table 2), especially *E. coli* AmpC  $\beta$ -lactamase (PDB ID 2R9W, rmsd 2.1Å, Dali Z-score 37.6) and Pab87 peptidase (PDB ID 2QMI, rmsd 2.6Å, Dali Z-score 37.2). The

following analysis therefore focuses on the structural comparison of ClbPpep, AmpC  $\beta$ -lactamases and S12 MEROPS peptidases.

Although their overall fold was clearly similar, there are significant differences affecting the accessibility of the active site from the surrounding solvent. In comparison to AmpC  $\beta$ -lactamases, the upper part of the catalytic pocket was largely open, because of the absence of two helices (h9 and h10) (Suppl.Fig. 2A and 2E), as observed in Pab87 peptidase and R61 carboxypeptidase (23,29). However, the remaining structural element of ClbP in this area, the loop between residues 295 to 315, b9 and b10 strands (residues 310 to 311 and 316 to 318 respectively), moved away from the catalytic pocket, resulting in a widening of the catalytic groove, which was unusually larger than the known structures of S12 enzymes. In the bottom part of the catalytic pocket, ClbPpep and Pab87 peptidases harbored the additional helix h4' (residues 166 to 170) unlike AmpC enzymes (Suppl.Fig. 2A and 2E). This helix slightly obstructed accessibility to the ClbPpep catalytic pocket. However, the entrance of the binding site was more open in ClbPpep than in S12 peptidases, which harbor one, two or several bulky helices in this area. Of note, between C337 and C367, ClbP harbored a disulfide bond, which links b12 strand (residues 335 to 341) with the C-terminal helix h11. To our best knowledge, this disulfide bond was not observed in other known structures of S12 enzymes. Overall, the ClbPpep structure was closely similar to that of S12 peptidases. However, ClbPpep had noticeable conformational differences in the active site, especially in the two extremities of the catalytic groove, which was unusually large (Suppl. Fig. 2B, 2C and 2D). This atypical shape suggests an adaptation of ClbP catalytic pocket to a specific substrate.

**The catalytic residues and implications for catalysis.** Striking structural similarities were observed between the active site of ClbPpep, class C  $\beta$ -lactamases and other S12 enzymes such as Pab87 peptidase(23,30,31). In particular, the location and geometry of catalytic residues such as S95, the nucleophile, Y186, the general base

and the main chain nitrogen atoms of positions 95 and 330 that form an oxyanion hole were similar (Fig. 1A and 1B). The positioning of the key residue Y186 and the presence of H327 (instead of the K residue of the KTG-box in  $\beta$ -lactamases) make, however, ClbPpep catalytic residues very similar to Pab87 peptidase. These results suggest that these enzymes and other S12 active-site serine enzymes share a catalytic mechanism.

Differences were observed in residues probably involved in substrate recognition. The motif YxN, which is observed in most S12 enzymes was replaced by the <sup>186</sup>YAS motif in ClbP (Fig. 1B). S188 in ClbP was hydrogen bonded to the K98 residue of motif SxxK. In class C  $\beta$ -lactamases, the structural analog is N152. This residue shares similar interaction with the motif SxxK, but can also interact with ligands unlike S188 residue of ClbP (21). H257 of ClbP, which points towards the vicinity of S188, may replace N152 for substrate binding in this area of the active site.

Opposite to the <sup>186</sup>YAS motif, the catalytic pocket of ClbP exhibits the HGG motif. The alignment of ClbP with class C  $\beta$ -lactamases revealed that HGG motif is a structural analog of KTG-box, which is a major wall of the catalytic cavity (Fig. 1B) (31). In class C  $\beta$ -lactamases, K can be replaced by H or R, and T may be replaced by S. A positively charged side chain followed by one bearing a hydroxyl group has therefore been suggested to be universally conserved. Of note, the last residue in ClbP motif is an atypical G328 residue instead of the canonical residue S or T, which are usually directly involved in substrate recognition *via* a hydrogen bond (Matagne et Al. 1998). However, at the vicinity (b11) (Fig. 1A and Suppl. Fig. 2E), positions 330 and 331 of ClbP harbor the unusual residues Q and N respectively, which points towards the catalytic pocket and may establish hydrogen bonds with substrates.

The residues E159 and F316 of ClbPpep are structural analogs of positions involved in the binding of the N-terminal and C-terminal parts of peptides in S12 amino and carboxy peptidases. These positions determine

the carboxypeptidase or aminopeptidase activity of S12 peptidases (32). According to previous works, the residues E159 and F316 of ClbP should favor aminopeptidase activity because the negative-charged residue E159 may accommodate the positive-charged N-terminal part of peptides, and the absence of positive-charged residue in position 316 may alter the accommodation of the negative-charged C-terminal part of peptides.

Overall, the analysis of residues located in ClbPpep catalytic pocket showed that the catalytic process mediated by this enzyme is based on the generally accepted mechanism for active-site serine peptidases. Despite the overall similarity of ClbP and AmpC  $\beta$ -lactamase structures, the structural elements surrounding the active serine more likely form a peptide-recognizing pocket with its own specificities, probably in relation with the biological role of ClbP.

**Active site investigation by site-directed mutagenesis.** The putative active site of ClbP was investigated by site-directed mutagenesis. Resulting mutants were assessed for their ability to restore the cytopathic activity of the *pks*-positive *E. coli*  $\Delta$ *clbP* isogenic mutant. Infected HeLa cells were analyzed for megalocytosis, G<sub>2</sub>/M cell cycle arrest, and phosphorylation of histone H2AX (3).

The Ser and Tyr residues of the two conserved motifs SxxK and YxN are involved in the hydrolytic mechanism mediated by S12 enzymes (33,34). To confirm the importance of these motifs and residues in ClbP, substitutions S95A, K98T and Y186G were introduced by site-directed mutagenesis residues into His-tagged ClbP. None of these mutants were able to restore the cytopathic activity of the *pks*-positive *E. coli*  $\Delta$ *clbP* isogenic mutant unlike the wild-type ClbP protein (Fig. 1C and 1D). These results confirmed the identification of the active site and the importance of ClbP main domain for the functionality of *pks* island.

The six positions 159, 188, 257, 316, 328 and 331 can participate in substrate binding in the ClbP catalytic pocket, because the site chain of these residues pointed

towards the catalytic pocket (Fig. 1A). The hydrogen bond donors S188 and H257 of ClbP-His were replaced by the residue A, which is not able to accept or give hydrogen bonds. The substitutions E159K and F316R were introduced into ClbP-His to modify the charges in the critical areas of the catalytic pocket. We also introduced the substitution G328S to check whether a small residue is required in the second position of the motif HGG to accommodate ClbP substrate. Substitution C337A was also introduced into ClbP-His to check the importance of the disulfide bond C337-C367 for stability or folding. No substitution altered the trans-complementing activity. ClbP activity may have been only weakly modified by the substitutions and the cellular assays are moderately sensitive. Unfortunately, efficient biochemical tests are not yet available to monitor ClbP activity. However, cellular tests are a good representation of the physiological context and the results are evidence of substrate tolerance of ClbP.

**ClbP is a peptidase and a NRP-recognizing protein.** As ClbP presents a S12 peptidase scaffold, peptidase activity was investigated using the reporters Gly-, D-Ala-, L-Ala-, and L-Ala-L-Ala-*p*-nitroanilides. The catalytic efficiency  $k_{cat}/K_m$  of ClbPpep against these substrates were 0.005, 0.018, 0.020, and 0.040  $\mu\text{M}^{-1} \text{min}^{-1}$  respectively. This peptidase activity is weak in comparison to that of enzymes really adapted to these substrates such as S12 aminopeptidases (32). However, in a ClbPpep mutant harboring the substitution Y186E, there was no activity, which confirms the involvement of the ClbP binding site in the peptidase activity.

Imipenem is a NRP-type compound belonging to the chemical group of  $\beta$ -lactams, which efficiently bind members of the S12 enzyme family, such as carboxypeptidase and AmpC  $\beta$ -lactamases (35,36). ClbPpep incubated with imipenem was analyzed by mass spectrometry (Suppl. Fig. 3). Imipenem induced a mass shift consistent with the formation of imipenem acyl-enzyme complex of ClbPpep, as observed for  $\beta$ -lactam-binding proteins (37). Despite a high concentration of imipenem (ratio imipenem/ClbPpep,  $>1/1000$ ),

part of ClbPpep exhibited a mass corresponding to the native protein. The results show that ClbP is able to bind NRP-type compounds, has a moderate  $\beta$ -lactam-recognizing activity, and was devoid of significant  $\beta$ -lactamase activity. The  $k_{cat}$  value ( $< 0.001 \mu\text{M}^{-1} \text{s}^{-1}$ ) was around 1,000,000-fold lower than typical values of AmpC enzymes (38). Consequently ClbP in contrast to AmpC was not able to increase the resistance level of *E. coli* to  $\beta$ -lactam antibiotics.

**ClbP is an SRP/Sec/YidC- and MreB-dependent inner membrane protein with an N-terminal periplasmic domain.**

The putative signal peptide and the translocation pathway of ClbP were investigated to determine its cellular location and topology. A ClbP-C-terminal His fusion protein was expressed in *E. coli* laboratory strain TG1 and western-blot analysis was performed on whole cell lysates and cellular fractions to localize ClbP (Suppl. Fig. 4A). We observed a reactive band with an apparent molecular mass compatible with ClbP-His mass since this band was missing from the whole cell lysate of *E. coli* carrying the empty plasmid. Suppl. Fig. 4A shows that ClbP-His was predominantly located in the inner membrane fraction, to a lesser extent in the spheroplast fraction but not in the periplasmic or cytoplasmic fractions. These results show that ClbP was associated to the inner membrane of *E. coli* and that its membrane insertion was independent of other *pks* island components.

The putative signal sequence of ClbP (ClbPss) plus the first six amino acids of the mature part was fused with a GFP reporter to give the fusion protein ClbPss-GFP. Western-blot analysis using a GFP-specific antibody was performed with *E. coli* TG1 cells expressing ClbPss-GFP and revealed the precursor (ClbPss-GFP) and mature (GFP) forms of the fusion protein (Suppl. Fig. 4B). This result confirmed the cleavage of the N-terminal part of the protein and the functionality of the signal sequence of ClbP.

SRP, SecY, and YidC are major components of the translocation/insertion pathways used by inner membrane proteins (39). MreB and FtsZ are proteins of the

bacterial cytoskeleton. MreB may be involved in cytoplasmic trafficking of proteins towards the Sec translocation system (40). To further analyze the membrane targeting of ClbPss, the translocation of the ClbPss-GFP fusion protein was analyzed by western blot analysis using *SRP*, *secY*, *yidC*, *mreB* and *ftsZ* mutants (Suppl. Fig. 4C). Processing was reduced in the *secY* and *yidC* conditional mutants and in the *mreB* mutant, but not in the wild-type strain and *ftsZ* mutant. No signal was observed with the SRP mutant, suggesting an absence of translocation and a degradation of the nascent peptide in the cytoplasm. Therefore, it is likely that the actin-like cytoskeleton MreB protein and the SRP/Sec/YidC pathway play an important role in the insertion of ClbP in the inner membrane and that the N-terminal cleavable signal sequence of ClbP is sufficient to address and translocate ClbP into *E. coli* inner membrane by the SRP/Sec/YidC insertion pathway.

GFP protein fused to signal sequences is a useful indicator to monitor the addressing of proteins to periplasm (40). GFP is incorrectly folded when it is translocated into the periplasm, and correctly folded when located in the cytoplasm. Only slight cytoplasmic fluorescence was observed in the *E. coli* TG1 strain after induction of ClbPss-GFP expression (Suppl. Fig. 5), suggesting that ClbPss address ClbP S12 large domain in the periplasm. Most of the mature GFP appeared to be in the periplasm (Suppl. Fig. 4B). This result confirmed that ClbP signal peptide is sufficient to address GFP to the inner membrane, after which the signal peptide is cleaved, resulting in the release of GFP into the periplasmic compartment. We can therefore propose that the signal peptide is released from the mature moiety of ClbP for a productive function of the N-terminal periplasmic S12 domain of the protein, which remains associated with the inner membrane via transmembrane helices.

**Identification of ClbP homologs across prokaryotic lineages.** Since the functionality of the *pks* island is dependent on ClbP in *E. coli* (3), similar proteins may be observed in other bacteria producing bioactive PK-NRP compounds. We therefore screened

public protein and genomic databases to identify proteins homologous to ClbP peptidase domain. 77 homologs were obtained including the protein ZmaM, which may be involved in the synthesis of the PK-NRP antibiotic zwittermicin A (ZmA) (41). These proteins only shared 29.3% (ZP\_04059476.1) to 42.7% (ZP\_04215168.1) identity and all belonged to the MEROPS S12 enzyme family. The amino acid sequences and the predicted topology of both ClbP and the homologs were compared with those of representative members of the S12 family that have been assigned to a biological function and/or a biochemical activity. The resulting phylogenetic tree and the topologies revealed that the proteins clustered in three major branches supported by high bootstrap values ( $\geq 64\%$ ) (Suppl. Fig. 6). One branch harbored the known representative members of S12 family, which were sub-clustered according to their biochemical activity ( $\beta$ -lactamases, peptidases, esterases and amidases). The second harbored the proteins FmtA-like ( $n=17$ ) of *Staphylococcus* species (27). These proteins did not cluster with ClbP, but were phylogenetically closer to ClbP than the other members of the S12 family. In addition, their predicted topology was identical to that of ClbP unlike the other S12 family members. The third branch exhibited the longest phylogenetic distance, which clearly defined a subgroup containing ClbP and most close-related proteins ( $n=60$ ). These proteins, designated ClbP-like, exhibited a predicted topology identical to that of ClbP. Some ( $n=14$ ), such as ZmaM, harbored in addition a putative ‘half-size’ ABC-type exporter domain in C-terminal. Except ClbP and its closest homolog (ClbP-like of *Hahella chejuensis*) produced by different  $\gamma$ -proteobacteria, all the other functional ClbP-like were identified in taxons of the Firmicute phylum (*Bacillus* and *Clostridium* geni), which is the main bacterial source of PK, NRP and/or hybrid PK-NRP compounds with the Actinobacteria phylum (42).

**ClbP-like-encoding genes cluster with NRP megasynthases.** PK, NRP and hybrid compounds are synthesized in bacteria by PK and NRP megasynthases, all or most of



which are encoded by genomic islands. The gene encoding ClbP and its homolog ZmaM are surrounded by such genomic islands (3,41). The genomic environment of 17 *fmtA*-like genes and 35 *clbP*-like genes were analyzed to find such genetic elements. No clues for the presence of megasynthase catalytic unit-encoding genes were revealed in the vicinity of any *fmtA*-like and 17 *clbP*-like genes. In contrast, putative NRPS- and PKS-encoding genes were observed in the vicinity of 18 *clbP*-like genes (Suppl. Fig. 7). Further investigations also revealed genes encoding for other putative proteins known to be involved in PK and NRP synthesis pathways such as PhosphoPantetheinyl Transferases (PPTases), tailoring enzymes and transporters (2,43). They form 22 to 93 kb apparent gene clusters. The organization of these genetic elements did not share obvious similarity, except for the eight ZmaM-encoding gene clusters and the two observed in *Bacillus pumilus*. However, putative NRPS-encoding genes were present in all of these genomic islands unlike PKS-encoding genes, suggesting that the cognate products contain peptide moiety that may be related to the peptidase activity of ClbP-like proteins.

**ClbP-like proteins display functional promiscuity.** The previous results suggest that ClbP-like proteins are involved in a general mechanism of NRP compound processing. To test if these enzymes are functionally equivalent, we analyze the ability of six representative ClbP-like proteins (originating from different phyla, different genes, and encoded by genes inserted or not in NRP biosynthesis cluster) and one FmtA-like to xeno-complement the cytopathic activity of the *pks*-positive *E. coli*  $\Delta clbP$  isogenic mutant. Megalocytosis, G<sub>2</sub>/M-cell cycle arrest and histone H2AX phosphorylation were restored by all ClbP-like proteins, whose gene is located in a NRPS-encoding genomic island (hereafter, ClbP-like proteins genetically linked to NRPS gene clusters from *H. chejuensis*, *Bacillus weihenstephanensis*, *Clostridium cellulolyticum* and ZmaM from *Bacillus cereus*) or not (hereafter, ClbP-like proteins genetically unlinked to NRPS gene clusters from *Bacillus mycoides* and *Bacillus*

*pseudomycooides*) (Suppl. Fig. 8). In contrast, the FmtA-like protein did not restore cytopathic activity, despite its full expression and translocation to the inner membrane (Suppl. Fig. 8). This xeno-complementation thus supports the functional promiscuity of ClbP-like proteins, which constitute a family of peptidases designated NRP-associated peptidases. Moreover, these results are evidence of substrate tolerance of NRP-associated peptidases, and confirm data obtained in experiments that have investigated residues of the ClbP catalytic pocket potentially involved in substrate binding.

**Sequence motifs and electrostatic surface properties of NRP-associated peptidases.** NRP-associated peptidases display sequence similarities with penicillin-recognizing proteins, notably the conserved catalytic motifs SxxK, Yx(N/S/T/C) and (K/H)(S/T/G)G (44). Only a few other residues are 100% conserved, but most are also observed in FmtA-like proteins (Suppl. Fig. 9) and cannot be therefore linked to the functional specificity of NRP-associated peptidases. Nine matrix-based MEME motifs were identified in the sequences of NRP-associated peptidases (Suppl. Fig. 10A), whose five were located in the catalytic pocket. Five motifs were also observed in FmtA-like proteins. Of the four others, specific to NRP-associated peptidases, three were located in the ClbP catalytic pocket and probably involved in substrate accommodation (Suppl. Fig. 10B).

We investigated the electrostatic surface properties of NRP-associated peptidases in an attempt to identify structural properties that constitute their specificity. The electrostatic potential maps obtained from the crystal structure of ClbPpep and four other closely related S12 enzymes were compared with those obtained from FmtA-like and NRP-associated peptidase homology models (Suppl. Fig. 11A-M). The catalytic pockets of NRP-associated peptidases were located in a large groove with a negative potential surface. In contrast, the binding site of FmtA-like and the other S12 peptidases presented an overall positive potential surface. The electrostatic-potential maps of NRP-associated peptidases

and FmtA-like models revealed a correlation between the catalytic pocket electrostatic potential and the capacity to restore the bioactivity of *pks* island defective for ClbP production. The catalytic pocket electrostatic potential of NRP-associated peptidases may therefore be a major determinant of their activity by affecting the accommodation of substrates and their processing.

The distribution of charges on protein surface was compared to that of substrates or putative ligands. NAG-NAM-pentapeptide, benzylpenicillin and cephalothin, which are the substrate or inhibitors of carboxypeptidases and class C  $\beta$ -lactamases, exhibited an overall negative electrostatic potential (Suppl. Fig. 11P-S). This can favor their binding into the positive-charged catalytic pocket of carboxypeptidases and class C  $\beta$ -lactamases. Interestingly, the final bioactive PK-NRP compound ZmA and its putative preprocessing counterpart pre-zwittermicin, which is proposed by Kevany and coworkers to be cleaved by ZmA<sub>M</sub> (41), exhibited overall a positive potential surface (Suppl. Fig. 11N-O). This positive potential surface matches with the negative potential surface of the NRP-associated peptidase binding site. These results highlight the structural and functional differences between NRP-associated peptidases and other S12 enzymes and suggest that the former are a subgroup of the MEROPS S12 peptidase family, which is adapted to process positive-charged NRP or hybrid PK-NRP compounds.

## DISCUSSION

The goal of this study was to identify the structure and function of the ClbP protein, which is essential to the genotoxic activity of the *pks* island. We extended investigations to ClbP-like proteins which are ClbP's closest related proteins (30 to 42% identity), and are clustered on a distinct phylogenetic branch in the MEROPS S12 enzyme family.

The crystal structure of the ClbP enzymatic domain together with mutagenesis experiments and biochemical assays revealed the same  $\alpha/\beta$  fold observed in the members of the MEROPS S12 family and a serine-reactive

active site conferring NRP-binding and peptidase activities necessary to colibactin bioactivity. The active site possessed original structural features. The shape of the ClbP active site forms an unusually wide, deep and negative-charged groove. Molecular modeling data suggest that these features are conserved in ClbP-like peptidases and differ from known structures of S12 members such as aminopeptidases, carboxypeptidases, Pab87 peptidases and AmpC  $\beta$ -lactamases (21,23,26,45). The catalytic cavity of these enzymes is smaller than that of ClbP and harbors positive charges, which match the negative charges of known substrates. In contrast, we observed that ClbP binding site is adapted to accommodate positive-charged and linear compounds such as certain PK-NRP compounds like ZmA. Bioinformatics and biochemical analyses of the ZmA biosynthesis enzymes performed by Kevany *et al.* suggest that the compound is initially biosynthesized as part of a large metabolite that is processed by a peptidase, resulting in the formation of ZmA(41). The only candidate enzyme encoded by the ZmA biosynthesis gene cluster to catalyze such a peptidase activity is ZmA<sub>M</sub>, a ClbP-like protein fused to cyclic peptide ABC exporter. These results highlight the structural differences between ClbP and other S12 peptidases and suggest that ClbP-like proteins can process positive-charged NRP or PK-NRP compounds.

ClbP and some ClbP-like-encoding genes belong to NRPS-encoding gene clusters. Of note, these clusters always contain NRPS-encoding genes. The ClbP-like proteins displayed functional promiscuity since they fully trans-complemented a *pks*<sup>+</sup> *E. coli* mutated for *clbP*. Likewise, ClbP-like peptidases shared a particular cellular localization and topology according to *in silico* analyses, namely a large N-terminal periplasmic peptidase domain anchored to the inner membrane by three C-terminal transmembrane helices. Hence, ClbP and the closest related ClbP-like proteins form a functional peptidase subgroup of S12 serine enzymes, which is probably involved in NRP or PK-NRP biosynthesis. We propose

therefore to designate this ClbP-like family the ‘NRP-associated peptidases’.

The major source of diversity in NRP compounds mainly derives from the building-block-inserting adenylation domain in each module of NRPSs (46). NRP-associated peptidases can contribute to the molecular diversity of NRP compounds. NRP cleavage by NRP-associated peptidases can be related to the action of tailoring enzymes in PK synthesis and confer another degree of structural scaffold diversity of NRP compounds. ClbP encoded by *pks* island and the other NRP-associated peptidases may process a pro-colibactin and other NRP compounds into biologically active molecules by cleaving a nonribosomal-borne peptide bond. The ClbP-like peptidase domain of ZmaM had been suggested to cleave a ‘pre-ZmA’ metabolite between the D-Asn and D-Ser, releasing ZmA and a fatty acyl-D-Asn (41). Interestingly, *in silico* analysis of NRPS substrate specificity showed that all gene clusters producing NRP-associated peptidases encode NRPS able to insert the close related residues Asn and/or Gln into putative NRP compounds, suggesting that these residues may be involved in substrate recognition by NRP-associated peptidases. Of note, ClbP-like proteins were the only proteins encoded by these islands that harbored a predicted translocation signal sequence and a predicted periplasmic catalytic domain, in contrast to other enzymes involved in biosynthesis of PK, NRP, and mixed compounds which are cytoplasmic proteins. Accordingly, NRP-associated peptidases may act outside the cytoplasm, after assembly line chains of NRP biosynthesis.

NRP-associated peptidases may delete a part of NRP compounds also involved in interactions with exporter and/or modification machinery. The antibiotic ribosomal peptides microcin V, bacteriocins and lantibiotics are produced as inactive pre-peptides including an N-terminal leader sequence, which presumably maintains the bacteriocins *lato sensu* in an inactive form within the producer cell and facilitates interactions with both modification machineries and exporters. This

leader sequence is usually cleaved during or after the export, respectively by cysteine peptidases fused to ABC transporters or by dedicated periplasmic serine peptidases (47,48).

Several NRP-associated peptidases harbor in C-term a typical ‘half-size’ ABC exporter domain, such as ZmaM (Suppl. Fig. 6). These ABC export systems are involved in the extrusion of noxious substances and the export of extracellular toxins (49). Five of the seven ABC exporters fused to NRP-associated peptidases belong to the family of cyclic peptide exporters (InterProScan accession number IPR005898). These ABC exporters were already proposed to transport NRP compounds (49). ClbP and other NRP-associated peptidases, did not harbor such an export domain, but their cognate gene clusters encode a MATE transporter (like ClbM in the *pks* island) or other efflux pumps. After the translocation of an inactive NRP compound from cytoplasm to periplasm *via* efflux pumps, the periplasmic location of NRP-associated peptidase domain may allow the maturation of the fully active toxin away from the potential cytoplasmic target of its producer cell. NRP-associated peptidases and efflux pumps can be therefore seen as a two-component resistance mechanism against the toxic activity generated by the cognate NRPS-encoding gene cluster (Suppl. Fig. 12).

NRP-associated peptidases genetically linked to NRPS gene clusters display a functional promiscuity with those genetically unlinked to NRPS gene clusters. These latter peptidases may also participate in the final step of the synthesis of a bioactive NRP compound whose most of genes encoding for its biosynthesis would be elsewhere in the genome. It had already been suggested that enzymes required for NRP and PK biosynthesis like PPTases are encoded in a distinct region of the main biosynthetic gene cluster in bacterial genomes (50). Thereby, we found one to three putative NRP biosynthetic clusters in 15 of 16 genomes possessing a NRP-associated peptidase genetically unlinked to a NRPS gene cluster.

In conclusion, ClbP is a prototype of a new subfamily of periplasmic serine-reactive peptidases probably acting as maturing enzymes for NRP compounds. We propose to designate these proteins NRP-associated peptidases. The location of ClbP and in particular its essential role in the production of the biological effect make this protein a key target for the control of the bioactivity of the PK-NRP-producing *pks* gene cluster, which

may affect commensalism, pathogenicity and constitute a predisposing factor for the development of colorectal cancer. Moreover, the engineering of NRP compounds should therefore take into account these new catalytic units, which can be interesting tools to modify the structures and the bioactivity of PK-NRP and NRP compounds, and can be a target for identifying bioactive molecules.

## REFERENCES

1. Li, J. W., and Vederas, J. C. (2009) *Science* **325**(5937), 161-165
2. Sattely, E. S., Fischbach, M. A., and Walsh, C. T. (2008) *Nat Prod Rep* **25**(4), 757-793
3. Nougayrede, J. P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006) *Science* **313**(5788), 848-851
4. Putze, J., Hennequin, C., Nougayrede, J. P., Zhang, W., Homburg, S., Karch, H., Bringer, M. A., Fayolle, C., Carniel, E., Rabsch, W., Oelschlaeger, T. A., Oswald, E., Forestier, C., Hacker, J., and Dobrindt, U. (2009) *Infect Immun* **77**(11), 4696-4703
5. Dubois, D., Delmas, J., Cady, A., Robin, F., Sivignon, A., Oswald, E., and Bonnet, R. (2010) *J Clin Microbiol*
6. Johnson, J. R., Johnston, B., Kuskowski, M. A., Nougayrede, J. P., and Oswald, E. (2008) *J Clin Microbiol* **46**(12), 3906-3911
7. Cuevas-Ramos, G., Petit, C. R., Marcq, I., Boury, M., Oswald, E., and Nougayrede, J. P. (2010) *Proc Natl Acad Sci U S A* **107**(25), 11537-11542
8. Pradel, N., Delmas, J., Wu, L. F., Santini, C. L., and Bonnet, R. (2009) *Antimicrob Agents Chemother* **53**(1), 242-248
9. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002) *Acta Crystallogr D Biol Crystallogr* **58**(Pt 11), 1955-1957
10. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J Appl Crystallogr* **40**(Pt 4), 658-674
11. Steiner, R. A., Lebedev, A. A., and Murshudov, G. N. (2003) *Acta Crystallogr D Biol Crystallogr* **59**(Pt 12), 2114-2124
12. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1), 2126-2132
13. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J App Cryst* **26**, 283-291
14. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50**(3), 437-450
15. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res* **25**(17), 3389-3402
16. Papadopoulos, J. S., and Agarwala, R. (2007) *Bioinformatics* **23**(9), 1073-1079

17. Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D. F., and Wright, F. (2009) *Bioinformatics* **25**(1), 126-127
18. Pedretti, A., Villa, L., and Vistoli, G. (2004) *J Comput Aided Mol Des* **18**(3), 167-173
19. Stewart, J. J. P. (2008) MOPAC2009. In., Stewart Computational Chemistry, Colorado Springs, CO, USA
20. Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., and Barrett, A. J. (2008) *Nucleic Acids Res* **36**(Database issue), D320-325
21. Beadle, B. M., Trehan, I., Focia, P. J., and Shoichet, B. K. (2002) *Structure* **10**(3), 413-424
22. Pratt, R. F. (2008) *Cell Mol Life Sci* **65**(14), 2138-2155
23. Delfosse, V., Girard, E., Birck, C., Delmarcelle, M., Delarue, M., Poch, O., Schultz, P., and Mayer, C. (2009) *PLoS One* **4**(3), e4712
24. Okazaki, S., Suzuki, A., Komeda, H., Asano, Y., and Yamane, T. (2008) *J Synchrotron Radiat* **15**(Pt 3), 250-253
25. Wagner, U. G., Petersen, E. I., Schwab, H., and Kratky, C. (2002) *Protein Sci* **11**(3), 467-478
26. Bompard-Gilles, C., Remaut, H., Villeret, V., Prange, T., Fanuel, L., Delmarcelle, M., Joris, B., Frere, J., and Van Beeumen, J. (2000) *Structure* **8**(9), 971-980
27. Komatsuzawa, H., Choi, G. H., Fujiwara, T., Huang, Y., Ohta, K., Sugai, M., and Suginaka, H. (2000) *FEMS Microbiol Lett* **188**(1), 35-39
28. Holm, L., and Sander, C. (1996) *Science* **273**(5275), 595-603
29. Meroueh, S. O., Minasov, G., Lee, W., Shoichet, B. K., and Mobashery, S. (2003) *J Am Chem Soc* **125**(32), 9612-9618
30. Dodson, G., and Wlodawer, A. (1998) *Trends Biochem Sci* **23**(9), 347-352
31. Matagne, A., Lamotte-Brasseur, J., and Frere, J. M. (1998) *Biochem J* **330** ( Pt 2), 581-598
32. Delmarcelle, M., Boursoit, M. C., Filee, P., Baurin, S. L., Frere, J. M., and Joris, B. (2005) *Protein Sci* **14**(9), 2296-2303
33. Chen, Y., Minasov, G., Roth, T. A., Prati, F., and Shoichet, B. K. (2006) *J Am Chem Soc* **128**(9), 2970-2976
34. Galleni, M., Lamotte-Brasseur, J., Raquet, X., Dubus, A., Monnaie, D., Knox, J. R., and Frere, J. M. (1995) *Biochem Pharmacol* **49**(9), 1171-1178
35. Matagne, A., Ghuysen, M. F., and Frere, J. M. (1993) *Biochem J* **295** ( Pt 3), 705-711
36. Beadle, B. M., and Shoichet, B. K. (2002) *Antimicrob Agents Chemother* **46**(12), 3978-3980
37. Drawz, S. M., Babic, M., Bethel, C. R., Taracila, M., Distler, A. M., Ori, C., Caselli, E., Prati, F., and Bonomo, R. A. (2010) *Biochemistry* **49**(2), 329-340
38. Bauvois, C., Ibuka, A. S., Celso, A., Alba, J., Ishii, Y., Frere, J. M., and Galleni, M. (2005) *Antimicrob Agents Chemother* **49**(10), 4240-4246
39. Yuan, J., Zweers, J. C., van Dijl, J. M., and Dalbey, R. E. (2010) *Cell Mol Life Sci* **67**(2), 179-199
40. Drew, D., Sjostrand, D., Nilsson, J., Urbig, T., Chin, C. N., de Gier, J. W., and von Heijne, G. (2002) *Proc Natl Acad Sci U S A* **99**(5), 2690-2695

41. Kevany, B. M., Rasko, D. A., and Thomas, M. G. (2009) *Appl Environ Microbiol* **75**(4), 1144-1155
42. Donadio, S., Monciardini, P., and Sosio, M. (2007) *Nat Prod Rep* **24**(5), 1073-1109
43. Martin, J. F., Casqueiro, J., and Liras, P. (2005) *Curr Opin Microbiol* **8**(3), 282-293
44. Goffin, C., and Ghuysen, J. M. (2002) *Microbiol Mol Biol Rev* **66**(4), 702-738, table of contents
45. Lee, W., McDonough, M. A., Kotra, L., Li, Z. H., Silvaggi, N. R., Takeda, Y., Kelly, J. A., and Mobashery, S. (2001) *Proc Natl Acad Sci U S A* **98**(4), 1427-1431
46. Fischbach, M. A., Walsh, C. T., and Clardy, J. (2008) *Proc Natl Acad Sci U S A* **105**(12), 4601-4608
47. Cotter, P. D., Hill, C., and Ross, R. P. (2005) *Nat Rev Microbiol* **3**(10), 777-788
48. Duquesne, S., Destoumieux-Garzon, D., Peduzzi, J., and Rebuffat, S. (2007) *Nat Prod Rep* **24**(4), 708-734
49. Saurin, W., Hofnung, M., and Dassa, E. (1999) *J Mol Evol* **48**(1), 22-41
50. Lu, Y. W., San Roman, A. K., and Gehring, A. M. (2008) *J Bacteriol* **190**(20), 6903-6908

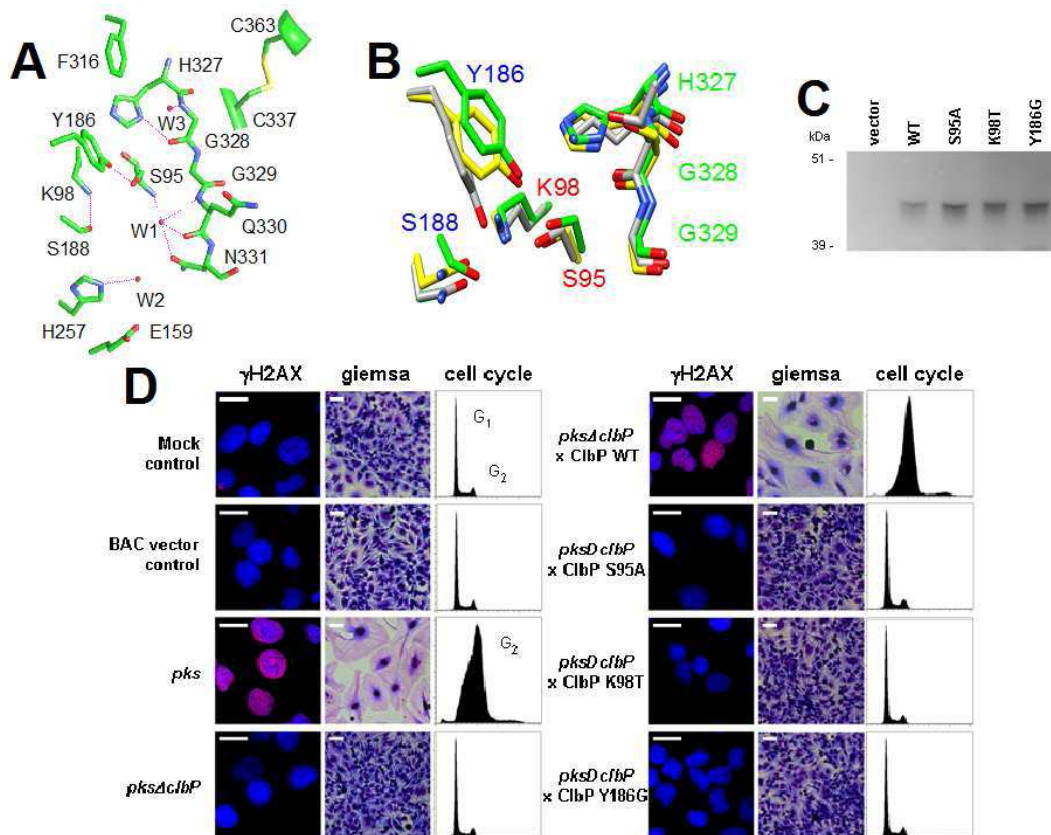
#### FOOTNOTES

This paper is dedicated to the memory of our recently deceased friend and colleague Dr. Olivier Baron. We thank R. Perroux, M. Jan and C. Habouzit for technical assistance, Alexei Sorokin (INRA, Jouy en Josas, France), Michael G. Thomas (University of Wisconsin, Madison, USA), and Jihyun F. Kim (Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea) for providing *B. weihenstephanensis* KBAB4 strain, *B. cereus* UW85 strain and fosmid clone containing *hch\_03446* locus tag of *H. chejuensis* KCTC 2396<sup>T</sup>, respectively. We also thank C. Blavignac and C. Sczepaniak (CICS, Université d'Auvergne, France) for technical assistance in flux cytometry. This work was supported by le Ministère Français de l'Education Nationale, de la Recherche et de la Technologie, grant number JE2526 and l'Institut National de la Recherche Agronomique, grant number USC-2018. Molecular graphics images were produced using the UCSF Chimera package (NIH P41 RR-01081).

## FIGURE LEGEND

**Fig. 1.** Key residues in ClbP active site. (A) Residues of ClbP located in close vicinity of the binding site. Hydrogen bonds are indicated by pink dash lines and the water molecules by the letter W. Carbon atoms of ClbP, Pab87 peptidase and AmpC  $\beta$ -lactamase are in green, yellow and grey respectively; oxygen atoms are in red and nitrogen atoms in blue. (B) Overlay of penicillin-recognizing motifs of ClbP, Pab87 peptidase and AmpC  $\beta$ -lactamase. The motifs SxxK, Yx(N/S/T) and (K/H)(S/T/G)G are numbered in red, blue and green respectively and residue numbering is according to ClbP sequence. (C, D) Phenotypic analysis of *pks* island activity in *E. coli* DH10B/pBAC*pks* mutated or not for the *clbP* gene and trans-complemented with ClbP wild-type (WT) protein or the S95A, K98T or Y186G active site ClbP mutants. HeLa cells were infected 4 hours with 100 *E. coli* per cells, and then the cells were washed and incubated with gentamicin for 4-72 hours. Panel C: anti-His western-blot analysis of the expression of His-tagged WT and active site ClbP mutants in *E. coli* DH10B/pBAC*pks* $\Delta$ *clbP* after infection. Panels D: Host histone H2AX Ser139 phosphorylation ( $\gamma$ H2AX) indicative of DNA double strand breaks was assayed by confocal immunofluorescence 4 hours after infection. DNA and  $\gamma$ H2AX are pseudocolored in blue and red respectively (Bars = 20  $\mu$ m). Cell swelling (megalocytosis) was observed following Giemsa staining 3 days after infection (Bars = 50  $\mu$ m). G<sub>2</sub>-cell cycle arrest following DNA damage was assessed by flow cytometric DNA-content analysis, 48 hours after infection.

Figure 1.



## SUPPLEMENTAL DATA

### Supplemental experimental procedures

**Strains, plasmids, and media.** The strains and plasmids used in this study are listed in Suppl. Tables 3 and 4, respectively. *E. coli* strains were routinely grown at 37°C in Luria Bertani (LB) broth or on LB agar plates. Ampicillin (Amp) (100 µg/ml), kanamycin (50 µg/ml), chloramphenicol (25 µg/ml), glucose (0.2%), arabinose (0.2%), lactose (0.2%), IPTG (1 mM), or tetracyclin (200 ng/ml) were added as required to the medium. Precultures were grown from single colonies and used at 100-fold dilutions (50-fold for the YLS-3 mutant) for inoculation of the cultures. Non *E. coli* strains were grown for phenol-chloroform DNA extraction as recommended by the DSMZ or CIP collections, or by the givers.

**DNA cloning for ClbP investigations.** The DNA sequences of *clbP* were amplified from the genome of *E. coli* strain IHE3034<sup>1</sup>. The DNA fragments encoding for the entire ClbP, the putative signal sequence of ClbP plus the first six amino acids of the mature part (ClbPss), and the peptidase domain of ClbP designated ClbPpep were amplified by PCR using clbPMK-F and clbPBN-R, clbPMK-F and clbPssBN-R, clbP-NdeI2-F and clbP-NotI2-R as primers (Suppl. Table 5), respectively. The PCR reactions for cloning experiments were performed using the High Fidelity Platinum<sup>®</sup> *Taq* polymerase or the Expand High Fidelity PCR System according to the manufacturer's instructions (Invitrogen, Cergy-Pontoise, France). The amplified fragments were purified, double digested by *MfeI/BamHI* for cloning at the *EcoRI/BamHI* sites of pASK-IBA33plus to obtain ClbP-RGS6His (ClbP + C-terminal in-frame RGS6His tag), double digested by *MfeI/NheI* for cloning at the *EcoRI/NheI* sites of p8761 to obtain p61ssClbP (ClbP signal sequence + C-terminal in-frame GFPmut2), or double digested by and cloned at the *NdeI/NotI* sites of pET9am to obtain pClbPpep (ClbP peptidase domain). In other experiments, ClbP-6His was subcloned from pMB702 (1) into pBRSK vector (2) to obtain pOB901. The constructions were checked by double-stranded DNA sequencing (Genome Express, Grenoble, France). Plasmids were subsequently electroporated into competent *E. coli* strains: p33ClbP into TG1 and DH10B/pBAC*pksΔclbP* strains for cellular localization of ClbP and HeLa cellular challenges, respectively; pOB901 into DH10B/pBAC*pksΔclbP* strain for HeLa cellular challenges; p61ssClbP into TG1, MC4100A, HPT406A, CU164A, MC1060, MYC-cs, YLS-3, and YLS-24 *E. coli* strains for ClbPss-GFP analysis; and pClbPpep into BL21(DE3) strain for ClbPpep and derivative overexpression (original strains are listed in Suppl. Table 3).

**DNA cloning of ClbP homolog-encoding genes.** The DNA fragments encoding for ClbP homologues of *H. chejuensis* KCTC 2396<sup>T</sup> (locus tag *HCH\_03446*), *B. cereus* UW85 (locus tag *zmaM*), *B. weihenstephanensis* KBAB4 (locus tag *BcerKBAB4\_1783*), *C. cellulolyticum* H10 (locus tag *Ccel\_2389*), *B. mycoides* DSM 2048 (locus tag *bmyco0001\_51990*), *B. pseudomycooides* DSM 12442 (locus tag *bpmx0001\_2150*), and *S. aureus* Mu50 (locus tag *SAV2441*) were amplified by PCR using primers listed in Suppl. Table 5, purified, double digested by *MfeI* or *EcoRI* (according to primer name) and *BamHI* or *XhoI* for cloning at the *EcoRI/BamHI* or *XhoI* sites of pASK-IBA33plus (C-terminal in-frame with RGS6His tag) resulting in plasmids listed in Suppl. Table 4. The constructions were checked by double-stranded DNA sequencing (Genome Express, Grenoble, France). Plasmids were electroporated into competent *E. coli* DH5αF' before isolation by miniprep. Plasmids were subsequently electroporated into competent final *E. coli* strain DH10B/pBAC*pksΔclbP*.

**Site-Directed Mutagenesis.** The site-directed mutagenesis experiments were performed using Quick-Change<sup>™</sup> Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA). The substitutions E159K, Y176G, Y186E, S188A, H257A, G328S, G328S+F316R, N331A and C337A were individually introduced into ClbP cloned into p33ClbP plasmid and the substitution Y186E were introduced into ClbPpep cloned into pClbPpep plasmid. Reaction mixtures consisted of 20 ng template DNA, 125 ng each primer, 500 mM dNTPs, 1x buffer, and 2.5 U of *PfuTurbo* DNA polymerase in a final volume of 50 µl. The PCR program was as follows: initial denaturing at 95°C for 30 s, followed by 18 cycles at 95°C for 30 s, 55 to 60°C (according to the primer pairs) for 1 min, and 68°C for 14 min, and completed by



an additional 14 min at 68°C. Upon completion, 10 units of *DpnI* were added directly to the PCR mixture and digested at 37°C for 1 h. The mixture (1 µl) was directly transformed into DH5α and plated on LB supplemented with antibiotic according to the vector resistance. In other experiments, ClbP S95A, K98T and Y186G mutants were generated by ‘reverse PCR’ using pOB901 as template. Reverse PCR products were purified, ligated and transformed into DH5α. Successful mutations were confirmed by double DNA sequencing and the final constructs were used to transform *E. coli* DH10B/pBAC*pksΔclbP* for p33clbP and pOB901 derivative vectors, or *E. coli* BL21(DE3) for pClbPpep derivative vector. The mutagenic primers are listed in Suppl. Table 6.

**Cell induction assays of translocation pathway, *MreB* and *FtsZ* mutants.** Cell induction assays of translocation pathway were performed as previously described (3-5). We used the cold-sensitive *secY-cs* and *yidC* conditional mutants CU164A and MYC-cs, in which protein export is partially slowed down at 37°C and is severely affected upon exposure to 23°C. The following cold-sensitive and -insensitive strains harboring p61ClbPssGFP were all incubated in LB-ampicillin-glucose (0.2%) medium at 37°C (permissive temperature for cold-sensitive strains) with shaking until early exponential phase: MC4100A, MC1060 and MC1000 wild-type, cold-sensitive conditional translocation mutant CU164A ( $\Delta secY$ ) and MYC-cs ( $\Delta yidC$ ), and HPT406A (SRP mutant), YLS-3 ( $\Delta mreB$ ) and YLS-24 ( $\Delta ftsZ$ ) strains. The cultures were put on ice for 3 min and then kept at 23°C (nonpermissive temperature) for 1 h for conditional mutant induction. The cells were harvested by centrifugation, washed once with LB, and resuspended in LB-ampicillin-arabinose 0.2% medium and incubated a further 3 h at 23°C with shaking for induction of ClbPss-GFP expression.

**Cellular fractionation experiments.** Cellular fractionation experiments were performed as previously described (4,5). Crude extracts were prepared by centrifugation of 1 ml culture at OD<sub>600</sub>=0.5. The pellets were suspended in 0.1 ml Tris-HCl buffer (40 mM, pH 8) and 0.02 ml 1% protamine sulfate. Periplasm and spheroplasts were prepared by lysozyme/EDTA/cold osmohock method and separated by centrifugation. Membranes and cytoplasmic soluble fractions were prepared from spheroplasts by breaking the cells by sonication and separated by ultracentrifugation at 120 000 x g for 90 min. Protease inhibitor Complete (Roche Diagnostics, Mannheim, Germany) was added at the maximum recommended concentration in all buffers used.

**Western-blot experiments.** Protein samples (20-25 µg for each sample) were resolved by SDS-PAGE separation electrophoresis on 12.5% polyacrylamide gels, immobilized onto polyvinylidene difluoride (PVDF) membranes and analyzed by immunoblot using the ECL blotting system according to the manufacturer’s instructions (Amersham Biosciences, Buckinghamshire, UK). Antisera were polyclonal rabbit anti-GFP (Invitrogen) at a dilution of 1:10,000 and monoclonal anti-RGS6His or anti-5His, which were used according to the manufacturer’s instructions (Qiagen, Courtaboeuf, France).

**ClbPpep and ClbPpepY186E production and purification.** pClbPpep plasmid encodes the peptidase domain of ClbP without the first 102 nucleotides and the last 390 nucleotides of the *clbP* gene. These sequences encode respectively for the signal peptide and three putative anchor hydrophobic transmembrane helices that were removed to facilitate cytoplasmic expression and protein purification, respectively. *E. coli* BL21(DE3) transformed with pClbPpep plasmid was cultured in Terrific Broth medium supplemented with 50 µg/ml kanamycin, 0.4 M D-sorbitol, and 2.5 mM β-betaine. The cell culture was grown at 37°C with shaking to an  $A_{600\text{ nm}} \sim 0.7$ . Overexpression of the *clbPpep*-encoding genes was induced with 0.2 mM isopropyl b-D-thiogalactopyranoside overnight at 25°C with shaking. Cells were harvested by centrifugation (10,000 g for ten minutes at 4°C), resuspended in 20 mM Tris-HCl buffer (pH 9.0) (4 mL/g of pellet). After addition of 5 µg lysozyme (2 hours at 4°C), cells were disrupted by ultrasonic treatment (four times for 30 seconds, each time at 20 W). The extract was clarified by centrifugation at 14,000g for 60 minutes at 4°C. After addition of 2 mg DNase I (15 min at room temperature; Roche), the supernatant (10,000 g for 60 minutes at 4°C) was dialyzed overnight against 20 mM Tris-HCl buffer (pH 9.0) and filtrated (0.22µm filter; Millex-GS, Millipore, Carrigtwohill, Ireland). Purification was carried out by ion-exchange chromatography onto a HiTrap™

Q Sepharose™ High Performance column (10 ml; Amersham Pharmacia Biotech) equilibrated with 20 mM Tris-HCl (pH 7.0) and eluted with a linear NaCl gradient (0 to 500 mM). The ClbPpep-containing elution peak was then purified by gel filtration chromatography onto a Superose™ 12 column (3.2 by 30 cm; Amersham Pharmacia Biotech.), which had been equilibrated and eluted with 100 mM NaCl 5 mM phosphate buffer (pH 8.0). ClbPpep-containing elution peak was extensively dialyzed against 50mM NaCl, 5 mM phosphate buffer (pH 8.0), concentrated by ultrafiltration to 10 mg/ml for crystallization. The total protein concentration was estimated by the Bio-Rad protein assay (Bio-Rad, Richmond, Calif.), with bovine serum albumin (Sigma Chemical Co.) used as a standard. The enzyme was more than 95% homogeneous as determined by Coomassie blue staining after SDS-PAGE. The N-terminal sequencing of the isolated protein was achieved by sequential Edman reaction carried out at the (Plate-forme de Protéomique Analytique et Fonctionnelle) Institut National de Recherche Agronomique (Nouzilly, France). Production and purification of ClbPpepY186E was performed in the same way using the plasmid pClbPpepY186E.

**Crystallization.** ClbP crystals were grown in hanging drops over a well solution of 0.8 M monosodium dipotassium phosphate buffer (pH 7.0). The drops comprised 2 µl of protein solution (10 mg/ml in 50 mM NaCl 5 mM phosphate buffer pH 8.0) and 2 µl of well solution. Crystals reached a size of 40 µm X 40 µm X 70 µm within four weeks. Before data collection, crystals were immersed for about 30 s in a cryoprotectant solution (20% sucrose, 1.0 M monosodium dipotassium phosphate pH 7.0) and were flashcooled in liquid nitrogen.

**Mass spectrometry. Sample preparation.** ClbPpep (20 µM) was incubated at 20°C for 24 h in a 100 mM Tris, 75 mM NaCl buffer (pH 7.5) alone, or with 20 mM imipenem (Sigma), in a 1 mL final volume. The reaction mixtures were desalted by dialysis against a 5 mM ammonium bicarbonate buffer, before analysis by matrix-assisted laser desorption/ionisation–time of flight (MALDI-TOF)-mass spectrometry. For this, the sample was spotted onto the stainless steel target (Applied Biosystems) and an equal volume of matrix solution (14mg/mL sinapinic acid in 50% acetonitrile in water containing 0.1% trifluoroacetic acid) was deposited into the sample drop. **Instrument settings.** MALDI-TOF mass spectra were acquired on a Voyager DE-PRO mass spectrometer (Applied Biosystems, Framingham, MA, USA) in linear positive mode, using an accelerating voltage of 25kV, a grid voltage of 93%, a pulse delay extraction of 700ns, in the 15,000–50,000 m/z mass range. Calibration was performed in a close external mode using a protein calibration mixture (LaserBio Labs, Sophia-Antipolis, France).

**Eukaryotic cell culture, bacterial infections,  $\gamma$ H2AX staining and cell cycle analysis.** Cell assays were performed as previously described, with minor modifications<sup>1</sup>. For bacterial infections, overnight LB-amoxicillin-tetracycline cultures of bacteria were diluted in interaction medium (DMEM, 5 % fetal calf serum (FCS), 25 mM HEPES, amoxicillin, tetracyclin). ~50% confluent HeLa cell cultures (ATCC CCL2) were then infected with a multiplicity of infection (MOI) of 100. Cells were washed 4 h after inoculation and incubated in DMEM 10% FCS 200 µg/ml gentamicin until analysis. Cell morphology was examined by Giemsa staining at 72 hours after infection. For  $\gamma$ H2AX staining, cells were fixed at 4 or 20 hours after infection in 4% formaldehyde, and incubated with anti-phospho(Ser139)-H2AX antibodies (JBW301, Upstate) followed by FITC- or rhodamine-conjugated secondary antibodies. DNA was stained with DAPI or TO-PRO-3 (Invitrogen) and images were acquired with a Zeiss LSM 510 Meta or Olympus IX70 laser scanning confocal microscope, the confocal aperture being set to achieve a z optical thickness of ~0.5 µm. For cell cycle analysis, cellular suspensions were fixed and permeabilized with ethanol 70% at 48 hours after infection, then treated with 50 µg/ml propidium iodide and 250 µg/ml RNase. DNA-content data were acquired with a SLR II or FACScalibur flow cytometer (Becton Dickinson) and analyzed with FlowJo software (Tree Star).

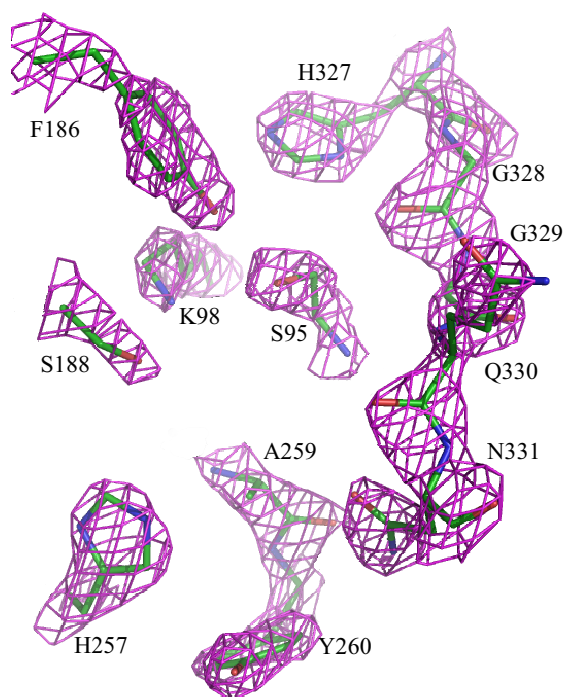
**Motif consensus elaboration.** The MEME (<http://meme.sdsc.edu/meme/meme.html>) software program (version 4.1.0) was used for the identification of motifs from ClbP-like sequences exhibiting trans-complementing activity (6,7). We then applied the MAST program (version 4.0)(8), using the motif weight matrix from the MEME program, to search for the motif in all FmtA-like and ClbP-like sequences as well as for a random sequence model based on the letter composition of the target

sequence. These scores are translated into position  $P$  values, which represent the probability of a randomly selected position in a randomly generated sequence having a match score at least as large as that of the given position. The best (i.e., lowest) position  $P$  values are then adjusted to take into account the length of the sequence. Only the motif hits with a position-specific goodness-of-fit  $P$  value of less than  $10^{-4}$  were considered to identify the motif in the sequence.

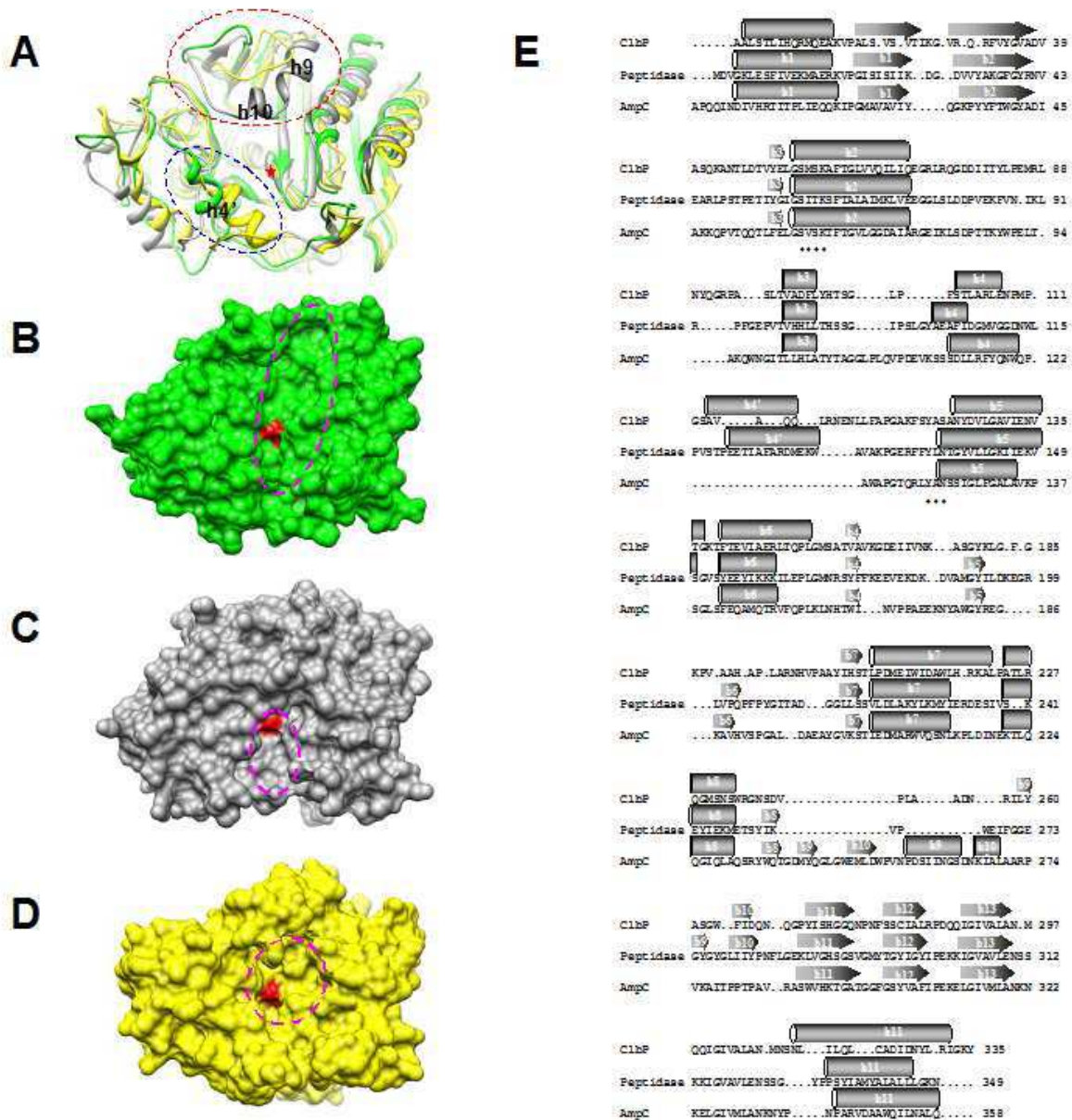
**Molecular modeling.** The crystal structure of representative MEROPS S12 enzymes (1EI5, 2EFU, 1HVB, 1CI8, 2QMI, 3BLS, 1FR1, 1GCE) and ClbPpep were used as inputs for the program MODELLER (9). Ten models were generated and then energy minimized using the CHARMM molecular dynamics and simulation procedure (10) in the MODELLER program. The models with the lowest value of the objective function score were selected as the representative models. The models were then protonated with PROPKA using the PDB2PQR server (11) and the protons were subjected to energy minimization (100 steps steepest descent) employing the MMTK routine (12) under CHIMERA (13) with AMBER parameters (14). The quality and stereochemistry of the models were evaluated using the PROCHECK program (15). The main-chain conformations were within the favored or allowed regions of the Ramachandran plot for >98% residues. The APBS program (16) was used to calculate the electrostatic surface potentials of protein models, which were visualized with the PyMol program.

The crystal structures of *N*-acetylglucosamine-*N*-acetylmuramic disaccharide-pentapeptides (NAG-NAM-pentapeptides), benzylpenicillin and cephalothin were obtained in RCSB data bank from entries 1P4N, 2EAX, 3HUO, and 1KVL. Coordinates for Zwittermicin A and Pre-Zwittermicin (17) were obtained using CACTVS (<http://www2.ccc.uni-erlangen.de/software/cactvs/>) and the CORINA server (18).

## Supplemental Figures

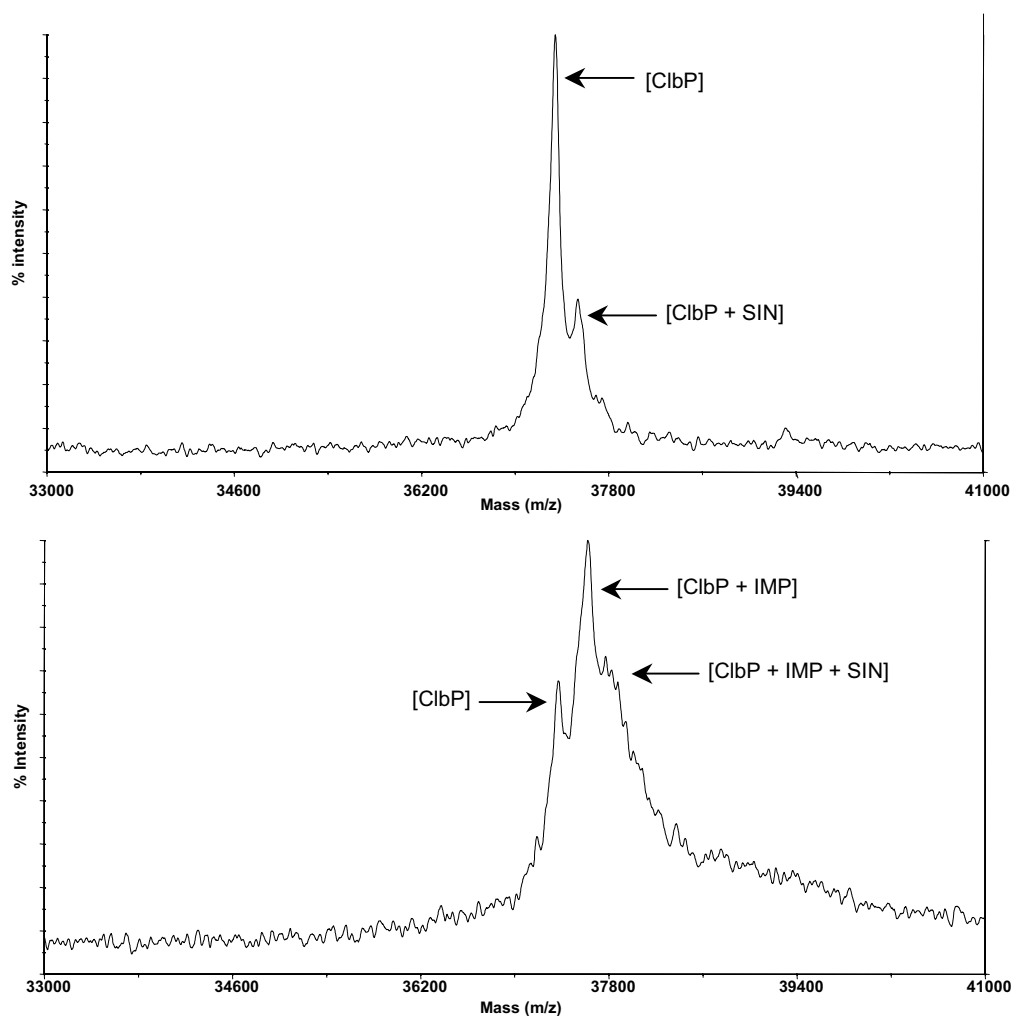


Supplemental Figure 1. Electron density map of ClbPpep in the vicinity of the active site. The 2Fo-Fc electron density of the refined model, contoured at  $1\sigma$ , is shown in mauve. Carbon atoms are in green, oxygen atoms in red, and nitrogen atoms in blue.



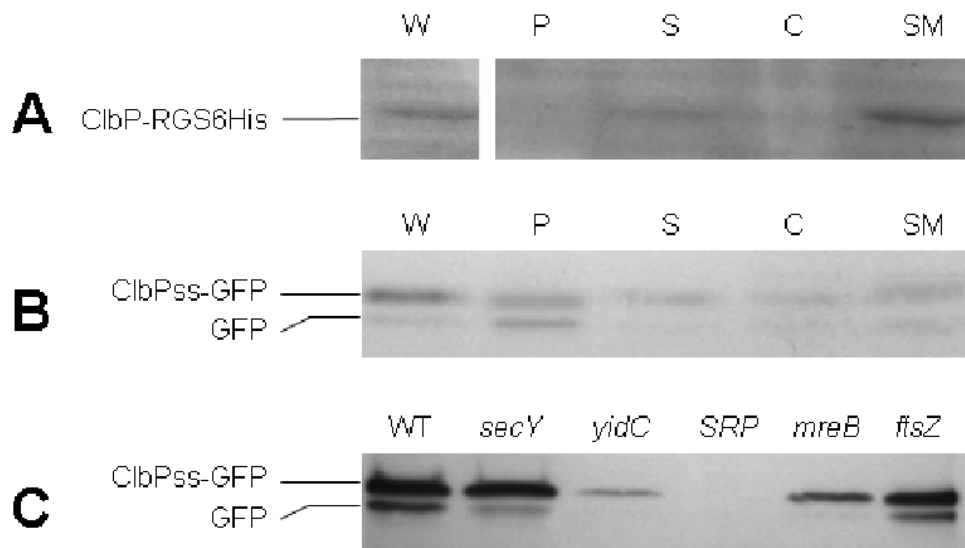
**Supplemental Figure 2.** Overall structure of ClbPpep, AmpC  $\beta$ -lactamase (1KVL) and Pab87 peptidase (2QMI).

(A) Superimposition of ClbPpep (green), AmpC (grey) (19) and Pab87 peptidase (yellow) (20) presented as ribbons. The loops including helices h9 and h10 are circled in red. Helices h4' are circled in blue. The active serine is indicated by a red star. ClbPpep (B), AmpC (C) and Pab87 peptidase (D) surfaces. The catalytic grooves are circled in pink and the active serine surfaces are shown in red. (E) Sequence alignments with secondary structures of ClbPpep with Pab87 peptidase and *E. coli* AmpC  $\beta$ -lactamase. The alignment was performed based on the superposed structures. The motifs SxxK and YxN, which defined the MEROPS S12 peptidase family, are indicated by a diamond symbol.



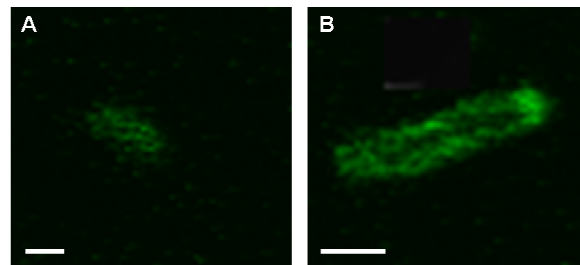
Supplemental Figure 3. Linear mode MALDI-TOF mass spectra of ClbPpep incubated alone (top) or with imipenem (bottom).

ClbP alone had a mass of 37341 Da, which is consistent with the theoretical mass (37345 Da). Heterogeneity in the spectrum was observed, due to a reaction with the sinapinic acid matrix. After incubation with imipenem (IMP), the spectrum was considerably modified. The compound was more heterogeneous, as indicated by a broader peak. The unmodified compound was present as a minor peak, whereas the major compound had + 246 Da mass shift, which is consistent with an acyl-enzyme complex of ClbPpep with imipenem, as observed for penicillin-binding proteins. In spite of a high concentration of imipenem (20 mM), part of ClbPpep exhibited a mass corresponding to the native protein, which suggests a low binding capacity.



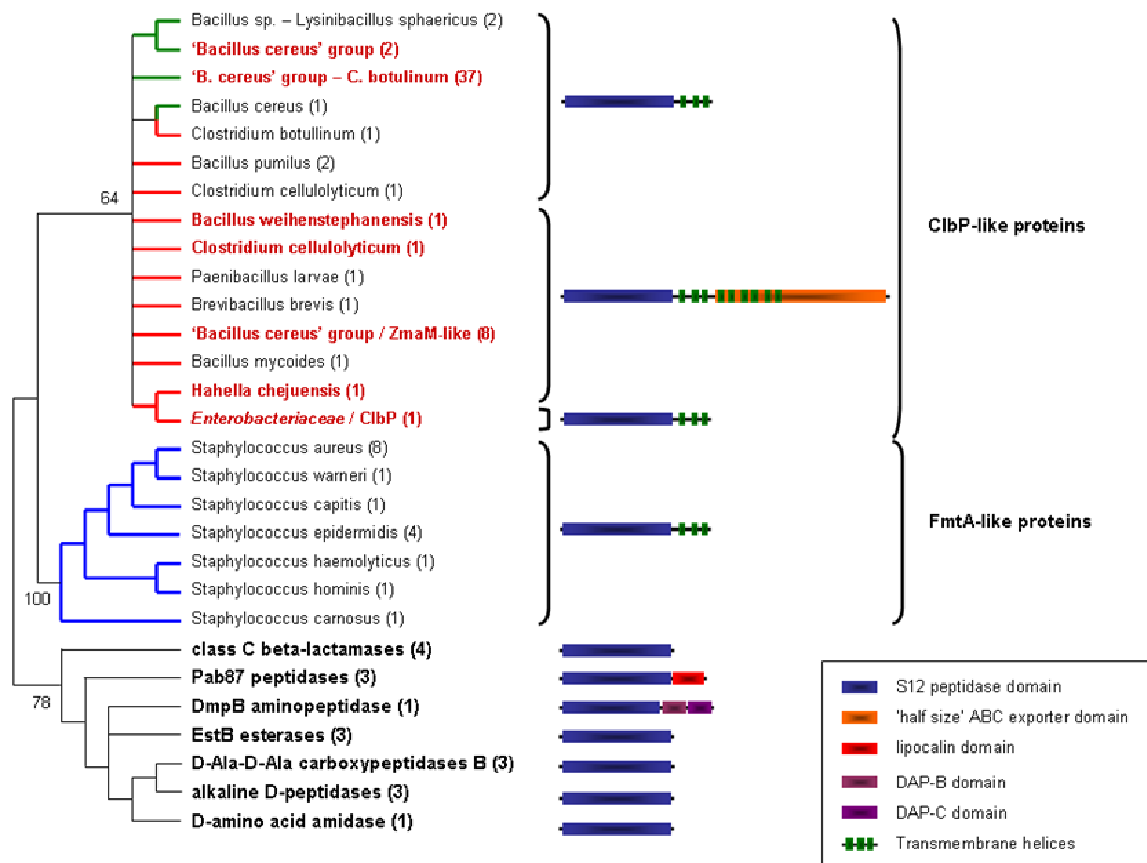
**Supplemental Figure 4. ClbP cellular localization and topology.**

(A) Inner membrane localization of ClbP-His in *E. coli* K-12 TG1 strain by cellular fractionation and western blot analysis using a monoclonal anti-His. (B) Cellular distribution of the ClbPss-GFP precursor and its processed mature counterpart GFP in *E. coli* TG1 by cellular fractionation and western blot analysis. (C) Processing of ClbPss-GFP fusions in both wild-type *E. coli* strain (WT) and *secY*, *yidC*, *SRP*, *mreB* and *ftsZ* *E. coli* mutant strains by cellular fractionation and western-blot analysis. W, whole cells, P, periplasmic, S, spheroplast, C, cytoplasmic, and SM, solubilized inner membrane fractions. Precursor (ClbPss-GFP) and mature forms (GFP) of the fusion protein are indicated.



**Supplemental Figure 5. Localization of ClbPss-GFP (A) and TorAss-GFP (B) fusions in *E. coli* TG1 strain.**

Scale bar, 1  $\mu$ m. The trimethylamine *N*-oxide reductase (TorA) of *E. coli* is a soluble periplasmic protein synthesized as a precursor possessing a cleavable N-terminal signal sequence which contains a twin-arginine motif (TorAss)(5). N-terminal twin-arginine motif-containing signal sequence allows protein translocation across the bacterial cytoplasmic membrane by Tat system in a folded state, unlike the Sec translocation machinery (21,22). Periplasmic fluorescence was observed in the *E. coli* TG1 strain with induced TorAss-GFP expression whereas only little cytoplasmic fluorescence was observed in the *E. coli* TG1 strain with induced ClbPss-GFP expression.

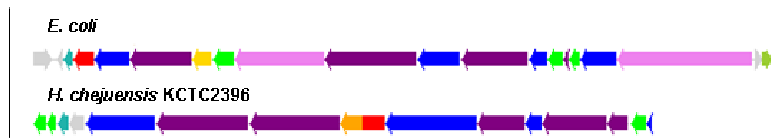


**Supplemental Figure 6.** Consensus phylogenetic tree comprising ClbP, its homologs and representative members of the MEROPS S12 peptidase family.

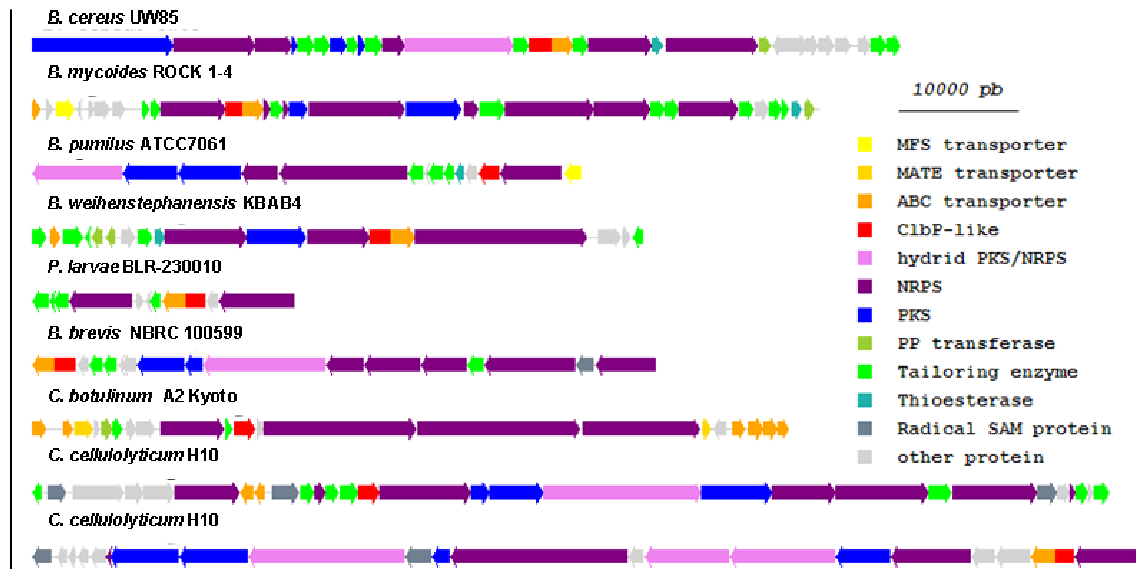
ClbP and its homologues are designated by the name of the encoding bacterial taxons. The other members of the MEROPS S12 peptidase family are designated by their biochemical activity. The figure summarizes 100 multiple bootstrapped replicates. Sequences sharing >90% identity were collapsed into one branch. The number of collapsed sequences is indicated in parentheses. Red and green branches indicate ClbP-like proteins genetically linked or not to NRPS-encoding gene cluster respectively. The taxon names in red correspond to proteins that xeno-complement the *pks*<sup>+</sup> *E. coli* *clbP* mutant strain. Protein predicted topology is depicted on the right.



**$\gamma$ -Proteobacteria**

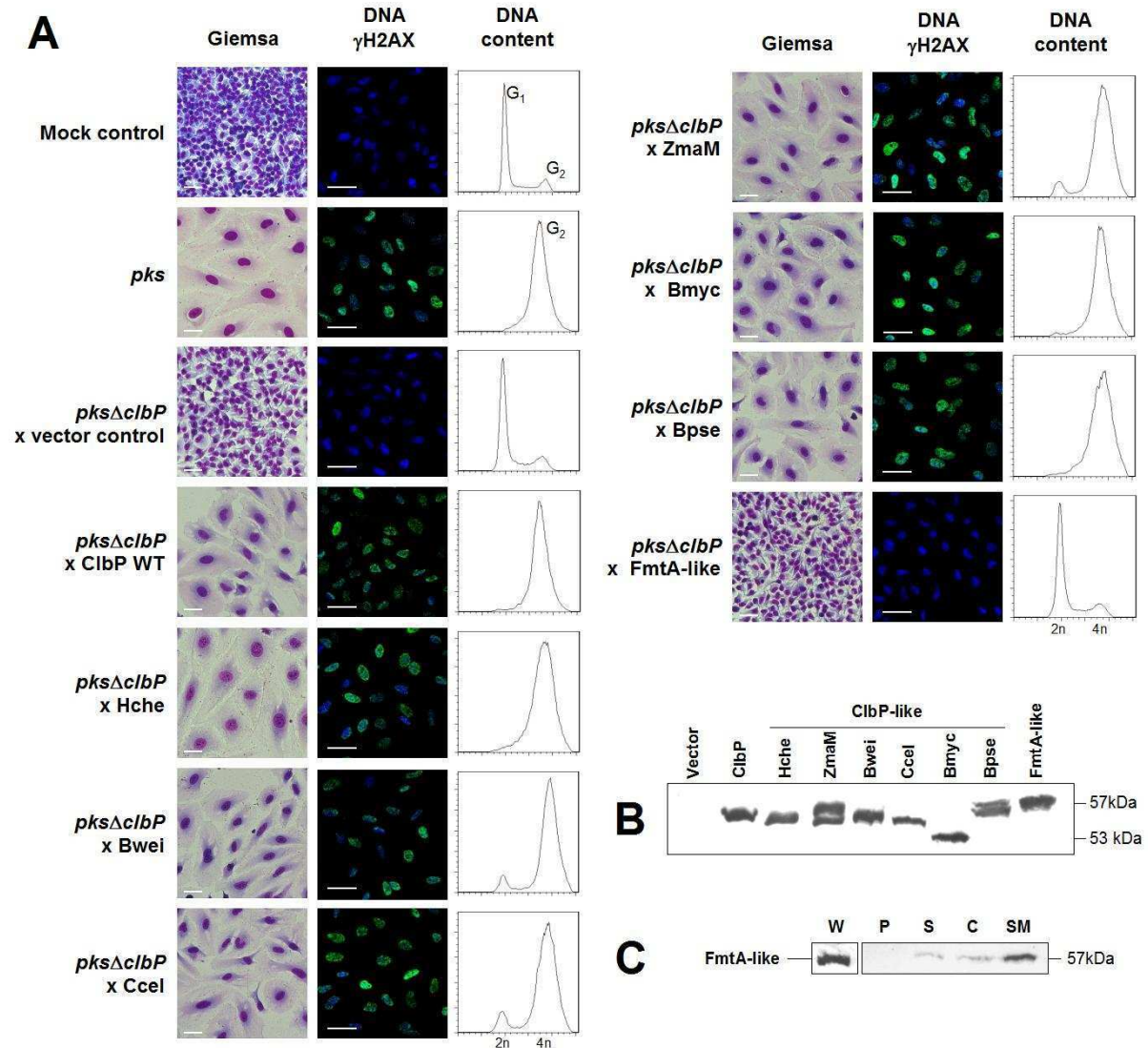


**Firmicutes**



Supplemental Figure 7. Gene clusters containing ClbP-like-encoding sequences (red).

Members are sorted by prokaryotic phylum. The ClbP-like-encoding genes produce both single-domain proteins and fusion proteins with putative ‘half size’ ABC exporters. Each cluster contains NRPS and tailoring enzymes. Relative gene length and directionality are shown.



**Supplemental Figure 8.** Functional promiscuity displayed by ClbP-like proteins.

(A) Phenotypic analysis of *pks* island activity in *E. coli* strains DH10B/pBAC*pks* (*pks*), DH10B/pBAC*pksDeltaClbP* trans-complemented with empty vector, ClbP wild-type (ClbP WT), six ClbP-like proteins from *H. chejuensis* (Hche), *B. weihenstephanensis* (Bwei), *C. cellulolyticum* (Ccel), *B. cereus* (ZmaM), *B. mycoides* (Bmyc) and *B. pseudomycooides* (Bpse), or one FmtA-like protein from *S. aureus*. ~50% confluent HeLa cells were infected 4 hours with MOI 100 of *E. coli* per cell. Left panels: The cell morphology was observed following Giemsa staining at 72 hours after infection (Bars = 50  $\mu$ m). Middle panels: Histone H2AX Ser139 phosphorylation ( $\gamma$ H2AX) indicative of DNA double strand breaks was assayed by confocal immunofluorescence at 20 hours after infection. DNA and  $\gamma$ H2AX are pseudocolored in blue and green, respectively (Bars = 50  $\mu$ m). Right panels: G2-cell cycle arrest following DNA damage was assayed by flow cytometry 48 hours after infection. (B) Anti-His western-blot analysis of the expression of His-tagged ClbP wild-type protein, ClbP-like proteins, and FmtA-like protein in *E. coli* DH10B/pBAC*pksDeltaClbP*. (C) Inner membrane localization of FmtA-like protein from *S. aureus* in *E. coli* DH10B/pBAC*pksDeltaClbP* by cellular fractionation and anti-His western-blot analysis. W, whole cells, P, periplasmic, S, spheroplast, C, cytoplasmic, and SM, solubilized inner membrane fractions.

ClbP .....aalstlihqrmqeakvpals.vs.vtikg.vr.q.rfvygvadv 39  
 ClbP-like H. chejuensis .....qtvself.rrwdasqiptgg.ls.vvung.qa.h.tlrlgaga 37  
 ClbP-like B. cereus (ZmaM) .....kiqkhiieeqisqagipgls.vviviqgne.ti.y.kknygyadi 39  
 ClbP-like B. weihenstephanensis KBAB4 .....qldslikkqmgeakipgms.vi.vvkgd.qa.vyksfgysnl 39  
 ClbP-like C. cellulolyticum 2 .....riedhihkymksgkipgls.vv.vvkdd.rdiy.kssfgysdl 39  
 ClbP-like B. cereus BL .....qidqfikeqqelsqipgls.vv.ivekg.kt.vyqkgfgyadl 39  
 ClbP-like B. mycooides .....aiqallddacrisgvpgms.is.iladd.ev.f.yfssgyadr 38  
 ClbP-like C. cellulolyticum BL .....kideliqkqmdkskiqgslvi.vsgdk.av.y.kkcygyaei 39  
 ClbP-like B. brevis .....aieqlvntmesekipgaa.vv.vvkeg.ktiy.kqsfgsdt 39  
 ClbP-like C. botulinum .....kiekfveenmdkkipgls.vt.ivkgd.kt.vyqkgfgyadi 39  
 ClbP-like P. larvae .....kmerwieeqmdkagipgls.vv.isgkd.st.lyqkgfgyagl 39  
 ClbP-like B. cereus CPT .....kiekfieenmdkkipgla.vtlvkgdk.tv.y.qkgfgyadi 39  
 ClbP-like L. sphaericus .....nielfieeqaiskipgls.vt.ivdkg.vt.vyqkgfgyani 39  
 FmtA-like S. aureus .....qidkiihdllqkghipgas.il.ivkngkvf.l.nkgygyadv 39  
 AmpC E. coli .....apqqindivhrtitplieqqkipgmavaviy....qgkpyyftgwadi 45  
 Peptidase Pab87 .....mdvgklesfivekmaerkvpgisisiik..dg..dvvyakgfgyrnv 43

**Motif 1** **Motif 7**

ClbP asqkantldtvYELGSMNSKAFtglvvqiliqegrllrqqDDIITYLPEMrl 88  
 ClbP-like H. chejuensis .....gvFELASCSKAFtgliiavleqegalsrqDAITRWIPELnq 78  
 ClbP-like B. cereus (ZmaM) .....dnkrqvtneftFELGSMNSKAFtalavqlqlekegkiklsDSLTKYVSWFem 88  
 ClbP-like B. weihenstephanensis KBAB4 .....etkqrvttektlFELGSMNSKAFtalaiyqlvqkgldidkDPVSKYLQWFqm 88  
 ClbP-like C. cellulolyticum 2 .....krkekvtentlFELGSMNSKAFtigilqlqekglkileDSINKYIPWLkl 88  
 ClbP-like B. cereus BL .....kttkppnedtlFELASCSKAFtglailqlekegllkqadDITKFLPWlql 88  
 ClbP-like B. mycooides .....ekglsasentlYELASVSKAFtgmimlleegqllsmtDPVQKYLWFTl 87  
 ClbP-like C. cellulolyticum BL .....etkkpvtedtFELGSMNSKAFtalgikledkglinldSVSKYIPWLkl 88  
 ClbP-like B. brevis .....ekqgepvtsdtlFELGSMNSKAFtalavhqlveqkrlsldDPVQNYIPWfsv 88  
 ClbP-like C. botulinum .....elqkvsdseslFELGSMNSKAFtalgildlqknglikfdDEVTKYIPWLkv 88  
 ClbP-like P. larvae .....nnkrpvtgrtlFELGSMNSKAFtglavqlqdgqiirlsDPVSAYLPWFkm 88  
 ClbP-like B. cereus CPT .....dakkpvtkslFELGSMNSKAVtalgifnleksqgikleDEVTKYIPWLkv 88  
 ClbP-like L. sphaericus .....kttkpvtsdtlFELGSMNSKAFtglailqlekgllrrsDDIITYIPWLtl 88  
 FmtA-like S. aureus .....dkkvkaspttkYELASNSKAFtglailklaqegrlnlnDDVSKHVPHFkm 88  
 AmpC E. coli .....akkqpvtqtlFELGSMNSKAFtgvllggdaiargeiklsDPTTKYWPPElt. 94  
 Peptidase Pab87 .....earlpstpetiYgigsitksftalaimklveegslsldDPVEKFNV.Ikl 91

\*\*\*

**Motif 4**

ClbP nyqg.....r.....pa...sltvaDFLYHTSG.....LP.... 111  
 ClbP-like H. chejuensis rpel.....g.....ys...airlnNLLFHSSG.....LS.... 101  
 ClbP-like B. cereus (ZmaM) eykg.....k.....py...eitlyDLLHHTSG.....IP.... 111  
 ClbP-like B. weihenstephanensis KBAB4 iyegnykqqqlnk.....nv...eitleQLLYHTSG.....IP.... 119  
 ClbP-like C. cellulolyticum 2 kykg.....e.....av...ditveQLFHHTSG.....IP.... 111  
 ClbP-like B. cereus BL tynd.....k.....pv...kinlnHLLYHTSG.....IP.... 111  
 ClbP-like B. mycooides kyqg.....k.....pvdmqsltnNLFHHTSG.....LTnirh 116  
 ClbP-like C. cellulolyticum BL nykg.....k.....sa...dltlnHFLHHTSG.....VS.... 111  
 ClbP-like B. brevis tykg.....q.....ge...tilikHLLYHTSG.....IP.... 111  
 ClbP-like C. botulinum kykg.....k.....ev...pitieQLVHHTSG.....VS.... 111  
 ClbP-like P. larvae hfkg.....e.....ehqgekidgv...ditleQLLHHTSG.....IP.... 119  
 ClbP-like B. cereus CPT kykg.....k.....ea...sitieQLLHHTSG.....IP.... 111  
 ClbP-like L. sphaericus tyng.....k.....ke...iitiqQLLSHTSG.....IP.... 111  
 FmtA-like S. aureus nyng.....q.....ne...titikQLLAQTSG.....IPs... 112  
 AmpC E. coli .....akqwngitllHLATYTAGglplqVPdevk 122  
 Peptidase Pab87 r.....pfgepvtvhhllthssg.....IPslgy 115

ClbP ..fstlarlenpmp.gsav....a...qq...lr...ne....n...l 135  
 ClbP-like H. chejuensis ..gdtldllrpsda.anal....e...qlpgllr...dt....p...l 128  
 ClbP-like B. cereus (ZmaM) ..ekaigyipitng.kdsi....e...kt...ir...nvmp...ig...l 138  
 ClbP-like B. weihenstephanensis KBAB4 ..fhtigdipisnd.ndal....e...nt...vreilnq....k...l 146  
 ClbP-like C. cellulolyticum 2 ..fesigyikpdqr.enalent...v...kt...lv...gs....e...l 138  
 ClbP-like B. cereus BL ..sttissipissa.ndal....ettikt...lk...qt....k...l 138  
 ClbP-like B. mycooides ..tgnipqgnpdm.lqkt....v...em...lv...da....e...l 141  
 ClbP-like C. cellulolyticum BL ..fltledikpsta.enal....e...qt...vr...tl...qeie...l 138  
 ClbP-like B. brevis ..fesigsipiss.enai....e...et...vr...ai....nhsel 138  
 ClbP-like C. botulinum ..fntidkipvsne.dnal...yetv...kt...li...gi....e...l 138  
 ClbP-like P. larvae ..fetikdipqgdg.ddslq.r.tv...kn...lv...nr....e...l 146  
 ClbP-like B. cereus CPT ..fktidkipvsne.dnal....e...et...vktlini....e...l 138  
 ClbP-like L. sphaericus ..snsivqipesta.knal...eltv...kt...ll...dq....t...l 138  
 FmtA-like S. aureus ..ditsedavtnknnrlnndv....t...ra...im...gd....e...l 138  
 AmpC E. coli sssdllrfyqnwqp..... 137  
 Peptidase Pab87 aeafidqmvqgdnlvpstpeetiafardmekw.....a 149

**Motif 2**

ClbP		lfapgakfsMASANYDVLG	gavienvtgktfteviaerltqplgmsatvav	185
ClbP-like H. chejuensis		aypvgareeyATLNYSLG	laaeratgksfavllrekiflplgmkdtfve	178
ClbP-like B. cereus (ZmaM)		nrepsssfeySTVNYDLG	lviekatnqsferyvqdhvlsklldllstyag	188
ClbP-like B. weihenstephanensis KBAB4		etypgeqfnVASINYDILG	lvigkvtngsfegyvqnnilnqfnmngntflf	196
ClbP-like C. cellulolyticum 2		qyppgteysYATINYDVLG	llikivtgesyekymhdnivktyglnttflf	188
ClbP-like B. cereus BL		nhkpgssfeYATINYDILG	lviekvttkpyevyiknildqigmknslv	188
ClbP-like B. mycooides		afppgeqynYGTVNYDVLG	lvieivsrqsyedfmkeqvfpplglhqtvyv	191
ClbP-like C. cellulolyticum BL		dtppgkqfgYATINYDVLG	lviekvtgksfedymqdevllpoglnttftt	188
ClbP-like B. brevis		vrkpgteyeyATINYDILG	gyivekvtgetyegyvqqviaklgtqtfl	187
ClbP-like C. botulinum		dnepgksfgYATINYDVLG	liiqevtgksyenyeenilkpmlsttyly	188
ClbP-like P. larvae		dfypgeqfgYATINYDVLG	lvieevtgsssfetyvrthvldtlgkktflf	196
ClbP-like B. cereus CPT		dseppgrdfgYATINYDVLG	lviekvtgksyegyieenlvkplghntlyl	188
ClbP-like L. sphaericus		nrkpgssfeYATINYDVLG	liieqvtklpydyvkvkqilqpigmyqsfvg	188
FmtA-like S. aureus		hhkpggeefYSNMNYDLG	liiqnvtkqsytkyitnswlklpmlhmtsfk	188
AmpC E. coli		awapgtcrlYANSSIGLFG	alavkpsglsfeqamqtrvfqplklnhtwi	186
Peptidase Pab87		vakpgerffYLNTGYVLLG	kiiiekvsgvsyeyikkileplgmrrsyff	199

♦♦♦

**Motif 3**

ClbP		kgdeiiivnk...asgyklg.f.gkpv.aah.ap.lARNHVPAAY	Ihstlp	227
ClbP-like H. chejuensis		gdpaatapr...isgykls.ftvarp.y.e.ap.rYLQNTPAGY	Ilstpe	220
ClbP-like B. cereus (ZmaM)		remaplndm...aggykrn...flqs.lpydap.nYRGNTPAGY	Fisnin	230
ClbP-like B. weihenstephanensis KBAB4		rkdvakym...skgykig.f.lkpief.n.ap.iYRGNTPAGY	Fisnne	238
ClbP-like C. cellulolyticum 2		reqaskkgv...iakgykvg...flrp.leynap.iYRGNTPAGY	Lisnin	231
ClbP-like B. cereus BL		ylgghpnem...atgyklr.fmrpyp.y.e.ap.iYRGNTPAGY	Iisntn	230
ClbP-like B. mycooides		kedaqatggl...aggyrssf.mttp.f.k.ap.dYAGNKPAGY	Iisntk	234
ClbP-like C. cellulolyticum BL		rnqavesrm...aagykvs...lmpv.k.ynap.vFLGNI	PAGYFitn	230
ClbP-like B. brevis		kaqapvegm...argnie.f.fspkv.y.d.ap.dYRGNTPAGY	Vvsnv	229
ClbP-like C. botulinum		rnenvnkhm...akgykvs.flrphe.y.d.ap.iYRGNKPAGY	Iisnge	230
ClbP-like P. larvae		rretagrdm...atgykhg...fmqs.ltynap.mYRGDTPAGY	Fitnan	238
ClbP-like B. cereus CPT		kneivneyi...akgykle.flkndv.y.d.ap.vYRGNKPAGY	Iisae	230
ClbP-like L. sphaericus		ihqvaifhm...asgykig...lmre.lpftpp.iYRGNI	PAGYIisan	230
FmtA-like S. aureus		qtnnkskhd...aigyelq.g.stpv.v.s.kpefNLWDT	PSAYMmste	230
AmpC E. coli		.nvppaeknyawgyreg.....kavhvsogaL.DAEAYG	Vkstie	224
Peptidase Pab87		keeverdk...dvamgyildkegr...lvpqpfpygITAD...GGLlssvl		241

**Motif 6**

ClbP		DMEIWDAWLh.rkalpatlrqgmsnswrgnsdv.....		260
ClbP-like H. chejuensis		DMARWLQFLQg.ppklsdsrldvlgwa.sqqakr.....		252
ClbP-like B. cereus (ZmaM)		DMERWLKLQLg.vekrpmfpeelindsheanrkv.....		263
ClbP-like B. weihenstephanensis KBAB4		DMEKWLRLMQLg.iyglstdhghkaiysthipnrsv.....		271
ClbP-like C. cellulolyticum 2		DIAGWLKIOMg.gtsnsvndkkliqeshipnrsv.....		263
ClbP-like B. cereus BL		DLAKWNNLQLg.ieinnsihtqlisdshipdktv.....		263
ClbP-like B. mycooides		DMARWMIQOMg.vgdipeifhtvikksr.gdm.....		266
ClbP-like C. cellulolyticum BL		DVQRWLSLHLg.siklpsidyklmeksheidtsvephdi.....		268
ClbP-like B. brevis		DMERWLQIQLg.le.asafrrpelvrasqlpdrsv.....		261
ClbP-like C. botulinum		DMGKWLKIOMg.tlsdlkfhknlieeshnandti.....		263
ClbP-like P. larvae		DMSKWLQIQLg.sgdginrligqshs.p.drvt.....		269
ClbP-like B. cereus CPT		DMAKWLKIQLg.tlndlkfdkdivgksqeanrtv.....		263
ClbP-like L. sphaericus		DIAKWLKLQLg.nseiysidattikeshnpdqtv.....		263
FmtA-like S. aureus		DLEHWLKFQLn.ppdkykslvq.q.shknlst.....		260
AmpC E. coli		DMARWVQSNLkpldinektlqqgiqlagsrywqtgdmgyqglgwemldwpv		274
Peptidase Pab87		DLAKYLKMYIerdesivs..keyiekmetisyik.....v		273

		<b>Motif 8</b>	<b>Motif 5</b>	
ClbP		pla....adn....riLYASGW..FIDQn..qgpyisHGGONPNFSSci		297
ClbP-like H. chejuensis		pyg....gds....daGYAYGW..DVEAg..attswsHPGNPNAGAYv		289
ClbP-like B. cereus (ZmaM)		apd....n.n....gaSYAAGW..EVYQs..gggeisHGGSNPNYSSyi		299
ClbP-like B. weihenstephanensis KBAB4		pps....edg....sSYAGGW..QVFQn..gpgeisHAGSNPNFSSfv		307
ClbP-like C. cellulolyticum 2		rpy....vdg....sSYAAGW..SVFQs..gggelsHGGSNPNYSSfi		299
ClbP-like B. cereus BL		qpi....knn....dYYASGW..IVRQe..dtksilHGGSNPNFSSfi		299
ClbP-like B. mycooides		svs....av.n....gmYYAAGW..SVNAe...qtieHTGNPNFRTev		302
ClbP-like C. cellulolyticum BL		gvl....ydk....geRYGAGW..ERAEG..rngf.fHGGANPNYSSym		304
ClbP-like B. brevis		apa....adg....fSYASGW..MIYQd..gggqisHGGNPNFSSym		297
ClbP-like C. botulinum		apl....rdg....sSYASGW..LVYEK..gsreisHAGNPNYSSfi		299
ClbP-like P. larvae		ppa....edg....sSYAAGW..SVYQl..gsgmlsHSGSNPNYSSql		305
ClbP-like B. cereus CPT		rpl....sdg....sFYASGW..FIYPk..senqisHAGNPNYSSfi		299
ClbP-like L. sphaericus		epf....d.t....ntYYATGWA..VVEKe..nkpyiyHAGNPTFTSyf		300
FmtA-like S. aureus		ig....ep.n....anAYASGW..FTN.n..dehlvfHSGTLDNFSSfi		295
AmpC E. coli		npdsiingsdnkialaarPVKAITPPTPAV..raswvhKTCATGGFGSv		322
Peptidase Pab87		p.....weifgGEGYGYGLIYYPnflgeklvgHSSVSGMYTgyi		312

♦♦♦

```

ClbP          alrpdqgigivalan.mns...n...l...ilql...cadidnyl.... 330
ClbP-like H. chejuensis   afdpqgtvgvvalgn.sns...p...q...viql...gravfehl.... 322
ClbP-like B. cereus (ZmaM) gfrdeeqigvvlan.lns...n...y...tyyi...gssllnm...id 333
ClbP-like B. weihenstephanensis KBAB4 vfhpqeklgvavman.mns...d...y...tqni...gqaimdtl.vge. 343
ClbP-like C. cellulolyticum 2   vfrpeeklgvavlan.lns...s...y...vqtl...gggiidiild.r. 335
ClbP-like B. cereus BL       imqpnkqgtgvavlsn.mks...a...vttsiggg...vidlwngk.... 335
ClbP-like B. mycooides       aillnertavslsngant...n...i...nmvlk...vkdilegn.... 337
ClbP-like C. cellulolyticum BL firpdqgmigimlcn.fgs...d...t...iqilgknlvdiidgn.... 340
ClbP-like B. brevis          vfrpeeklgavvltl.lnslnas...v...iaeg...avnlvlgk.... 333
ClbP-like C. botulinum      lfrpedkvgvailsn.tss...q...y...ipai...tggineil.q.g. 334
ClbP-like P. larvae          vllpgeeigiavlan.lns...d...y...tevi...gngiaail.... 338
ClbP-like B. cereus CPT      vfnpkdkvgvavlsn.tns...q...y...veli...ggginkil.... 332
ClbP-like L. sphaericus     imqpdeqigvailan.mnt...sstta...iggg...imdlwegk.... 336
FmtA-like S. aureus         llnpkqnygivvvan.lns...e...y...vpkl...vehlntqi.... 328
AmpC E. coli                afipekelgivmlanknyp.....nparvdaawqilnalq.... 358
Peptidase Pab87             gyipekkigvavlen.ssg.....yppsyiamyalalllgkn.... 349

```

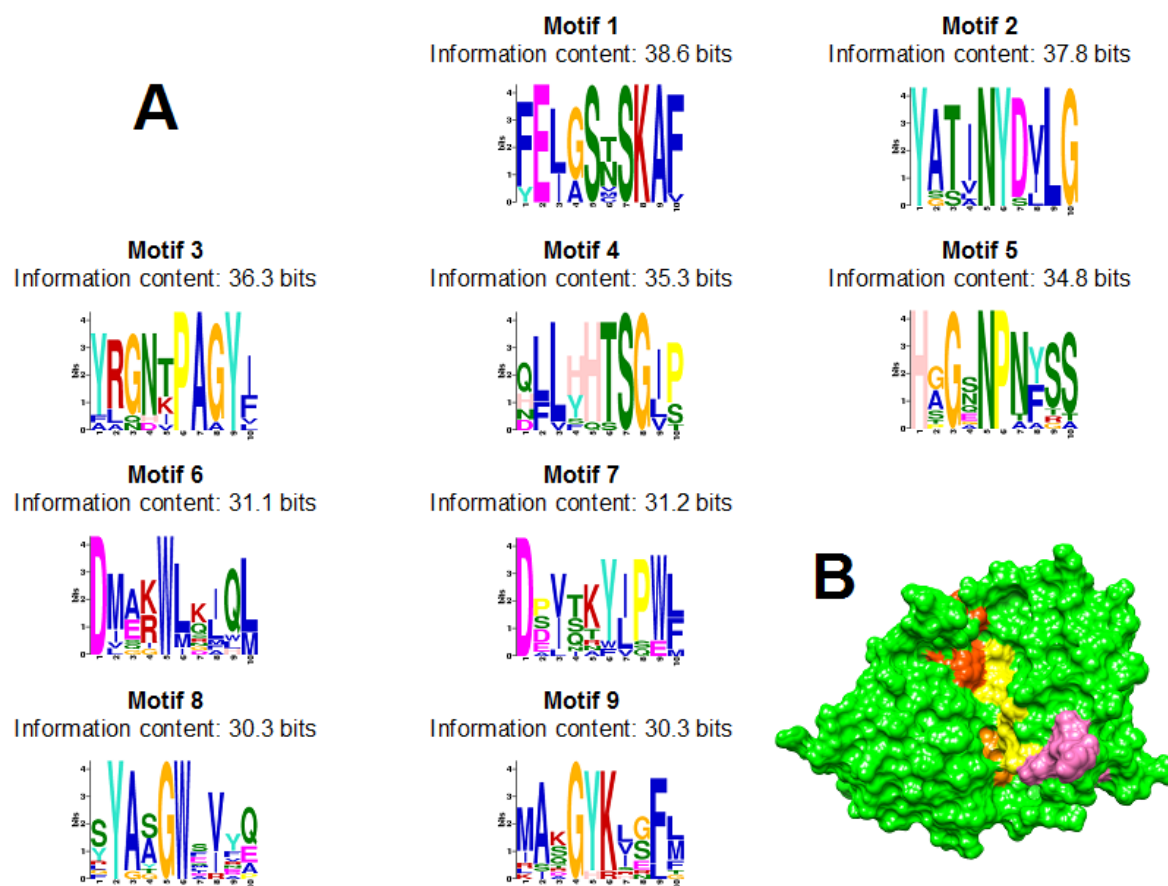
```

ClbP          .r.igk.....y... 335
ClbP-like H. chejuensis   .r.gas.p..qpl... 330
ClbP-like B. cereus (ZmaM) kevk.p.....l... 340
ClbP-like B. weihenstephanensis KBAB4 .s.vvt.....n... 348
ClbP-like C. cellulolyticum 2   .e.taa.....p... 340
ClbP-like B. cereus BL       .n.isd.....t... 340
ClbP-like B. mycooides       .l.tqs.....y... 342
ClbP-like C. cellulolyticum BL .g.eps.....l... 345
ClbP-like B. brevis          .e.apv.....v... 338
ClbP-like C. botulinum      .rvynk.....d... 340
ClbP-like P. larvae          .q.gka..pep.l... 346
ClbP-like B. cereus CPT      .e.egt.....ynkn 340
ClbP-like L. sphaericus     .t.ttp.....l... 341
FmtA-like S. aureus         .v.nhkry..s.t... 336
AmpC E. coli                .....
Peptidase Pab87             .....

```

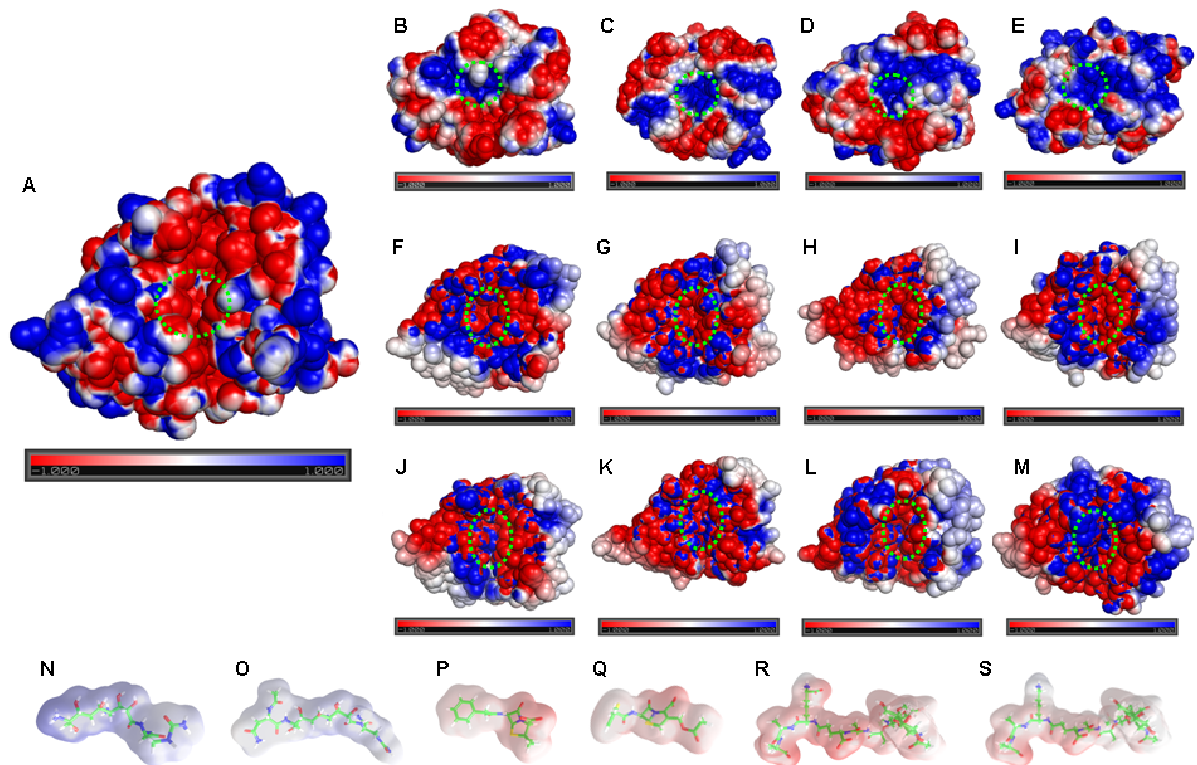
**Supplemental Figure 9.** Sequence alignment of NRP-associated peptidases with FmtA-like protein, AmpC  $\beta$ -lactamase and Pab87 peptidase.

The sequences corresponding to motifs are in upper case. Residues emphasized by black shading are 100% conserved in NRP-associated peptidases. PRP specific motifs are indicated by black diamonds.

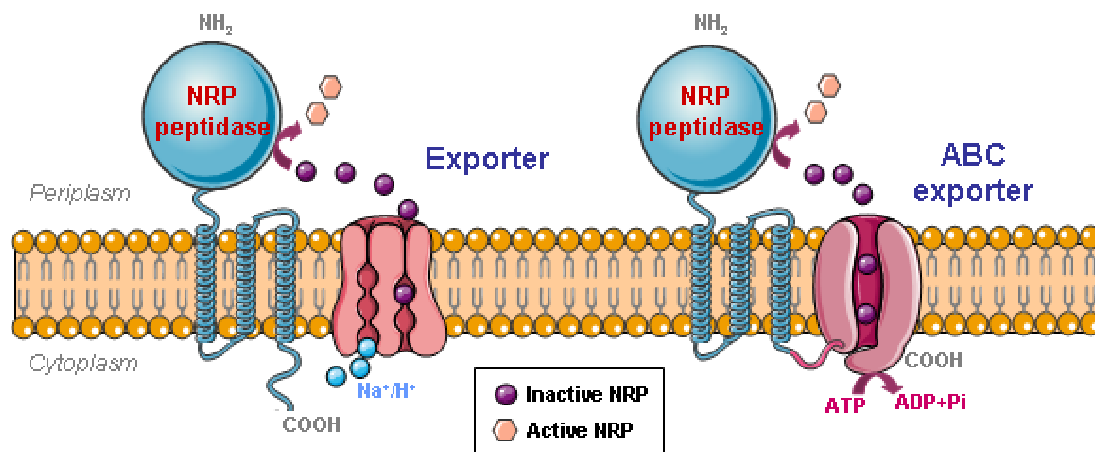


Supplemental Figure 10. The motifs obtained with MEME program from NRP-associated peptidase sequences.

(A) The motifs are numbered according to information content and are represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif. These are displayed as ‘sequence LOGOS’, containing stacks of letters at each position in the motif. The total height of the stack is the information content of that position in the motif in bits. The height of the individual letters in a stack is the probability of the letter at that position multiplied by the total information content of the stack. (B) Location of NRP-associated peptidase specific motifs predicted by MEME program on ClbPpep surface. Motif 2 is colored in orange, motif 5 in yellow, motif 8 in orange-red, and motif 9 in pink.



**Supplemental Figure 11.** Electrostatic potential surface map of MEROPS S12 peptidases, and their putative and known ligands including the crystal structures of ClbPpep (A), Pab87 peptidase (B) (23), aminopeptidase (C) (24), R61 carboxypeptidase (D) (19), and AmpC  $\beta$ -lactamase of *E. coli* (E), the homology models of NRP-associated peptidases of *H. chejuensis* (F), *B. cereus* (ZmaM) (G), *B. weihenstephanensis* (H), *C. cellulolyticum* (I), *B. pseudomycooides* (J), *B. mycooides* (K), *Lysinibacillus sphaericus* C3-41 (L), and of FmtA-like protein of *S. aureus* (M), and the NRP-associated peptidase putative ligands (N, O) and known ligands of the MEROPS S12 enzymes AmpC and carboxypeptidases (P-S): (N) zwittermicin A, (O) pre-zwittermicin, (P) benzylpenicillin, (Q) cephalothin, (R) NAG-NAM-pentapeptides of Gram negative bacteria, and (S) NAG-NAM-pentapeptides of Gram positive bacteria. The catalytic pockets are indicated by green circles. Electrostatic surface potential in blue is positive, negative in red and neutral in white.



**Supplemental Figure 12.** Model of secretion and maturation of NRP compounds involving NRP-associated peptidases.

As a two-component resistance mechanism, after translocation of an immature inactive NRP compound from cytoplasm to periplasm *via* efflux pumps such as  $\text{Na}^+/\text{H}^+$  energy-driven MATE exporters or ATP-driven ABC exporters, the periplasmic location of the NRP-associated peptidase main domain may allow the maturation of the fully active NRP toxin, away from the potential cytoplasmic target of its producer cell. ABC exporters may be linked covalently to NRP peptidase as with *B. cereus* ZmaM. (Illustration made with the assistance of Servier Medical Art.)



## Supplemental Tables

**Supplemental Table 1.** Data collection and structure determination statistics for ClbPpep crystal structure.

<b>Data collection</b>	
Space group	C2
Cell dimensions	
a, b, c (Å)	103.9, 149.9, 87.3
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 123.9, 90.0
Resolution (Å)	20 - 2.4 (2.5 - 2.4) <sup>a</sup>
R <sub>merge</sub> (%) <sup>b</sup>	8.6 (28.0) <sup>a</sup>
[I / $\sigma$ (I)] (°)	11.5 (3.7) <sup>a</sup>
Completeness (%)	90.9 (94.3) <sup>a</sup>
Redundancy	3 (3) <sup>a</sup>
<b>Refinement</b>	
No. reflections	39,788
R <sub>work</sub> / R <sub>free</sub>	19.85/ 22.47
No. atoms	
Protein	7719
Water	249
B-factors	
Protein	22.6
Water	19.2
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angle (°)	0.919
Ramachandran plot:	
Favored region (%)	97.2
Allowed region (%)	2.8

<sup>a</sup> Values in parentheses represent the highest resolution shells.

**Supplemental Table 2.** The 20 best non-redundant neighbours of ClbPpep structure recovered from DALI server.

<b>PDP ID<sup>a</sup></b>	<b>Z-score<sup>b</sup></b>	<b>RMSD<sup>c</sup> (Å)</b>	<b>Aligned residues<sup>d</sup></b>	<b>Sequence identity (%)</b>	<b>Description</b>
2r9w	37.6	2.1	311	25	AmpC beta-lactamase
2qmi	37.2	2.6	324	24	Pab87 peptidase
2zc7	37.1	2.1	310	25	AmpC beta-lactamase
1fr1	36.9	2.2	311	24	AmpC beta-lactamase
2qz6	36.7	2.2	306	23	AmpC beta-lactamase
1zkj	36.1	2.4	310	21	AmpC beta-lactamase
1ga0	35.3	2.4	311	24	AmpC beta-lactamase
1ei5	34.5	2.5	310	22	D-Amino peptidase
1ci8	34.0	2.6	315	21	Carboxylesterase
1cef	33.6	2.6	306	22	D-Ala-D-Ala Carboxypeptidase
2efu	33.0	2.8	306	18	D-Amino Acid Amidase
2dns	33.0	2.7	309	18	D-Amino Acid Amidase
1wyc	26.7	3.0	283	21	6-Aminohexanoate-dimer hydrolase
3i7j	25.0	2.7	248	20	AmpC beta-lactamase
1e25	16.7	3.3	217	11	PER-1 beta-lactamase
1hzo	15.9	3.3	212	8	PVUL beta-lactamase
1iyq	15.8	2.9	206	11	TOHO-1 beta-lactamase
1ylj	15.8	3.1	207	12	CTX-M-9 beta-lactamase
3byd	15.7	3.0	205	12	OXY-1 beta-lactamase
3bfd	15.6	3.0	204	11	SED-1 beta-lactamase

<sup>a</sup>, PDP identity of the subject structure in RCSB protein data bank; <sup>b</sup>, quality of alignment; <sup>c</sup>, root-mean square deviation of C $\alpha$  atoms; <sup>d</sup>, Number of aligned residues.

Supplemental Table 3. Strains used in this study.

Strain or Plasmid	Genotype/Description	Source
<i>E. coli</i>		
DH5 $\alpha$ F'	F'/endA1 hsdR17 (r <sub>K</sub> -m <sub>K</sub> <sup>+</sup> ) glnV44 thi-1recA1 gyrA (Nal <sup>r</sup> ) relA1 $\Delta$ (lacIZYAargF)U169 deoR ( $\phi$ 80dlac $\Delta$ (lacZ)M15)	Invitrogen, France
BL21(DE3)	F <sup>-</sup> ompT gal [dcm] [lon] hsdSB (r <sub>B</sub> - m <sub>B</sub> -; an E. coli B strain) with DE3, a $\lambda$ prophage carrying the T7 RNA polymerase gene	Novagen, Darmstadt, Germany
TG1	$\Delta$ (lac-pro) supE thi hsdD5/F' traD36 proA+B+ lacIq lacZ $\Delta$ M15	Amersham Pharmacia biotech, France
IHE3034	Extra-intestinal pathogenic E. coli O18:K1:H7	(1)
DH10B pVector	DH10B containing the BAC vector pBeloBAC11	1
DH10B pBACpks	DH10B containing pBACpks	(1)
DH10B pBACpks $\Delta$ clbP	DH10B containing pBACpks $\Delta$ clbP	(1)
MC4100A	F' lac $\Delta$ (lacZYA-argF)U169 araD139 rpsL150 thi flbB5301 deoC7 ptsF25 relA1 rbsR ara <sup>+</sup>	(5)
HPT406A	$\Delta$ SRP isogenic mutant of MC4100A	(25)
CU164A	secY39-cs, zhd-33::Tn10, ara $\square$ <sup>+</sup>	(5)
MC1060	F- DE(codB-lacI)3 galK16 galE15 LAM- e14- mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2	(3)
MYC-cs	cold-sensitive yidC derivative of MC1060	(3)
MC1000	araD139 (ara, leu)7697 lacX74 galU galK atrA	(26)
YLS-3	$\Delta$ mreB isogenic mutant of MC1000	(26)
YLS-24	$\Delta$ ftsZ isogenic mutant of MC1000	(21)
<i>Bacillus weihenstephanensis</i> KBAB4	bearing clbP-like gene (locus tag "BcerKBAB4_1783")	(27)
<i>Bacillus cereus</i> UW85	bearing clbP-like gene (locus tag "zmaM")	(17)
<i>Hahella chejuensis</i> KCTC 2396 <sup>T</sup>	bearing clbP-like gene (locus tag "HCH_03446")	(28)
<i>Bacillus mycoides</i> DSM 2048	bearing clbP-like gene (locus tag "bmyco0001_51990")	DSMZ
<i>Bacillus pseudomycooides</i> DSM 12442	Bearing clbP-like gene (locus tag "bpmx0001_2150")	DSMZ
<i>Clostridium cellulolyticum</i> H10	bearing clbP-like genes (locus tag "Ccel_0973" and "Ccel_2389")	DSMZ
<i>Staphylococcus aureus</i> Mu50	bearing fmtA-like gene (locus tag "SAV2441")	CIP

Supplemental Table 4. Plasmids used in this study.

Plasmids	Description	Source
pBeloBAC11 (or pVector)	Cm <sup>R</sup>	New England BioLabs
pBAC <i>pks</i>	pBeloBAC11 harboring the <i>pks</i> island	(1)
pBAC <i>pks</i> Δ <i>clbP</i>	pBAC <i>pks</i> island mutated for <i>clbP</i>	(1)
p8761	pBAD24 encoding TorA twin-arginine signal peptide + C-terminal in-frame GFPmut2, <i>Para</i>	(4)
p61ClbPssGFP	pBAD24 encoding ClbP signal peptide plus 6 amino acid (1-36) + C-terminal in-frame GFPmut2, <i>Para</i>	This study
pET9a	T7 promoter, Kan <sup>R</sup>	Novagen
pClbPpep	pET9a encoding ClbP peptidase domain (ClbPpep : amino acid 35 to 375)	This study
pClbPpepY186E	pET9a encoding the mutant Y186E of ClbPpep	This study
pMB702	pCRScript derivative encoding <i>clbP</i> plus 570 bp upstream	(1)
pBRSK	pBR328 derivative containing P <sub>lac</sub> and P <sub>T7</sub> promoters, and the multiple cloning site from the Bluescript SK vector (Stratagene)	(2)
pOB901	pBRSK encoding ClbP	This study
pOB902	pOB901 derivative encoding the mutant S95A of ClbP	This study
pOB903	pOB901 derivative encoding the mutant K98T of ClbP	This study
pOB904	pOB901 derivative encoding the mutant Y186G of ClbP	This study
pASK-IBA33plus	Tetracycline promoter, C-terminal in-frame <i>RGS6His</i> tag, Amp <sup>R</sup>	IBA GmbH, Germany
p33ClbP	pASK-IBA33plus encoding <i>clbP</i> - <i>RGS6His</i>	This study
pS95A	p33ClbP derivative encoding the mutant S95A of ClbP	This study
pE159K	p33ClbP derivative encoding the mutant E159K of ClbP	This study
pY186E	p33ClbP derivative encoding the mutant Y186E of ClbP	This study
pS188A	p33ClbP derivative encoding the mutant S188A of ClbP	This study
pH257A	p33ClbP derivative encoding the mutant H257A of ClbP	This study
pF316R	p33ClbP derivative encoding the mutant G328S F316R of ClbP	This study
pG328S	p33ClbP derivative encoding the mutant G328S of ClbP	This study
pN331A	p33ClbP derivative encoding the mutant N331A of ClbP	This study
pC337A	p33ClbP derivative encoding the mutant C337A of ClbP	This study
p33BweihenstephanensisT	pASK-IBA33plus encoding ClbP-like-encoding sequence of the locus tag <i>BcerKBAB4_1783</i>	This study
p33ZmaMT	pASK-IBA33plus encoding ClbP-like-encoding sequence of the locus tag <i>zmaM</i>	This study
p33HchejuensisT	pASK-IBA33plus encoding ClbP-like-encoding sequence of the locus tag <i>HCH_03446</i>	This study
p33Ccellulolyticum2T	pASK-IBA33plus encoding ClbP-like-encoding sequence of the locus tag <i>Ccel_2389</i>	This study
p33Bmycoides1	pASK-IBA33plus encoding ClbP-like-encoding sequence of the locus tag <i>bmyco0001_51990</i>	This study
p33Bpseudomycoides	pASK-IBA33plus encoding ClbP-like-encoding sequence of the locus tag <i>bpmx0001_2150</i>	This study
p33FmtA-like	pASK-IBA33plus encoding ClbP-like-encoding sequence of the locus tag <i>SAV2441</i>	This study

**Supplemental Table 5.** Primers used in this study for clonings (with restriction sites bolded and underlined).

<b>Primer name</b>	<b>Sequence</b>	<b>Target</b>
clbPMK-f	AAC <b><u>CAATTG</u></b> GTACCATGACAATAATGGAACACGTTAGC	ClbP-encoding sequence
clbPBN-r	AGC <b><u>GGATCC</u></b> GCTAGCCTCATCGTCCACTCCTTGTG	ClbP-encoding sequence
clbPssBN-r	AGC <b><u>GGATCC</u></b> GCTAGCGATAGGCTCATGTTCTTGCGCC	ClbP signal sequence- encoding sequence
clbP-NdeI2-f	GGAATTCC <b><u>CATATG</u></b> CCTATCGGGCGCAAGATG	ClbPpep-encoding sequence
clbP-NotI2-r	AGCATAGTTTAG <b><u>GCGGCCGC</u></b> CTTAATATTTGCCAATGCGCAGA	ClbPpep-encoding sequence
KBAB4_1783-MfeI-f	AAC <b><u>CAATTG</u></b> GTACCATGATGAGAGTAAAAATTGTTTTG	locus tag <i>BcerKBAB4_1783</i>
KBAB4_1783-XhoI-r	AGC <b><u>CTCGAG</u></b> GCTAGCATTCTTTCTTATTATGCAAAGG	locus tag <i>BcerKBAB4_1783</i>
ZmaM-EcoRI-f	AAC <b><u>GAATTC</u></b> GTACCATGAAGTTAAACATATGGTTGAAG	locus tag <i>zmaM</i>
ZmaM-BamHI-r	AGC <b><u>GGATCC</u></b> GCTAGCAAGCGACTTTCCCTTTTGTCTTGG	locus tag <i>zmaM</i>
HCH_3446-EcoRI-f	AAC <b><u>GAATTC</u></b> GTACCATGAATAATGTGGCTTAATCC	locus tag <i>HCH_03446</i>
HCH_3446-BamHI-r	AGC <b><u>GGATCC</u></b> GCTAGCCAGCGACGAAAAGGTTTGCTG	locus tag <i>HCH_03446</i>
Ccel_2389-MfeI-f	AAC <b><u>CAATTG</u></b> GTACCATGAAATTAACAAAAATGG	locus tag <i>Ccel_2389</i>
Ccel_2389-BamHI-r	AGC <b><u>GGATCC</u></b> GCTAGCAAAGAGAAAGTTCAAATT	locus tag <i>Ccel_2389</i>
Bmy_51990-EcoRI-f	AAC <b><u>GAATTC</u></b> GTACCATGAAAAAGATTATTG	locus tag <i>bmyco0001_51990</i>
Bmy_51990-BamHI-r	AGC <b><u>GGATCC</u></b> GCTAGCTCGGTGAGTGATATACAAACC	locus tag <i>bmyco0001_51990</i>
Bpmy_2150-EcoRI-f	AAC <b><u>GAATTC</u></b> GTACCATGAATAAAAAGCAAAGAAAAC	locus tag <i>bpmyx0001_2150</i>
Bpmy_2150-BamHI-r	AGC <b><u>GGATCC</u></b> GCTAGCAGAAATAAATCTATTCATTTC	locus tag <i>bpmyx0001_2150</i>
SAV_2441-EcoRI-f	AAC <b><u>GAATTC</u></b> GTACCATGACTACTAAAAACTG	locus tag <i>SAV2441</i>
SAV_2441-BamHI-r	AGC <b><u>GGATCC</u></b> GCTAGCTGTCTTCGTTATCTTTG	locus tag <i>SAV2441</i>

**Supplemental Table 6.** Primers used in this study for site-directed mutagenesis of *clbP* gene (with mutated bases bolded and underlined).

Mutation	Primer name	Sequence
S95A	IHAPJPN105	gatgagtaaggcgtttaccgg
	IHAPJPN106	<u>g</u> ctcccagctcgtaaactgtg
K98T	IHAPJPN107	<u>t</u> gcgtttaccggacttgtggg
	IHAPJPN108	<u>g</u> tactcatcgatcccagctcgta
E159K	clbP-E159Kf	caacactggctcggctg <u>a</u> aaaaccctatgcctgg
	clbP-E159Kr	ccagggcatagggtttt <u>t</u> cagccgagccagtgtg
Y186G	IHAPOB3 Y186G	<u>c</u> gcctccgccaattatgatgtgtgggcgcggtgattgaaaatgtgacgggaaaaacctttac
	IHAPOB4	<u>c</u> cgctaaacttagcaccggagcaaacagcagattctcgttgcgcagtgtg
Y186E	clbP-Y186Ef	gaagtttagc <u>g</u> aagcctccgccaattatgatgtgtgggc
	clbP-Y186Er	cataattggcggaggc <u>t</u> tcgctaaacttcgcaccggcgc
S188A	clbP-S188Af	gtttagctatgcc <u>g</u> ccgccaattatgatgtgtgggcgcg
	clbP-S188Ar	catcataattggcggg <u>g</u> ccatagctaaacttcgcaccggcgc
H257A	clbP-H257Af	ggcccgaac <u>g</u> ctgttctcgcgcctatatccatagcactctgcc
	clbP-H257Ar	ggcggcaggaacag <u>g</u> cttccgggcccagaggcgcatgaaacagaacgg
F316R/G328S	clbP-G328Ss	cccttacatcagtcac <u>a</u> gtgggcagaatccaaacttttcttcttgattgcg
	clbP-G328SF316Rr	cccact <u>t</u> gtgactgatgtaagggccctgattctggctgat <u>c</u> ctccaaccgc
G328S	clbP-G328Sf	gcccttacatcagtcac <u>a</u> gtgggcagaatccaaac
	clbP-G328Sr	gtttggattctgccact <u>t</u> gtgactgatgtaagggc
N331A	clbP-N331Af	cacggtgggcag <u>g</u> ctccaaacttttcttcttgattgcggttgcgaccg
	clbP-N331Ar	gaagaaaagtttggag <u>c</u> ctgcccaccgtgactgatgtaagggccttgattctgg
C337A	clbP-C337Af	gcagaatccaaacttttcttct <u>g</u> ccattgcggtgacaccggatcagc
	clbP-C337Ar	gctgatccggtcgcaacgcaat <u>g</u> gcagaagaaaagtttgattctg

## Supplemental References

1. Nougayrede, J. P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006) *Science* **313**(5788), 848-851
2. Schmitt, C. K., Darnell, S. C., Tesh, V. L., Stocker, B. A., and O'Brien, A. D. (1994) *J Bacteriol* **176**(2), 368-377
3. Chen, M., Xie, K., Jiang, F., Yi, L., and Dalbey, R. E. (2002) *Biol Chem* **383**(10), 1565-1572
4. Pradel, N., Delmas, J., Wu, L. F., Santini, C. L., and Bonnet, R. (2009) *Antimicrob Agents Chemother* **53**(1), 242-248
5. Santini, C. L., Ize, B., Chanal, A., Muller, M., Giordano, G., and Wu, L. F. (1998) *Embo J* **17**(1), 101-112
6. Bailey, T. L., and Elkan, C. (1994) *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36
7. Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006) *Nucleic Acids Res* **34**(Web Server issue), W369-373
8. Bailey, T. L., and Gribskov, M. (1998) *Bioinformatics* **14**(1), 48-54
9. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) *Curr Protoc Bioinformatics* **Chapter 5**, Unit 5 6
10. Sali, A., and Blundell, T. L. (1993) *J Mol Biol* **234**(3), 779-815
11. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) *Nucleic Acids Res* **32**(Web Server issue), W665-667
12. Hinsén, K. (2000) *J Comp Chem* **21**, 79-85
13. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J Comput Chem* **25**(13), 1605-1612
14. Case, D. A., Darden, T. A., Cheatham, I., T.E. , Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. A. (2008) AMBER 10. In., University of California, San Francisco, CA, USA
15. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J App Cryst* **26**, 283-291
16. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc Natl Acad Sci U S A* **98**(18), 10037-10041
17. Kevany, B. M., Rasko, D. A., and Thomas, M. G. (2009) *Appl Environ Microbiol* **75**(4), 1144-1155
18. Gasteiger, J., Rudolph, C., and Sadowski, J. (1990) *Tetrahedron Comp Methodology* **3**, 537-547
19. Beadle, B. M., Trehan, I., Focia, P. J., and Shoichet, B. K. (2002) *Structure* **10**(3), 413-424
20. Delfosse, V., Girard, E., Birck, C., Delmarcelle, M., Delarue, M., Poch, O., Schultz, P., and Mayer, C. (2009) *PLoS One* **4**(3), e4712
21. Yuan, J., Zweers, J. C., van Dijk, J. M., and Dalbey, R. E. (2010) *Cell Mol Life Sci* **67**(2), 179-199
22. Thomas, J. D., Daniel, R. A., Errington, J., and Robinson, C. (2001) *Mol Microbiol* **39**(1), 47-53
23. Bompard-Gilles, C., Remaut, H., Villeret, V., Prange, T., Fanuel, L., Delmarcelle, M., Joris, B., Frere, J., and Van Beeumen, J. (2000) *Structure* **8**(9), 971-980
24. Lee, W., McDonough, M. A., Kotra, L., Li, Z. H., Silvaggi, N. R., Takeda, Y., Kelly, J. A., and Mobashery, S. (2001) *Proc Natl Acad Sci U S A* **98**(4), 1427-1431
25. Tian, H., and Beckwith, J. (2002) *J Bacteriol* **184**(1), 111-118
26. Shih, Y. L., Kawagishi, I., and Rothfield, L. (2005) *Mol Microbiol* **58**(4), 917-928
27. Lapidus, A., Goltsman, E., Auger, S., Galleron, N., Segurens, B., Dossat, C., Land, M. L., Broussolle, V., Brillard, J., Guinebretiere, M. H., Sanchis, V., Nguen-The, C., Lereclus, D., Richardson, P., Wincker, P., Weissenbach, J., Ehrlich, S. D., and Sorokin, A. (2008) *Chem Biol Interact* **171**(2), 236-249

28. Jeong, H., Yim, J. H., Lee, C., Choi, S. H., Park, Y. K., Yoon, S. H., Hur, C. G., Kang, H. Y., Kim, D., Lee, H. H., Park, K. H., Park, S. H., Park, H. S., Lee, H. K., Oh, T. K., and Kim, J. F. (2005) *Nucleic Acids Res* **33**(22), 7066-7073





## II. Article 2 – Les cyclomodulines des *E. coli* d'urosepsis

### Cyclomodulins in Urosepsis Strains of *Escherichia coli*.

Damien Dubois, Julien Delmas, Anne Cady, Frédéric Robin, Adeline Sivignon, Eric Oswald et Richard Bonnet.

*Journal of Clinical Microbiology* (2010), Vol. **48** (6), p2122-9.

L'îlot *pks* est quasi exclusivement restreint aux souches de *E. coli* de groupe phylogénétique B2. Les ExPEC et en particulier les UPEC appartiennent majoritairement au phylogroupe B2 et dans une moindre mesure au phylogroupe D [21]. En raison de la découverte récente de cet îlot génomique *pks*, aucune donnée épidémiologique n'avait été rapportée dans un contexte d'infection urinaire, *E. coli* restant le premier micro-organisme responsable d'infections urinaires nosocomiales comme communautaires.

Nous avons déterminé la prévalence des gènes codant pour les quatre types de cyclomodulines (CM) retrouvés chez *E. coli*, à savoir l'îlot *pks* codant la colibactine, les différents types de *cnf* et de *cdtB* et le gène *cif*, dans une série de 146 souches isolées d'hémocultures de patients ayant présenté un épisode d'urosepsis communautaire et admis aux CHU de Clermont-Ferrand ou de Rennes, comparée à une série de 51 souches fécales commensales. Ces dernières étaient issues d'individus sains communautaires de la région clermontoise n'ayant présenté aucun signe d'infection aiguë, ou de troubles bénins ou malins du tractus gastro-intestinal et n'ayant reçu aucun antibiotique dans le mois précédent le prélèvement.

Le groupe phylogénétique de chaque souche de *E. coli* a été déterminé par PCR multiplex. Le phylogroupe B2 était prédominant dans les souches isolées d'urosepsis (54%), suivi par les phylogroupes D (23%), A (18%) et B1 (5%), comme précédemment observé dans plusieurs études. Les isolats commensaux appartenaient le plus fréquemment au phylogroupe A (37%), suivi des phylogroupes B1 et B2 (24 % pour les deux groupes), résultats concordants avec certaines études précédentes mais différant d'autres études françaises et d'autres pays industrialisés où le phylogroupe B2 prédominait dans les isolats commensaux. Ainsi le phylogroupe A, et dans une moindre mesure le phylogroupe B1 étaient

significativement plus prévalents parmi les souches commensales que parmi les souches d'origine urinaire (37% versus 18%,  $p=0,011$  ; et 24% versus 5%  $p<0,001$ ). En revanche, le phylogroupe B2 était significativement prédominant dans les souches urinaires (54% versus 24%,  $p<0,001$ ).

Les gènes codant les différentes CM ont été dépistés par hybridation d'ADN dot-blot à l'aide de quatre sondes ou mélange de sondes détectant les gènes médians *clbJ-K* de l'îlot *pks*, le gène *cif* et les différents types de gènes de *cnf* et de *cdtB*. Les gènes détectés par hybridation ont été confirmés par PCR en ciblant les gènes des extrémités de l'îlot *pks*, *clbA* et *clbQ*, le gène *cif*, et en typant les gènes *cnf* (par PCR multiplex mise au point au cours de cette étude) et les gènes *cdtB*. La distribution des gènes des CM a été analysée en fonction des phylogroupes. Les gènes codant les différentes CM les plus fréquents dans la cohorte étaient l'îlot *pks* (27%) et le *cnf1* (18%), qui restaient fortement liés au phylogroupe B2. Ces deux facteurs de virulence se retrouvaient fréquemment associés dans le même isolat. Les gènes *cdtB* (6%) et *cif* (1%) étaient faiblement représentés et ne présentaient pas de distribution phylogénétique particulièrement significative. Les gènes *cnf2* et *cnf3* étaient absents.

L'îlot *pks*, le gène *cnf1* et l'association de ces deux facteurs de virulence étaient prédictifs d'une origine uroseptique de l'isolat ( $p=0,005$ ,  $p<0,001$  et  $p=0,001$  respectivement). Cependant, la prévalence de l'îlot *pks* parmi les isolats du phylogroupe B2 commensaux et uroseptiques n'était significativement pas différente, contrairement à celle du gène *cnf1*, suggérant que la colibactine est importante pour le processus de colonisation des épithélia et CNF1 pour la virulence. En outre, l'absence de gènes codant les différentes CM associée aux phylogroupes B1 ou A était significativement prédictive d'une origine commensale ( $p=0,003$  et  $p=0,018$  respectivement). Les gènes *cdtB* et *cif* ne présentaient pas de distribution clinique particulière.

L'expression phénotypique des CM a été investiguée sur culture cellulaire. Les isolats possédant les gènes *cnf1* et *cdtB* produisaient une CM fonctionnelle. Un seul des deux isolats *cif*-positifs était cytopathogène. En raison du caractère hémolytique de nombreuses souches *pks*-positives, seulement 40% de ces souches étaient analysables. Parmi celles-ci, 14% soit 3 souches sur 21 ont présenté un îlot *pks* non fonctionnel ; deux d'entre elles provenaient des fécès, soit 33% des souches *pks*-positives isolées de fécès. Ce défaut de fonctionnalité de l'îlot *pks* peut induire une surestimation de la prévalence du facteur de virulence « colibactine » lors des analyses génétiques.

Enfin, nous avons analysé la prévalence des résistances aux antibiotiques habituellement utilisés en pratique clinique parmi les isolats d'origine urinaire, la résistance étant décrite par plusieurs auteurs comme opposée à la virulence. Les isolats des phylogroupes A étaient significativement plus résistants aux fluoroquinolones, à l'amoxicilline et aux céphalosporines de 3<sup>ème</sup> génération. Les isolats des phylogroupes B2 étaient significativement plus sensibles à ces mêmes antibiotiques et au cotrimoxazole. De même, les isolats *pks*-positifs et *cnf1*-positifs étaient significativement plus sensibles aux fluoroquinolones, à l'amoxicilline et au cotrimoxazole, et à l'amoxicilline et au cotrimoxazole, respectivement.



## Cyclomodulins in Urosepsis Strains of *Escherichia coli*<sup>∇</sup>

Damien Dubois,<sup>1,2</sup> Julien Delmas,<sup>1,2</sup> Anne Cady,<sup>3</sup> Frédéric Robin,<sup>1,2</sup> Adeline Sivignon,<sup>2</sup>  
Eric Oswald,<sup>4,5,6,7</sup> and Richard Bonnet<sup>1,2\*</sup>

CHU Clermont-Ferrand, Centre de Biologie, Laboratoire de Bactériologie, Clermont-Ferrand F-63003,<sup>1</sup> Clermont Université, Université d'Auvergne, Evolution des bactéries pathogènes et susceptibilité de l'hôte, JE 2526 usc INRA 2018, Clermont-Ferrand F-63001,<sup>2</sup> CHU Rennes, Hôpital Pontchaillou, Pôle des Micro-organismes, Service de Bactériologie, Virologie et Hygiène Hospitalière, Rennes F-35033,<sup>3</sup> Université de Toulouse, UPS, Faculté de Médecine, Toulouse F-31400,<sup>4</sup> Laboratoire de Bactériologie-Hygiène, CHU de Toulouse, Institut Fédératif de Biologie, Toulouse F-31059,<sup>5</sup> INRA, UMR 1225, Toulouse F-31076,<sup>6</sup> and Université de Toulouse, ENVT, UMR 1225, Toulouse F-31076,<sup>7</sup> France

Received 3 December 2009/Returned for modification 15 January 2010/Accepted 29 March 2010

**Determinants of urosepsis in *Escherichia coli* remain incompletely defined. Cyclomodulins (CMs) are a growing functional family of toxins that hijack the eukaryotic cell cycle. Four cyclomodulin types are actually known in *E. coli*: cytotoxic necrotizing factors (CNFs), cycle-inhibiting factor (Cif), cytolethal distending toxins (CDTs), and the *pks*-encoded toxin. In the present study, the distribution of CM-encoding genes and the functionality of these toxins were investigated in 197 *E. coli* strains isolated from patients with community-acquired urosepsis ( $n = 146$ ) and from uninfected subjects ( $n = 51$ ). This distribution was analyzed in relation to the phylogenetic background, clinical origin, and antibiotic resistance of the strains. It emerged from this study that strains harboring the *pks* island and the *cnfI* gene (i) were strongly associated with the B2 phylogroup ( $P, <0.001$ ), (ii) frequently harbored both toxin-encoded genes in phylogroup B2 (33%), and (iii) were predictive of a urosepsis origin ( $P, <0.001$  to  $0.005$ ). However, the prevalences of the *pks* island among phylogroup B2 strains, in contrast to those of the *cnfI* gene, were not significantly different between fecal and urosepsis groups, suggesting that the *pks* island is more important for the colonization process and the *cnfI* gene for virulence. *pks*- or *cnfI*-harboring strains were significantly associated with susceptibility to antibiotics (amoxicillin, cotrimoxazole, and quinolones [ $P, <0.001$  to  $0.043$ ]). Otherwise, only 6% and 1% of all strains harbored the *cdtB* and *cif* genes, respectively, with no particular distribution by phylogenetic background, antimicrobial susceptibility, or clinical origin.**

The bacterial species *Escherichia coli* comprises a wide diversity of strains belonging to the commensal intestinal flora of humans and warm-blooded animals. Among these strains, several pathogenic variants cause intestinal or extraintestinal infections in humans and animals (33). Population genetic studies based on multilocus enzyme electrophoresis and various DNA markers (10, 20, 44) classify the *E. coli* strains into four major phylogenetic groups (A, B1, B2, and D). The groups are diversely associated with certain ecological niches and propensities to cause disease.

Extraintestinal pathogenic *E. coli* (ExPEC) strains are facultative pathogens that are not yet fully described. They are reported to belong mainly to phylogroups B2 and D, and they possess high numbers of virulence genes that belong to a flexible gene pool (43, 53). Among ExPEC strains, uropathogenic *E. coli* (UPEC) strains take advantage of host behavior and susceptibility by employing virulence factors that facilitate bacterial growth and persistence in the urinary tract (5, 28–30). Important virulence mechanisms are adhesion, invasion, subversion of host defenses, and direct interference with host cellular functions via secreted effectors (33, 69).

These effectors include the cyclomodulins (CMs), a functional class of toxins that hijack the cell cycle, a fundamental

host cell process (48). In the species *E. coli*, four kinds of CMs have been identified: the rho GTPase-targeting toxins CNF-1 to CNF-3 (cytotoxic necrotizing factors) (34), the cycle-inhibiting factor (Cif) (38), and two kinds of genotoxins, cytolethal distending toxins (CDTs) I to V (19) and the recently discovered colibactin (41). Colibactin is probably a hybrid polyketide-nonribosomal peptide toxin, whose activity is encoded by the genomic *pks* island (41). CDTs, Cif, and colibactin block mitosis, whereas CNFs promote DNA replication without cytokinesis. CM production can therefore be detected by the analysis of the cytopathic effects induced (41, 50).

*E. coli* CMs are encoded by mobile elements (genomic islands, plasmids, and bacteriophages) that belong to the flexible gene pool of *E. coli* (22). The aim of the present study was to compare the prevalences of CMs in *E. coli* strains that differed in their clinical origins (community-acquired urosepsis or feces), phylogroups, and susceptibilities to antimicrobial agents.

### MATERIALS AND METHODS

**Recruitment of patients and strains.** One hundred forty-six *E. coli* strains were collected from blood cultures of 146 adults with community-acquired urosepsis between May 2006 and July 2008 in two university hospitals in France (Clermont-Ferrand [ $n = 86$ ] and Rennes [ $n = 60$ ]). Community-acquired urosepsis was defined as the association of a urinary tract infection ( $>10^4$  leukocytes/ml and  $>10^5$  CFU/ml) with bacteremia due to the same *E. coli* strain in patients who had not previously been admitted to hospital and whose hospital stay did not exceed 48 h. Patients with histories of urine disorders (previous hospitalization in a urology or nephrology unit, urethral instrumentation, or nephrostomy) were excluded. Of the 146 isolates, 123 came from samples taken in emergency wards, 5 in intensive care units, and 18 in other medical units. The median age of

\* Corresponding author. Mailing address: Laboratoire de Bactériologie, 28 place H. Dunant, 63001 Clermont-Ferrand, France. Phone: 33 4 73 75 49 20. Fax: 33 4 73 75 49 22. E-mail: rbonnet@chu-clermontferrand.fr.

<sup>∇</sup> Published ahead of print on 7 April 2010.

TABLE 1. Archetypal *E. coli* control strains used in this study

<i>E. coli</i> strain <sup>a</sup>	Gene(s) or classification method	Host origin	Source or reference
28C	<i>hly</i> , <i>cnf1</i> , <i>cdt-IV</i> , <i>clbA</i> , <sup>b</sup> <i>clbK</i> , <sup>b</sup> <i>clbJ</i> , <sup>b</sup> <i>clbQ</i> <sup>b</sup>	Porcine septicemia	13
1404	<i>cnf2</i> , <i>cdt-III</i>	Bovine septicemia	47
IHE3034 $\Delta$ <i>clbP</i>	<i>cdt-I</i> , <i>clb</i> $\Delta$ <i>clbP</i>	Newborn meningitis	41
DH5 $\alpha$ (pCP2123)	<i>cdt-II</i>	Laboratory strain	54
C48a	<i>cnf3</i> , <i>eae</i> , <i>cif</i> <sup>b</sup>	Healthy lamb	45
DH10 $\beta$ (pBAC <i>pks</i> )	<i>clb</i>	Laboratory strain	41
E22	<i>cif</i> , <i>eae</i>	Rabbit, EPEC	38
RS218	PCR-based phylotyping	Newborn meningitis	61
EDL933	<i>eae</i> , <i>stx</i> <sub>1</sub> , <i>stx</i> <sub>2</sub>	EHEC	51
DH5 $\alpha$		Laboratory strain	Novagen

<sup>a</sup> Strain C48a was kindly provided by José Antonio Orden, and strain RS218 was kindly provided by Philippe Bidet and Edouard Bingen.

<sup>b</sup> The *clbA*, *clbK*, *clbJ*, *clbQ*, and *cif* genes were detected during this study.

patients with urosepsis was 76 years (range, 20 to 96 years). There were 38 males (26%) (median age, 71 years; range, 22 to 93 years) and 108 females (74%) (median age, 76 years; range, 20 to 96 years).

Fecal isolates ( $n = 51$ ) were collected by anal swabs from adults in the community who had no evidence of acute infection or other disorders of the gastrointestinal tract and who had not received antibiotics in the preceding month. The healthy individuals had no health care association within the past 6 months. Healthy individuals who had had urinary tract infections in the previous 12 months and patients with benign or malignant gastrointestinal disease were excluded. The median age of the 51 patients was 46.5 years (range, 23 to 62 years). There were 18 males (35%) (median age, 30.5 years; range, 24 to 59 years) and 33 females (65%) (median age, 50 years; range, 23 to 62 years). One arbitrarily selected *E. coli* colony per sample was analyzed. Previous data show that there is an 86% probability that an arbitrarily selected fecal *E. coli* colony represents the quantitatively predominant clone in the sample (35).

The *E. coli* strains were identified with the automated Vitek II system

(bioMérieux). The experimental guidelines of the authors' institutions were followed in the conduct of clinical research. Epidemiological and laboratory results for each episode were recorded anonymously in a computer database in accordance with French law. *E. coli* archetypal reference strains and the *E. coli* K-12 strain DH5 $\alpha$ , which are used as positive and negative controls, respectively, for genotypic and phenotypic analyses, are given in Table 1. All strains were stored on 15% glycerol-supplemented Luria-Bertani (LB) medium at  $-80^{\circ}\text{C}$ .

**DNA extraction.** Template DNA was extracted by the boiling-water method. Briefly, three to five bacterial colonies from a freshly grown culture were suspended in 150  $\mu\text{l}$  sterile distilled water and were incubated for 15 min at  $95^{\circ}\text{C}$ . After chilling on ice, bacterial debris was removed by centrifugation at  $15,000 \times g$  for 5 min at  $4^{\circ}\text{C}$ . The supernatant was transferred to a fresh microcentrifuge tube and was stored at  $-20^{\circ}\text{C}$  until required.

**PCR phylogenetic grouping and fingerprint profiling.** *E. coli* strains were classified according to the *Escherichia coli* Reference Collection (ECOR) system (23) into phylogenetic groups A, B1, B2, and D by using a multiplex PCR technique (10). The *uidA* gene was used as an internal control of amplification, and the anonymous gene *svg* was detected to distinguish the B2<sub>1</sub> ribotyping/ST29 multilocus sequence type (MLST) isolates from the B2 strains (3). Three subtypes of phylogroup B2 strains were differentiated by this technique and were designated B2<sub>1</sub>, Tsp-negative B2, and Tsp-positive B2. A DNA positive control was performed with strain RS218, which harbors all the genes targeted by the multiplex PCR. In order to investigate the clonal relationships among the *pks*- and *cnfI*-positive strains, fingerprint profiles had been generated for each isolate using the enterobacterial repetitive intergenic consensus (ERIC)-PCR scheme reported by Adamus-Bialek et al. (1).

**Detection of CM-producing genes by dot blot DNA hybridization.** CM-encoding genes were detected by dot blot DNA hybridization experiments. The probes were obtained by PCR with previously published primers (Table 2) by using the PCR DIG probe synthesis kit (Roche Molecular Biochemicals, France) according to the manufacturer's instructions. Two-microgram DNA samples were fixed onto positively charged nylon membranes by UV illumination for 20 min. Hybridization was performed using the Roche labeling and detection kit (Roche Molecular Biochemicals) as indicated by the manufacturer. Each spot was checked with a 16S rRNA gene probe. The *pks* island, which contains the

TABLE 2. Primers used in this study

Primer	Oligonucleotide sequence (5'-3')	$T_m$ of PCR ( $^{\circ}\text{C}$ )	Specificity	PCR product size (bp)	Source or reference
91E	TCAAA(G,T)GAATTGACGGGGGC	54	16S rRNA gene	473	58
13B	GCCCGGGAACGTATTAC				18
<i>pks</i> ORF9-10.1KJ	ATTCGATAGCGTCACCCAAC	58	<i>clbK-clbJ</i>	2,119	41
<i>pks</i> ORF9-10.2KJ	TAAGCGTCTGGAATGCAGTG				
IHAPJPN42	CAGATACACAGATACCATTCA	55	<i>clbA</i>	1,002	27
IHAPJPN46	CTAGATTATCCGTGGCGATTCC				
IHAPJPN55	TTATCCTGTAGCTTTCGTTCC	55	<i>clbQ</i>	821	27
IHAPJPN56	CTTGTATAGTTACAACTATTTCC				
CNF-1s	GGGGGAAGTACAGAAGAATTA	48	<i>cnf1</i>	1,112	64
CNF-1as	TTGCCGTCCACTCTCACCAGT				
CNF-2s	TATCATACGGCAGGAGGAAGCACC	48	<i>cnf2</i>	1,241	66
CNF-2as	GTCACAATAGACAATAATTTCCG				
CNF3-3D	TAACGTAATTAGCAAAGA	48	<i>cnf3</i>	757	45
CNF-3as	GTCTTACTTACTTACAGT				This study
CDT-s1	GAAAGTAAATGGAATATAAATGTCCG	60	<i>cdtB-II</i> , <i>cdtB-III</i> , <i>cdtB-V</i>	467	64
CDT-as1	AAATCACCAAGAATCATCCAGTTA				
CDT-IIas <sup>a</sup>	TTTGTGTTGCCGCGCTGGTGAAA	62	<i>cdtB-II</i>	556	64
CDT-IIIas <sup>a</sup>	TTTGTGTCGGTGCAGCAGGGAAAA	62	<i>cdtB-III</i> , <i>cdtB-V</i>	555	64
CDT-s2	GAAAATAAATGGAACACACATGTCCG	56	<i>cdtB-I</i> , <i>cdtB-IV</i>	467	64
CDT-as2	AAATCTCCTGCAATCATCCAGTTA				
CDT-Is	CAATAGTCGCCACAGGA	56	<i>cdtB-I</i>	411	64
CDT-Ias	ATAATCAAGAACACCACCAC				
CDT-IVs	CCTGATGGTTCAGGAGGCTGGTTC	56	<i>cdtB-IV</i>	350	64
CDT-IVas	TTGCTCCAGAATCTATACCT				
P105	GTCAACGAACATTAGATTAT	49	<i>cdtC-V</i>	748	25
c2767r	ATGGTCATGCTTTGTTATAT				
<i>cif-int-s</i>	AACAGATGGCAACAGACTGG	55	<i>cif</i>	383	38
<i>cif-int-as</i>	AGTCAATGCTTTATGCGTCAT				

<sup>a</sup> Used with primer CDT-s1.

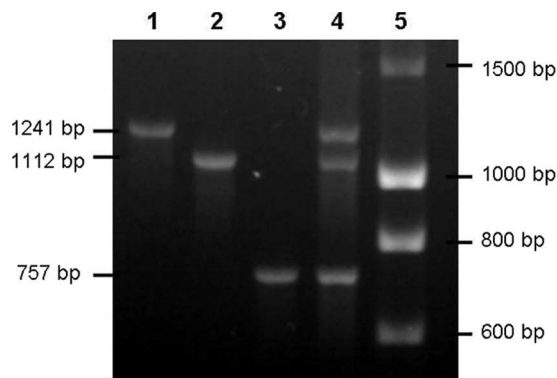


FIG. 1. Standardization of *cnf* gene typing by multiplex PCR. Lanes: 1, strain 1404 (*cnf2*); 2, strain 28C (*cnf1*); 3, strain C48a (*cnf3*); 4, mixture of equal quantities of 1404, 28c, and C48a DNAs (*cnf1*, *cnf2*, *cnf3*); 5, DNA ladder (Eurogentec).

colibactin-producing gene cluster *clb*, was screened with a probe overlapping the *clbK* and *clbJ* genes. The *cnf* genes were detected with a mixture of probes specific to *cnf1*, *cnf2*, and *cnf3*. The *cdtB* genes were detected by two hybridization experiments with the *cdtB-II-cdtB-III* and *cdtB-I-cdtB-IV* probe mixtures. The *cif* gene was detected using an internal specific probe. The sensitivities and specificities of the probes were checked on each membrane by spotting DNA extracts of all CM control strains.

**Identification of the CM-producing genes and other virulence factors by PCR assays.** Positive hybridizations with a CM probe were subjected to confirmatory PCR assays using the primers given in Table 2. The reaction mixture contained 50 ng DNA sample, 0.2 mM each deoxynucleoside triphosphate (dNTP), 0.4  $\mu$ M each primer, 3 mM MgCl<sub>2</sub>, and 1.0 U RedGoldStar DNA polymerase (Eurogentec, France) in the corresponding reaction buffer. Primers located in the 5' and 3' regions of the *pks* island (the *clbA* and *clbQ* genes) were used to confirm the full presence of the colibactin-producing island. A multiplex PCR was used to distinguish the *cnf1*, *cnf2*, and *cnf3* genes. Figure 1 shows the standardization of the *cnf* multiplex PCR. *cdtB-I*, *cdtB-II*, *cdtB-III/cdtB-V*, and *cdtB-IV* were differentiated by simplex PCR. The *cdt-III* and *cdt-V* operons were differentiated by *cdtC* gene sequencing (Cogenics, Meylan, France) (Table 2) (4). Isolates harboring *cdt* or *cif* were further tested by PCR for the additional virulence factor genes *stx1*, *stx2*, and *eae*, as previously described (8).

**Phenotypic tests.** The cytopathic effects of CNF and CDTs were investigated in all strains, and those of colibactin and Cif were investigated in nonhemolytic strains, as previously described (41, 50). Briefly, the effects of CDT and CNF were detected with a cell-lysate-interacting test. After a 48-h culture at 37°C with shaking in Luria-Bertani broth medium, bacterial cells were sonicated and were sterile filtered separately using 0.22- $\mu$ m-pore-size filters. HeLa cells were treated with the sterile sonicated lysates until analysis. The effects of colibactin and Cif were detected using a cell-bacterium-interacting test, which was based on the interaction between HeLa cells and bacteria. Overnight Luria-Bertani broth cultures of *E. coli* were diluted in interaction medium, and HeLa cell cultures were infected at multiplicities of infection (numbers of bacteria per cell at the onset of infection) of 100 and 200. Cells were washed 3 to 6 times 4 h after inoculation and were incubated in cell culture medium with 200  $\mu$ g/ml gentamicin until analysis. After 72 h of incubation at 37°C under a 5% CO<sub>2</sub> atmosphere, the medium was removed by several washes of the HeLa cell monolayers. The morphological changes characteristic of CDT, CNF, colibactin, and Cif were observed after staining with Giemsa stain, as described elsewhere (41).

The detection of alpha-hemolysin was performed for all strains studied by overnight growth at 37°C on Columbia sheep blood (5%) agar (Oxoid, Dardilly, France). Antibiotic susceptibility tests were performed by the disc diffusion method (Bio-Rad), using CA-SFM interpretative criteria (7). *E. coli* 25922 (ATCC) was used as the reference strain.

**Statistical analysis.** Statistical analysis was performed using the Fisher exact and chi-square tests. (For multiple-group comparisons, an initial chi-square test for heterogeneity was done, and only if this yielded a *P* value of <0.05 were the individual pairwise comparisons tested.) The threshold for statistical significance was a *P* value of <0.05.

No statistical difference in the distribution of phylogenetic groups, CM genotypes, or antimicrobial resistance phenotypes was observed between strain

TABLE 3. Distribution of phylogenetic groups among 197 *E. coli* isolates recovered from patients with urosepsis and from the feces of healthy individuals

Phylogenetic group	No. (%) of <i>E. coli</i> isolates		<i>P</i> <sup>a</sup>
	Fecal ( <i>n</i> = 51)	Urosepsis ( <i>n</i> = 146)	
A	19 (37)	27 (18)	(0.011)
B1	12 (24)	7 (5)	(<0.001)
B2	12 (24)	79 (54)	<0.001
D	8 (16)	33 (23)	

<sup>a</sup> *P* values (by Fisher's exact test) are shown where *P* is <0.05. *P* values given in parentheses indicate negative associations between the prevalence of the urosepsis strains and the particular phylogroup.

groups of the same clinical origin from different geographic sources (except for resistance to extended-spectrum cephalosporins [C3G], where the *P* value was 0.041). The strains were therefore grouped together and were analyzed according to their urosepsis or fecal origin.

## RESULTS AND DISCUSSION

***E. coli* phylogroups and clinical origin.** *E. coli* strains isolated from the fecal and urosepsis groups differed significantly in the prevalences of phylogenetic groups A, B1, and B2 but not in that of phylogroup D (Table 3). Phylogroup B2 was predominant in urosepsis strains (54%), followed by phylogroups D (23%), A (18%), and B1 (5%), as previously observed (5, 28–30). Fecal isolates belonged most frequently to phylogenetic group A (37%), followed by phylogroups B1 and B2, with identical prevalences (24%). These results are different from those of other studies, which found B2 strains predominant in fecal samples of healthy subjects, notably in industrialized countries (14, 16, 17, 29, 42, 71). This difference might be explained by the impact of geographic/climatic conditions, dietary factors, and/or the use of antibiotics or host genetic factors on the commensal flora (14, 16, 62). Hence, phylogroup A, and to a lesser extent phylogroup B1, was significantly more prevalent among fecal strains than among urosepsis strains (37% versus 18% [*P*, 0.011] and 24% versus 5% [*P*, <0.001], respectively). In contrast, phylogroup B2, known to encompass the most virulent ExPEC strains (15), was significantly more predominant in urosepsis strains than in fecal strains (54% versus 24% [*P*, <0.001]).

**Distribution of CM-encoding genes according to phylogenetic background.** Table 4 shows a clear heterogeneity in the

TABLE 4. Phylogenetic distribution of cyclomodulin genes and Hly among 197 *E. coli* isolates recovered from patients with urosepsis and from the feces of healthy individuals

Virulence factor	No. (%) of <i>E. coli</i> isolates					<i>P</i> <sup>a</sup>
	Total ( <i>n</i> = 197)	Group A ( <i>n</i> = 46)	Group B1 ( <i>n</i> = 19)	Group B2 ( <i>n</i> = 91)	Group D ( <i>n</i> = 41)	
<i>pks</i>	53 (27)	0	0	53 (58)	0	<0.001
<i>cnf1</i> -Hly	36 (18)	1 (2)	1 (5)	34 (37)	0	<0.001
Hly alone	15 (8)	5 (11)	0	8 (9)	2 (5)	
<i>cnf1</i> -Hly, <i>pks</i>	30 (15)	0	0	30 (33)	0	<0.001
<i>cdtB-I</i>	5 (2)	1 (2)	0	4 (4)	0	
<i>cdtB-IV</i>	6 (3)	0	0	5 (5)	1 (2)	
<i>cdtB-V</i>	1 (1)	0	1 (5)	0	0	
<i>cif</i>	2 (1)	0	2 (11)	0	0	

<sup>a</sup> *P* values (by Fisher's exact test) are shown where *P* is <0.05 and are for comparisons of group B2 versus all other groups combined.



TABLE 5. Distribution of cyclomodulin genes and Hly among 197 *E. coli* isolates according to phylogenetic group and clinical origin

CM gene(s)	No. (%) of all isolates (n = 197)		P value for all isolates <sup>a</sup>	No. (%) of isolates						No. (%) of group D isolates (n = 41)		
	Fecal (n = 51)	Urosepsis (n = 146)		Group A (n = 46)		Group B1 (n = 19)		Group B2 (n = 91)		P value for group B2 <sup>a</sup>	Fecal (n = 8)	Urosepsis (n = 33)
				Fecal (n = 19)	Urosepsis (n = 27)	Fecal (n = 12)	Urosepsis (n = 7)	Fecal (n = 12)	Urosepsis (n = 79)			
<i>pks</i>	6 (12)	47 (32)	0.005	0	0	0	0	6 (50)	47 (59)	0.028	0	0
<i>cnf1</i> -Hly	1 (2)	35 (24)	<0.001	0	1 (4)	0	1 (14)	1 (8)	33 (42)		0	0
Hly alone	3 (6)	12 (8)		2 (11)	3 (11)	0	0	1 (8)	7 (9)		0	2 (6)
<i>cnf1</i> -Hly, <i>pks</i>	1 (2)	29 (20)	0.001	0	0	0	0	1 (8)	29 (37)		0	0
<i>cdtB-I</i>	2 (4)	3 (2)		1 (5)	0	0	0	1 (8)	3 (4)		0	0
<i>cdtB-IV</i>	3 (6)	3 (2)		0	0	0	0	2 (17)	3 (4)		1 (12)	0
<i>cdtB-V</i>	1 (2)	0		0	0	1 (8)	0	0	0		0	0
<i>cif</i>	0	2 (1)		0	0	0	2 (29)	0	0		0	0

<sup>a</sup> P values (by Fisher's exact test) are shown where P is <0.05.

prevalences of CM genes, of which the most frequent were the *pks* island (27%) and *cnf1* (18%), and in that of alpha-hemolysin expression (26%), three traits that were strongly associated with phylogroup B2.

All *pks*-harboring strains belonged to phylogroup B2, as previously observed (27, 41, 56). This association was deep-seated (P, <0.001 for B2 versus A, B1, and D combined or separately). Fifty-eight percent of B2 strains possessed this trait, which was observed in all B2 subtypes according to phylogenetic grouping by PCR (3 among 12 B<sub>2</sub> strains, 5 among 7 Tsp-negative B2 strains, and 45 among 75 Tsp-positive B2 strains).

Multiplex PCR-based *cnf* typing revealed only *cnf1* genes. No *cnf2* or *cnf3* genes, which were first observed in *E. coli* strains isolated from animals (12, 45), were detected in our human strains, a finding consistent with the absence or very weak prevalence of these genes observed in previous studies (9, 12, 32, 37, 55, 63, 64). These results suggest that *cnf2*- and *cnf3*-positive strains are almost entirely absent in humans and thus are probably involved only in animal diseases. *cnf1*-harboring strains were highly associated with phylogroup B2, including 37% of *E. coli* B2 strains, but were barely present in the other phylogroups (P, <0.001 for B2 versus other groups), as previously observed (5, 26, 29, 30). Moreover, 33% of B2 strains harbored both the *pks* island and the *cnf1* gene (P, <0.001 for B2 versus other groups), as previously observed (27).

All *cnf1*-positive strains exhibited the alpha-hemolytic phenotype. This association was probably due to the presence of the pathogenicity island (PAI) II<sub>J96</sub>-like domain, in which the *cnf1* gene is located just downstream of the *hlyCABD* operon, which encodes the alpha-hemolytic phenotype (6, 34). We observed two isolates from one patient, one hemolytic *cnf1*-positive and one nonhemolytic *cnf1*-negative isolate, sharing the same randomly amplified polymorphic DNA (RAPD) 1283 typing and antimicrobial susceptibility profiling, suggesting the spontaneous loss of a PAI II<sub>J96</sub>-like domain in the second isolate, as a result of the instability of the PAI (39). Only the hemolytic *cnf1*-positive isolate was retained for statistical analysis. Only 8% of strains exhibited an alpha-hemolytic phenotype without the *cnf1* gene. These strains, which accounted for 29% of the alpha-hemolytic strains, had no particular phylogenetic distribution pattern. They probably possessed another

PAI containing the *hly* operon, similar to the PAI I<sub>CFT073</sub>-like domain (21).

*cdtB* genes were observed in 6% of strains, with only one *cdtB* subtype per strain. Although 75% of *cdtB*-positive strains belonged to phylogroup B2, no significant phylogenetic distribution pattern clearly emerged, even among *cdtB* subtypes. In our study, *cdtB-I* (n = 5) and *cdtB-IV* (n = 6) types were overrepresented compared to the *cdtB-V* (n = 1) type. In contrast, no *cdt-II*- or *cdt-III*-positive strains were found. Of the *cdt* genes, *cdt-I* and *cdt-IV* are the most closely homologous, and both genes, framed with lambdoid prophage genes (65), might have been acquired from a common ancestor by phage transduction. *cdt* genes did not possess any particular association with the *pks* island or the *cnf1* gene (see Table 6). Since *cdt* genes have been extensively investigated in Shiga toxin-producing *E. coli* (STEC) strains, *cdt*-harboring strains were screened for the *eae* and *stx* genes (4, 25, 46, 49). No *eae* or *stx* genes were detected in our *cdt*-encoding strains.

Two *cif*-harboring strains were observed. They belonged to phylogroup B1 and harbored no other CM-encoding gene. *Cif* is an effector protein of the type 3 secretion system (T3SS) encoded by the locus of enterocyte effacement pathogenicity island, and previous observations of the *cif* gene were restricted to enteropathogenic *E. coli* (EPEC) and enterohemorrhagic *E. coli* (EHEC) strains (36, 38). The two *cif*-positive strains of this study were confirmed as EPEC strains by detection of the *eae* gene but not of *stx* genes.

**Distribution of CM-encoding genes according to clinical origin.** Overall, the urosepsis strains had a significantly higher prevalence of the *pks* island than the fecal strains (32% versus 12%; P, 0.005) (Table 5). The *pks* island could therefore be involved in uropathogenesis; however, the prevalences of the *pks* island among phylogroup B2 strains were not significantly different between the fecal and urosepsis groups. The urosepsis strains were also significantly more likely to harbor Hly (P, <0.001) and *cnf1* associated with Hly (P, <0.001) than the fecal strains. Likewise, the B2 urosepsis strains harbored *cnf1* associated with Hly significantly more frequently than their B2 fecal counterparts (P, 0.028).

The weak prevalence of *cdtB* genes in urosepsis strains (4%) has been documented elsewhere (28, 29, 31). These observations suggest that CDTs are not major virulence factors for urosepsis. *cdt* genes in *E. coli* urosepsis strains were repre-

TABLE 6. Distribution of phylogenetic group-cyclomodulin gene profiles among 197 *E. coli* isolates recovered from patients with urosepsis and from the feces of healthy individuals

Phylogenetic group	CM gene profile	No. (%) of isolates		<i>P</i> <sup>a</sup>
		Fecal ( <i>n</i> = 51)	Urosepsis ( <i>n</i> = 146)	
B1	None	11 (21.6)	4 (2.7)	(0.003)
	<i>cdtB-V</i>	1 (2)	0	
	<i>cif</i>	0	2 (1.4)	
	<i>cnf1</i>	0	1 (0.7)	
A	None	18 (35.3)	26 (17.8)	(0.018)
	<i>cnf1</i>	0	1 (0.7)	
	<i>cdtB-I</i>	1 (2)	0	
B2	None	4 (7.8)	26 (17.8)	0.003
	<i>cnf1</i>	0	4 (2.7)	
	<i>pks</i>	4 (7.8)	17 (11.6)	
	<i>pks, cnf1</i>	1 (2)	26 (17.8)	
	<i>pks, cnf1, cdtB-I/cdtB-IV<sup>b</sup></i>	0	3 (2)	
	<i>pks, cdtB-I</i>	1 (2)	1 (0.7)	
	<i>cdtB-I/cdtB-IV<sup>b</sup></i>	2 (4)	2 (1.4)	
D	None	7 (13.7)	33 (22.6)	
	<i>cdtB-IV</i>	1 (2)	0	

<sup>a</sup> *P* values (by Fisher's exact test) are shown where *P* is <0.05. *P* values given in parentheses indicate negative associations between the prevalence of the urosepsis strains and the particular phylogenetic group-cyclomodulin gene profile.

<sup>b</sup> Because of their small numbers, high level of homology, and close epidemiology, *cdt-I* and *cdt-IV* were aggregated.

sented exclusively by the *cdtB-I* and *cdtB-IV* types. *cdtB* genes were more prevalent in the fecal isolates than in the urosepsis isolates (12% versus 4%), but not to the level of significance. This is in contrast with other studies in which *cdtB* prevalence in fecal strains ranged from 0.9% to 5%, with similar prevalences in urosepsis strains (29, 30, 32, 64). This difference may be explained in part by the exhaustive research into *cdt* subtypes that was performed in our study. In several studies, CDTs were detected in EPEC or STEC/EHEC strains isolated from patients with diarrhea (2, 4, 46, 49). Our CDT-producing strains were not isolated during bouts of enteric disease and did not possess *eae* or *stx* genes. A case-control study of CDT-producing *E. coli* showed no association between CDT-positive *E. coli* and diarrhea (2). Our *cdtB-V*-positive fecal strain did not possess *stx* and *eae* genes, while the *cdtB-V* gene has been reported only in STEC collections (4, 46). Finally, the two *cif*-positive, *eae*-positive urosepsis strains suggest the potential involvement of EPEC as an opportunistic organism in extraintestinal infections. The patients were both female and >80 years old, two risk factors for urinary sepsis.

**Distribution of phylogroup-CM gene profiles according to clinical origin.** CM gene profiles and phylogroups were coanalyzed to determine whether combinations of CMs and phylogroups can also differentiate between the two clinical groups of strains (Table 6). Phylogroup B1 and A strains with no CM-encoding genes were significantly more prevalent in feces than in the blood cultures of patients with urosepsis (*P*, 0.003 and 0.018, respectively), whereas the association of the CM profile *pks cnf1* with phylogroup B2 strains was more widespread in urosepsis strains (*P*, 0.003). Overall, both the *pks* island and the *cnf1* gene, whether associated with another CM or not, were

highly predictive of a urosepsis origin (20% versus 2%; *P*, 0.001) (Table 5). Three fingerprint profiles (encompassing 17, 10, and 3 strains) were obtained from the 30 *pks*-, *cnf1*-, and group B2-positive strains, suggesting that these urosepsis strains can belong to a distinct genetic background.

During urosepsis, colibactin and CNF1 may induce profound changes in cellular signaling pathways. Colibactin induces DNA double-strand breaks (41). CNF1 modulates a high number of cellular functions by hijacking rho GTPases (34), particularly attenuates polymorphonuclear leukocyte functions (11), and induces a severe and controlled inflammatory response (40, 57, 59, 60). By affecting the immune response, CNF1 could lengthen the brief time window between the establishment of bacteriuria and the activation of a host defense during urosepsis, consequently enhancing UPEC survival and allowing invasion of the parenchyma and bacteremia. CNF1 and probably colibactin may also favor host colonization, since their encoding genes have been found together in group B2 strains responsible for asymptomatic bacteriuria (70). In this study, the difference in the prevalences of the *pks* island and the *cnf1* gene in B2 fecal strains suggest that colibactin may act mainly as a colonization factor and CNF-1 as a virulence factor (Table 5). Further experimental investigations are needed to shed light on the pathogenic and/or colonizer roles of these CMs.

**Phenotypic detection of CMs.** Interpretation of the four CM-related phenotypes in eukaryotic cells was straightforward in all but two strains, which harbored the *pks* island and the *cdt* gene and for which the phenotypes are similar in the cell-bacterium-interacting test. Thus, we were not able to affirm that the *pks* islands of these two strains were functional. Nevertheless, all *cdt*- and *cnf*-positive strains produced a cytopathic effect, and a cytopathic effect induced by the *pks* island was observed in all but three strains. This result suggests nonfunctional *pks* islands. However, we cannot rule out the possibility that other genes, potentially involved in the secretion of colibactin or in the synthesis of precursors used by *pks* island-encoded enzymes, were altered. It is noteworthy that two of these nonfunctional *pks* islands did not belong to ExPEC strains and accounted for 33% of *pks*-positive fecal isolates. Only one of the two *cif* genes produced a cytopathic effect. Loukiadis et al. reported that 66% of *cif*-positive *E. coli* strains did not induce a Cif-related phenotype in eukaryotic cells due to frameshift mutations or an insertion sequence in the *cif* gene (36). However, a nonfunctional T3SS may also explain the absence of a cytopathic effect. In addition, all strains that induced cytopathic effects harbored CM-encoding genes, suggesting that the genomic techniques used had reliable sensitivity.

**Antibiotic susceptibilities of urosepsis strains according to CM-encoding genes and phylogenetic background.** The prevalence of antibiotic resistance among urosepsis strains was as follows: 59% were resistant to amoxicillin, 5% to extended-spectrum cephalosporins (C3G), 25% to cotrimoxazole, 18% to nalidixic acid and norfloxacin, 14% to ciprofloxacin, 2% to gentamicin, and 0% to amikacin. Table 7 shows the prevalences of CM-encoding genes and phylogenetic groups according to antibiotic susceptibility status. Quinolone susceptibility was associated with phylogenetic group B2 (59% of susceptible versus 33% of resistant isolates; *P*, 0.019) and the *pks* island

TABLE 7. Distribution of antibiotic susceptibility according to cyclomodulin genes/Hly and phylogenetic background among 143 *E. coli* isolates recovered from patients with urosepsis<sup>a</sup>

Variable	Quinolones–fluoroquinolones <sup>b</sup>			Amoxicillin			C3G <sup>c</sup>			Gentamicin		Cotrimoxazole		<i>P</i> <sup>d</sup>
	No. (%) of isolates		<i>P</i> <sup>d</sup>	No. (%) of isolates		<i>P</i> <sup>d</sup>	No. (%) of isolates		<i>P</i> <sup>d</sup>	No. (%) of isolates		No. (%) of isolates		
	R ( <i>n</i> = 27)	S ( <i>n</i> = 119)		R ( <i>n</i> = 86)	S ( <i>n</i> = 60)		R ( <i>n</i> = 7)	S ( <i>n</i> = 139)		R ( <i>n</i> = 3)	S ( <i>n</i> = 143)	R ( <i>n</i> = 36)	S ( <i>n</i> = 110)	
Phylogenetic groups														
A	11 (41)	16 (13)	(0.002)	24 (28)	3 (5)	(<0.001)	6 (86)	21 (15)	(<0.001)	2 (67)	25 (17)	9 (25)	18 (16)	0.020
B1	2 (7)	5 (4)		5 (6)	2 (3)		1 (14)	6 (4)		1 (33)	6 (4)	3 (8)	4 (4)	
B2	9 (33)	70 (59)	0.019	32 (37)	47 (78)	<0.001	0 (0)	79 (57)	0.003	0 (0)	79 (55)	13 (36)	66 (60)	
D	5 (19)	28 (24)		25 (29)	8 (13)	(0.028)	0 (0)	33 (24)		0 (0)	33 (23)	11 (31)	22 (20)	
CMs														
<i>pks</i>	3 (11)	44 (37)	0.011	18 (21)	29 (48)	<0.001	0	47 (34)		0	47 (33)	2 (6)	45 (41)	<0.001
<i>cnfI</i> -Hly	7 (26)	28 (24)		14 (16)	21 (35)	0.011	2 (29)	33 (24)		0	35 (24)	4 (11)	31 (28)	0.043
Hly alone	1 (4)	11 (9)		8 (9)	4 (7)		0	12 (9)		0	12 (8)	2 (6)	10 (9)	
<i>cdtB-I</i>	0	3 (3)		1 (1)	2 (3)		0	3 (2)		0	3 (2)	0	3 (3)	
<i>cdtB-IV</i>	0	3 (3)		1 (1)	2 (3)		0	3 (2)		0	3 (2)	0	3 (3)	
<i>cif</i>	0	2 (2)		1 (1)	1 (2)		0	2 (1)		0	2 (1)	0	2 (2)	

<sup>a</sup> R, resistant; S, susceptible.

<sup>b</sup> All resistant isolates were resistant both to quinolone (nalidixic acid) and to fluoroquinolones (either norfloxacin alone or norfloxacin and ciprofloxacin).

<sup>c</sup> C3G, extended-spectrum cephalosporins.

<sup>d</sup> *P* values (by Fisher's exact test) are shown where *P* is <0.05. *P* values shown in parentheses indicate negative associations between the prevalence of the susceptible strains and the particular trait (phylogenetic group or CM).

(37% versus 11%; *P*, 0.011), whereas quinolone resistance was associated with group A (41% of resistant versus 13% of susceptible isolates; *P*, 0.002). Likewise, amoxicillin susceptibility was associated with group B2 (78% versus 37%; *P*, <0.001), the combination of the alpha-hemolytic phenotype and the *cnfI* gene (35% versus 16%; *P*, 0.011), and the *pks* island (48% versus 21%; *P*, <0.001), whereas amoxicillin resistance was associated with groups A and D (28% versus 5% [*P*, <0.001] and 29% versus 13% [*P*, 0.028], respectively). Cotrimoxazole susceptibility was also associated with group B2 (60% versus 36%; *P*, 0.02), the alpha-hemolytic phenotype (36% versus 17%; *P*, 0.024), the combination of the alpha-hemolytic phenotype and the *cnfI* gene (28% versus 11%; *P*, 0.043), and the *pks* island (41% versus 6%; *P*, <0.001). C3G susceptibility was associated with group B2 (57% versus 0%; *P*, 0.003), whereas C3G resistance (six strains with overexpression of chromosome-mediated cephalosporinases and one with an extended-spectrum β-lactamase, CTX-M-14) was associated with group A (86% versus 15%; *P*, <0.001).

These results suggest that resistance to quinolones is associated with the less virulent phylogenetic groups and with a weak prevalence of virulence factors, as previously observed (24, 26, 52, 67, 68). We obtained similar results for resistance against cotrimoxazole, amoxicillin, and C3G. The acquisition of antibiotic resistance by horizontal gene transfers or mutations may therefore require a particular genetic background, as observed for virulence factors (15). Some studies have suggested that the low-virulence group A isolates are more exposed to antibiotic selection pressure within the intestinal tract (24, 26), which indicates the possible importance of environmental factors.

**Conclusion.** The distribution of CM-encoding genes, including the recently described *pks* genomic island, and the functionality of these toxins were investigated in *E. coli* strains in relation to their phylogenetic background, clinical origin, and antibiotic resistance. One finding to emerge from the present study was the frequent association of the *pks* island with the

*cnfI* gene and the alpha-hemolytic phenotype, and their presence in amoxicillin-, cotrimoxazole-, and quinolone-susceptible *E. coli* strains of the B2 phylogenetic background isolated from patients with urosepsis. The widespread diffusion of the *pks* island and the *cnfI* gene in *E. coli* help to distinguish ExPEC from commensal strains and reinforce the idea that these genes of the *E. coli* flexible gene pool are involved in pathogenicity and/or in the ability to survive in new ecological niches, such as the human urinary tract.

#### ACKNOWLEDGMENTS

This work was supported by le Ministère Français de l'Éducation Nationale, de la Recherche et de la Technologie (grant JE2526) and by l'Institut National de la Recherche Agronomique (grant USC-2018).

We thank Rolande Perroux and Marlène Jan for helpful technical assistance.

#### REFERENCES

- Adamus-Bialek, W., A. Wojtasik, M. Majchrzak, M. Sosnowski, and P. Parniewski. 2009. (CGG)4-based PCR as a novel tool for discrimination of uropathogenic *Escherichia coli* strains: comparison with enterobacterial repetitive intergenic consensus-PCR. *J. Clin. Microbiol.* **47**:3937–3944.
- Albert, M. J., S. M. Faruque, A. S. Faruque, K. A. Bettelheim, P. K. Neogi, N. A. Bhuiyan, and J. B. Kaper. 1996. Controlled study of cytolethal distending toxin-producing *Escherichia coli* infections in Bangladeshi children. *J. Clin. Microbiol.* **34**:717–719.
- Bidet, P., A. Metais, F. Mahjoub-Messai, L. Durand, M. Dehem, Y. Aujard, E. Bingen, X. Nassif, and S. Bonacorsi. 2007. Detection and identification by PCR of a highly virulent phylogenetic subgroup among extraintestinal pathogenic *Escherichia coli* B2 strains. *Appl. Environ. Microbiol.* **73**:2373–2377.
- Bielaszewska, M., M. Fell, L. Greune, R. Prager, A. Fruth, H. Tschape, M. A. Schmidt, and H. Karch. 2004. Characterization of cytolethal distending toxin genes and expression in Shiga toxin-producing *Escherichia coli* strains of non-O157 serogroups. *Infect. Immun.* **72**:1812–1816.
- Bingen-Bidois, M., O. Clermont, S. Bonacorsi, M. Terki, N. Brahimi, C. Loukil, D. Barraud, and E. Bingen. 2002. Phylogenetic analysis and prevalence of urosepsis strains of *Escherichia coli* bearing pathogenicity island-like domains. *Infect. Immun.* **70**:3216–3226.
- Blum, G., V. Falbo, A. Caprioli, and J. Hacker. 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. *FEMS Microbiol. Lett.* **126**:189–195.
- CA-SFM. 2007. Comité de l'Antibiogramme de la Société Française de Microbiologie, communiqué 2007. CA-SFM, Paris, France.
- Chassagne, L., N. Pradel, F. Robin, V. Livrelli, R. Bonnet, and J. Delmas.



2009. Detection of *stx1*, *stx2*, and *eae* genes of enterohemorrhagic *Escherichia coli* using SYBR Green in a real-time polymerase chain reaction. *Diagn. Microbiol. Infect. Dis.* **64**:98–101.
9. Clark, C. G., S. T. Johnson, R. H. Easy, J. L. Campbell, and F. G. Rodgers. 2002. PCR for detection of *cdt-III* and the relative frequencies of cytotolethal distending toxin variant-producing *Escherichia coli* isolates from humans and cattle. *J. Clin. Microbiol.* **40**:2671–2674.
  10. Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **66**:4555–4558.
  11. Davis, J. M., H. M. Carvalho, S. B. Rasmussen, and A. D. O'Brien. 2006. Cytotoxic necrotizing factor type 1 delivered by outer membrane vesicles of uropathogenic *Escherichia coli* attenuates polymorphonuclear leukocyte antimicrobial activity and chemotaxis. *Infect. Immun.* **74**:4401–4408.
  12. De Rycke, J., A. Milon, and E. Oswald. 1999. Necrotizing *Escherichia coli* (NTEC): two emerging categories of human and animal pathogens. *Vet. Res.* **30**:221–233.
  13. Dozois, C. M., S. Clement, C. Desautels, E. Oswald, and J. M. Fairbrother. 1997. Expression of P, S, and F1C adhesins by cytotoxic necrotizing factor 1-producing *Escherichia coli* from septicemic and diarrheic pigs. *FEMS Microbiol. Lett.* **152**:307–312.
  14. Duriez, P., O. Clermont, S. Bonacorsi, E. Bingen, A. Chaventre, J. Elion, B. Picard, and E. Denamur. 2001. Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology* **147**:1671–1676.
  15. Escobar-Páramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur. 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol. Biol. Evol.* **21**:1085–1094.
  16. Escobar-Páramo, P., K. Grenet, A. Le Menac'h, L. Rode, E. Salgado, C. Amorin, S. Gouriou, B. Picard, M. C. Rahimy, A. Andreumont, E. Denamur, and R. Ruimy. 2004. Large-scale population structure of human commensal *Escherichia coli* isolates. *Appl. Environ. Microbiol.* **70**:5698–5700.
  17. Escobar-Páramo, P., A. Le Menac'h, T. Le Gall, C. Amorin, S. Gouriou, B. Picard, D. Skurnik, and E. Denamur. 2006. Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ. Microbiol.* **8**:1975–1984.
  18. Gauduchon, V., L. Chalabreysse, J. Etienne, M. Celard, Y. Benito, H. Lepidi, F. Thivolet-Bejui, and F. Vandenesch. 2003. Molecular diagnosis of infective endocarditis by PCR amplification and direct sequencing of DNA from valve tissue. *J. Clin. Microbiol.* **41**:763–766.
  19. Ge, Z., D. B. Schauer, and J. G. Fox. 2008. In vivo virulence properties of bacterial cytolethal-distending toxin. *Cell. Microbiol.* **10**:1599–1607.
  20. Gordon, D. M., O. Clermont, H. Tolley, and E. Denamur. 2008. Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. *Environ. Microbiol.* **10**:2484–2496.
  21. Guyer, D. M., J. S. Kao, and H. L. Mobley. 1998. Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. *Infect. Immun.* **66**:4411–4417.
  22. Hacker, J., U. Hentschel, and U. Dobrindt. 2003. Prokaryotic chromosomes and disease. *Science* **301**:790–793.
  23. Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J. Bacteriol.* **172**:6175–6181.
  24. Houdouin, V., S. Bonacorsi, P. Bidet, M. Bingen-Bidois, D. Barraud, and E. Bingen. 2006. Phylogenetic background and carriage of pathogenicity island-like domains in relation to antibiotic resistance profiles among *Escherichia coli* urosepsis isolates. *J. Antimicrob. Chemother.* **58**:748–751.
  25. Janka, A., M. Bielaszewska, U. Dobrindt, L. Greune, M. A. Schmidt, and H. Karch. 2003. Cytolethal distending toxin gene cluster in enterohemorrhagic *Escherichia coli* O157:H– and O157:H7: characterization and evolutionary considerations. *Infect. Immun.* **71**:3634–3638.
  26. Jauréguy, F., E. Carboneille, S. Bonacorsi, C. Clec'h, P. Casassus, E. Bingen, B. Picard, X. Nassif, and O. Lortholary. 2007. Host and bacterial determinants of initial severity and outcome of *Escherichia coli* sepsis. *Clin. Microbiol. Infect.* **13**:854–862.
  27. Johnson, J. R., B. Johnston, M. A. Kuskowski, J. P. Nougayrède, and E. Oswald. 2008. Molecular epidemiology and phylogenetic distribution of the *Escherichia coli* *pks* genomic island. *J. Clin. Microbiol.* **46**:3906–3911.
  28. Johnson, J. R., M. A. Kuskowski, A. Gajewski, S. Soto, J. P. Horcajada, M. T. Jimenez de Anta, and J. Vila. 2005. Extended virulence genotypes and phylogenetic background of *Escherichia coli* isolates from patients with cystitis, pyelonephritis, or prostatitis. *J. Infect. Dis.* **191**:46–50.
  29. Johnson, J. R., K. Owens, A. Gajewski, and M. A. Kuskowski. 2005. Bacterial characteristics in relation to clinical source of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women. *J. Clin. Microbiol.* **43**:6064–6072.
  30. Johnson, J. R., F. Scheutzel, P. Ulleryd, M. A. Kuskowski, T. T. O'Bryan, and T. Sandberg. 2005. Host-pathogen relationships among *Escherichia coli* isolates recovered from men with febrile urinary tract infection. *Clin. Infect. Dis.* **40**:813–822.
  31. Johnson, J. R., and A. L. Stell. 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* **181**:261–272.
  32. Kadhum, H. J., D. Finlay, M. T. Rowe, I. G. Wilson, and H. J. Ball. 2008. Occurrence and characteristics of cytotoxic necrotizing factors, cytolethal distending toxins and other virulence factors in *Escherichia coli* from human blood and faecal samples. *Epidemiol. Infect.* **136**:752–760.
  33. Kaper, J. B., J. P. Nataro, and H. L. Mobley. 2004. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2**:123–140.
  34. Lemonnier, M., L. Landraud, and E. Lemichez. 2007. Rho GTPase-activating bacterial toxins: from bacterial virulence regulation to eukaryotic cell biology. *FEMS Microbiol. Rev.* **31**:515–534.
  35. Lidin-Janson, G., B. Kaijser, K. Lincoln, S. Olling, and H. Wedel. 1978. The homogeneity of the faecal coliform flora of normal school-girls, characterized by serological and biochemical properties. *Med. Microbiol. Immunol.* **164**:247–253.
  36. Loukiadis, E., R. Nobe, S. Herold, C. Tramuta, Y. Ogura, T. Ooka, S. Morabito, M. Kerouredan, H. Brugere, H. Schmidt, T. Hayashi, and E. Oswald. 2008. Distribution, functional expression, and genetic organization of Cif, a phage-encoded type III-secreted effector from enteropathogenic and enterohemorrhagic *Escherichia coli*. *J. Bacteriol.* **190**:275–285.
  37. Mainil, J. G., E. Jacquemin, and E. Oswald. 2003. Prevalence and identity of *cdt*-related sequences in necrotizing *Escherichia coli*. *Vet. Microbiol.* **94**:159–165.
  38. Marchès, O., T. N. Ledger, M. Boury, M. Ohara, X. Tu, F. Goffaux, J. Mainil, I. Rosenshine, M. Sugai, J. De Rycke, and E. Oswald. 2003. Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks cell cycle G<sub>2</sub>/M transition. *Mol. Microbiol.* **50**:1553–1567.
  39. Middendorf, B., B. Hochhut, K. Leipold, U. Dobrindt, G. Blum-Oehler, and J. Hacker. 2004. Instability of pathogenicity islands in uropathogenic *Escherichia coli* 536. *J. Bacteriol.* **186**:3086–3096.
  40. Munro, P., G. Flatau, A. Doye, L. Boyer, O. Oregioni, J. L. Mege, L. Landraud, and E. Lemichez. 2004. Activation and proteasomal degradation of rho GTPases by cytotoxic necrotizing factor-1 elicit a controlled inflammatory response. *J. Biol. Chem.* **279**:35849–35857.
  41. Nougayrède, J. P., S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, G. Gottschalk, C. Buchrieser, J. Hacker, U. Dobrindt, and E. Oswald. 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313**:848–851.
  42. Nowrouzian, F. L., I. Adlerberth, and A. E. Wold. 2006. Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. *Microbes Infect.* **8**:834–840.
  43. Nowrouzian, F. L., A. E. Wold, and I. Adlerberth. 2005. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J. Infect. Dis.* **191**:1078–1083.
  44. Ochman, H., and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* **157**:690–693.
  45. Orden, J. A., G. Dominguez-Bernal, S. Martinez-Pulgarin, M. Blanco, J. E. Blanco, A. Mora, J. Blanco, and R. de la Fuente. 2007. Necrotizing *Escherichia coli* from sheep and goats produce a new type of cytotoxic necrotizing factor (CNF3) associated with the *eae* and *ehxA* genes. *Int. Microbiol.* **10**:47–55.
  46. Orth, D., K. Grif, M. P. Dierich, and R. Wurzner. 2006. Cytolethal distending toxins in Shiga toxin-producing *Escherichia coli*: alleles, serotype distribution and biological effects. *J. Med. Microbiol.* **55**:1487–1492.
  47. Oswald, E., J. de Rycke, P. Lintermans, K. van Muylem, J. Mainil, G. Daube, and P. Pohl. 1991. Virulence factors associated with cytotoxic necrotizing factor type two in bovine diarrheic and septicemic strains of *Escherichia coli*. *J. Clin. Microbiol.* **29**:2522–2527.
  48. Oswald, E., J. P. Nougayrède, F. Taieb, and M. Sugai. 2005. Bacterial toxins that modulate host cell-cycle progression. *Curr. Opin. Microbiol.* **8**:83–91.
  49. Pandey, M., A. Khan, S. C. Das, B. Sarkar, S. Kahali, S. Chakraborty, S. Chattopadhyay, S. Yamasaki, Y. Takeda, G. B. Nair, and T. Ramamurthy. 2003. Association of cytolethal distending toxin locus *cdtB* with enteropathogenic *Escherichia coli* isolated from patients with acute diarrhea in Calcutta, India. *J. Clin. Microbiol.* **41**:5277–5281.
  50. Pèrès, S. Y., O. Marchès, F. Daigle, J. P. Nougayrède, F. Herault, C. Tasca, J. De Rycke, and E. Oswald. 1997. A new cytolethal distending toxin (CDT) from *Escherichia coli* producing CNF2 blocks HeLa cell division in G<sub>2</sub>/M phase. *Mol. Microbiol.* **24**:1095–1107.
  51. Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grobeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamowski, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**:529–533.
  52. Piatti, G., A. Mannini, M. Balistreri, and A. M. Schito. 2008. Virulence

- factors in urinary *Escherichia coli* strains: phylogenetic background and quinolone and fluoroquinolone resistance. *J. Clin. Microbiol.* **46**:480–487.
53. Picard, B., J. S. Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion, and E. Denamur. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* **67**:546–553.
  54. Pickett, C. L., D. L. Cottle, E. C. Pesci, and G. Bikah. 1994. Cloning, sequencing, and expression of the *Escherichia coli* cytolethal distending toxin genes. *Infect. Immun.* **62**:1046–1051.
  55. Pickett, C. L., R. B. Lee, A. Eyigor, B. Elitzur, E. M. Fox, and N. A. Strockbine. 2004. Patterns of variations in *Escherichia coli* strains that produce cytolethal distending toxin. *Infect. Immun.* **72**:684–690.
  56. Putze, J., C. Hennequin, J. P. Nougayrède, W. Zhang, S. Homburg, H. Karch, M. A. Bringer, C. Fayolle, E. Carniel, W. Rabsch, T. A. Oelschlaeger, E. Oswald, C. Forestier, J. Hacker, and U. Dobrindt. 2009. Genetic structure and distribution of the colibactin genomic island among *Enterobacteriaceae*. *Infect. Immun.* **77**:4696–4703.
  57. Real, J. M., P. Munro, C. Buisson-Touati, E. Lemichez, P. Boquet, and L. Landraud. 2007. Specificity of immunomodulator secretion in urinary samples in response to infection by alpha-hemolysin and CNF1 bearing uropathogenic *Escherichia coli*. *Cytokine* **37**:22–25.
  58. Relman, D. A. 1993. The identification of uncultured microbial pathogens. *J. Infect. Dis.* **168**:1–8.
  59. Rippere-Lampe, K. E., M. Lang, H. Ceri, M. Olson, H. A. Lockman, and A. D. O'Brien. 2001. Cytotoxic necrotizing factor type 1-positive *Escherichia coli* causes increased inflammation and tissue damage to the prostate in a rat prostatitis model. *Infect. Immun.* **69**:6515–6519.
  60. Rippere-Lampe, K. E., A. D. O'Brien, R. Conran, and H. A. Lockman. 2001. Mutation of the gene encoding cytotoxic necrotizing factor type 1 (*cnf*<sub>1</sub>) attenuates the virulence of uropathogenic *Escherichia coli*. *Infect. Immun.* **69**:3954–3964.
  61. Silver, R. P., W. Aaronson, A. Sutton, and R. Schneerson. 1980. Comparative analysis of plasmids and some metabolic characteristics of *Escherichia coli* K1 from diseased and healthy individuals. *Infect. Immun.* **29**:200–206.
  62. Skurnik, D., D. Bonnet, C. Bernede-Bauduin, R. Michel, C. Guette, J. M. Becker, C. Balaire, F. Chau, J. Mohler, V. Jarlier, J. P. Boutin, B. Moreau, D. Guillemot, E. Denamur, A. Andreumont, and R. Ruimy. 2008. Characteristics of human intestinal *Escherichia coli* with changing environments. *Environ. Microbiol.* **10**:2132–2137.
  63. Tavechio, A. T., L. R. Marques, C. M. Abe, and T. A. Gomes. 2004. Detection of cytotoxic necrotizing factor types 1 and 2 among fecal *Escherichia coli* isolates from Brazilian children with and without diarrhea. *Mem. Inst. Oswaldo Cruz* **99**:81–83.
  64. Tóth, I., F. Hérault, L. Beutin, and E. Oswald. 2003. Production of cytolethal distending toxins by pathogenic *Escherichia coli* strains isolated from human and animal sources: establishment of the existence of a new *cdt* variant (type IV). *J. Clin. Microbiol.* **41**:4285–4291.
  65. Tóth, I., J. P. Nougayrède, U. Dobrindt, T. N. Ledger, M. Boury, S. Morabito, T. Fujiwara, M. Sugai, J. Hacker, and E. Oswald. 2009. Cytolethal distending toxin type I and type IV genes are framed with lambdoid prophage genes in extraintestinal pathogenic *Escherichia coli*. *Infect. Immun.* **77**:492–500.
  66. Van Bost, S., E. Jacquemin, E. Oswald, and J. Mainil. 2003. Multiplex PCRs for identification of necrotogenic *Escherichia coli*. *J. Clin. Microbiol.* **41**:4480–4482.
  67. Velasco, M., J. P. Horcajada, J. Mensa, A. Moreno-Martinez, J. Vila, J. A. Martinez, J. Ruiz, M. Barranco, G. Roig, and E. Soriano. 2001. Decreased invasive capacity of quinolone-resistant *Escherichia coli* in patients with urinary tract infections. *Clin. Infect. Dis.* **33**:1682–1686.
  68. Vila, J., K. Simon, J. Ruiz, J. P. Horcajada, M. Velasco, M. Barranco, A. Moreno, and J. Mensa. 2002. Are quinolone-resistant uropathogenic *Escherichia coli* less virulent? *J. Infect. Dis.* **186**:1039–1042.
  69. Wiles, T. J., R. R. Kulesus, and M. A. Mulvey. 2008. Origins and virulence mechanisms of uropathogenic *Escherichia coli*. *Exp. Mol. Pathol.* **85**:11–19.
  70. Zdziarski, J., C. Svanborg, B. Wullt, J. Hacker, and U. Dobrindt. 2008. Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? *Infect. Immun.* **76**:695–703.
  71. Zhang, L., B. Foxman, and C. Marrs. 2002. Both urinary and rectal *Escherichia coli* isolates are dominated by strains of phylogenetic group B2. *J. Clin. Microbiol.* **40**:3951–3955.

### **III.Article 3 - Les cyclomodulines des *E. coli* de tumeurs**

## **Prévalence des cyclomodulines de *Escherichia coli* isolées de tumeurs colo-rectales, de diverticuloses et de lésions iléales de maladie de Crohn.**

Damien Dubois, Emmanuel Buc, Pierre Sauvanet, Nathalie Pradel, Marie-Agnès Bringer,  
Arlette Darfeuille-Michaud, Denis Pezet et Richard Bonnet

*A soumettre*

**Prévalence des cyclomodulines de *Escherichia coli*  
isolées de tumeurs colorectales, de diverticuloses  
et de lésions iléales de maladie de Crohn**

Damien Dubois<sup>a,c</sup>, Emmanuel Buc<sup>b,c</sup>, Pierre Sauvanet<sup>c</sup>, Nathalie Pradel<sup>c</sup>, Marie-Agnès Bringer<sup>c</sup>, Arlette Darfeuille-Michaud<sup>c</sup>, Denis Pezet<sup>b,c</sup> et Richard Bonnet<sup>a,c</sup>

<sup>a</sup>CHU Clermont-Ferrand, Hôpital G. Montpied, Centre de Biologie, Laboratoire de bactériologie clinique, Clermont-Ferrand, F-63003, France.

<sup>b</sup>CHU Clermont-Ferrand, Hôpital Estaing, Service de chirurgie générale et digestive, Clermont-Ferrand, F-63003, France.

<sup>c</sup>Clermont Université, Université d'Auvergne, Evolution des bactéries pathogènes et susceptibilité de l'hôte, JE 2526, USC-INRA 2018, Clermont-Ferrand, F-63001, France.

## Résumé

De nombreuses études ont évalué l'association entre le cancer colorectal et des agents infectieux. Bien que l'implication des cyclomodulines de *E. coli* ait été suggérée, aucune étude n'a abordé la colonisation du côlon par des *E. coli* produisant ces toxines chez les patients atteints de cancer colorectal.

La diversité clonale quantitative et phylogénétique de *E. coli* et la prévalence des gènes des cyclomodulines et leur fonctionnalité ainsi que la prévalence de l'alpha-hémolysine de *E. coli* ont été investiguées dans des isolats issus de biopsies de tumeurs coliques (n=39) et rectales (n=38), et de biopsies de diverticuloses (contrôles, n=30). De plus, la prévalence de ces facteurs de virulence a été déterminée chez des isolats de biopsies iléales de maladie de Crohn. Ces prévalences ont été comparées entre elles et à celle d'isolats commensaux précédemment rapportées.

Dix isolats de chaque biopsie ont été typés par deux techniques génomiques. Chaque clone déterminé a été ensuite étudié. Le groupe phylogénétique de chaque clone a été déterminé par PCR. Les gènes des cyclomodulines, c'est à dire, l'îlot *pks* codant la colibactine et les gènes *cdtB*, *cnf1-3* et *cif* ont été détectés par hybridation d'ADN dot-blot et confirmés ou typés par PCR. La fonctionnalité de ces cyclomodulines a été évaluée sur cultures cellulaires. Les conclusions majeures de cette étude sont :

- (1) la prévalence d'un clone de *E. coli* du phylogroupe B2 significativement supérieure dans les biopsies de tumeurs coliques comparé aux autres origines cliniques ( $p$ , <0,001 à 0,015), chez des patients de même origine géographique,
- (2) la prévalence de l'îlot *pks* significativement supérieure dans la série de tumeurs coliques comparé aux autres origines cliniques ( $p$ , <0,001 à 0,009),
- (3) la prévalence de l'opéron *hly-cnf1* significativement supérieure dans les tumeurs coliques et rectales que dans les fèces ( $p$ <0,001 et  $p$ =0,019, respectivement),
- (4) la prévalence de l'alpha-hémolysine significativement supérieure dans les tumeurs coliques que dans les fèces, les diverticuloses et la MC ( $p$ <0,001,  $p$ =0,019 et  $p$ =0,042, respectivement),
- (5) la prévalence du gène *cdtB* significativement plus élevée dans les tumeurs coliques que dans les diverticules ( $p$ =0,030),
- (6) la présence de trois *E. coli* EntéroPathogènes présentant un gène *cif* fonctionnel isolés de trois tumeurs coliques.

La majorité de ces facteurs de virulence étaient fonctionnels.

Notre étude épidémiologique montre que les différents facteurs de virulence de *E. coli* investigués pourraient être impliqués dans la tumorigenèse colorectale.



## Introduction

*Escherichia coli* est une espèce d'Entérobactérie diversifiée et versatile. Des études de génétique des populations reposant principalement sur l'analyse du core-génome ont permis de classer les souches de cette espèce en quatre groupes phylogénétiques majeurs A, B1, B2 et D [1]. Ces groupes sont associés à la capacité de ces souches à coloniser des niches écologiques et ont des capacités différentes à induire une pathologie. Aussi, les souches de *E. coli* peuvent également être classées selon leur comportement vis-à-vis de l'hôte. On peut distinguer les souches commensales non pathogènes, les souches pathogènes intestinales et les souches pathogènes extra-intestinales.

Les souches de *E. coli* non pathogènes et son hôte humain coexistent pendant des décennies avec un bénéfice mutuel. La colonisation de l'hôte se fait les heures qui suivent la naissance et leur niche préférentielle est la couche muqueuse du côlon. *E. coli* est un micro-organisme très compétitif sur ce site très peuplé et constitue la bactérie anaérobie facultative la plus abondante de la microflore intestinale humaine. Ces souches de *E. coli* commensales entraînent rarement une maladie, sauf chez les sujets immunodéprimés ou lorsque la barrière digestive normale est altérée, comme dans la péritonite [2].

Parmi les pathogènes intestinaux, six pathovars sont bien caractérisés et responsables de diarrhées : les *E. coli* entéropathogènes (EPEC), les *E. coli* entérohémorragiques (EHEC), les *E. coli* entérotoxiques (ETEC), les *E. coli* entéroaggrégatifs (EAEC), les *E. coli* entéroinvasifs (EIEC) et les *E. coli* d'adhésion diffuse (DAEC). Un ensemble hétérogène de souches appelées *E. coli* adhérents et invasifs (AIEC) [3] est associé à la maladie de Crohn (MC). Elles se répliquent dans les macrophages et induisent la libération de grandes quantités de TNF- $\alpha$ , une caractéristique qui pourrait contribuer à l'inflammation intestinale chronique observée dans la MC. Les *E. coli* pathogènes extra-intestinales (ExPEC) constituent un ensemble de souches encore mal défini qui comprend notamment les *E. coli* uropathogènes (UPEC) et les *E. coli* associés aux méningites néonatales (MNEC) [2].

Ces souches de *E. coli* pathogènes se distinguent principalement par l'acquisition de gènes impliqués dans la virulence par transfert horizontal *via* des plasmides, des bactériophages et/ou des îlots génomiques. Ce pool de gènes flexible de *E. coli* représente au moins la moitié du nombre total de gènes présents dans chaque génome, soit 2 000 à 3 400 gènes, et environ 88% des gènes recensés chez *E. coli* à ce jour [4]. Les facteurs de virulence codés par ces gènes du pool flexible peuvent affecter une grande variété de processus cellulaires critiques de l'hôte, comme la synthèse des protéines, les signaux de transduction, les fonctions du cytosquelette, la sécrétion d'ions, la transcription, l'apoptose, la

fonction mitochondriale et la division cellulaire [2]. Les protéines effectrices qui perturbent ces processus sont accompagnées par de nombreux facteurs de colonisation et de « fitness » qui permettent aux bactéries d'adhérer et/ou d'envahir les cellules hôtes. Parmi ces effecteurs, il a été observé une classe croissante de toxines appelées cyclomodulines (CM) [5]. Elles agissent sur le cycle cellulaire eucaryote, un processus fondamental pour l'homéostasie cellulaire et tissulaire, et ainsi induisent des effets cytopathiques qui peuvent être mis à profit pour leur détection [6, 7]. Quatre types de CM ont été identifiées puis caractérisées chez *E. coli* : les toxines protéiques CNF-1 à CNF-3 (facteurs nécrosants cytotoxiques) ciblant les Rho GTPases [8, 9], le facteur d'inhibition du cycle (protéine Cif) ciblant les complexes ubiquitine ligases cullin-RING [10, 11], et deux types de génotoxines, les toxines de distension cytoléthales (protéines CDT de cinq types, I à V) [12] et la colibactine, de découverte récente [6]. La colibactine est probablement une toxine de type polycétide-peptide non ribosomal hybride, dont l'activité est codée par l'îlot génomique *pks* [6]. Les toxines CDT, Cif et la colibactine bloquent la mitose, alors les toxines CNF favorisent la réplication de l'ADN sans cytocinèse. Ces toxines peuvent affecter le système immunitaire de l'hôte mais aussi le développement de cancer.

Actuellement, le cancer colorectal est le troisième cancer le plus fréquent, chez l'homme et la femme, et la quatrième cause de décès par cancer à travers le monde, avec 1,2 million de cas estimés et 609.000 décès estimés en 2008 [13]. Basé sur les tendances démographiques, l'incidence annuelle devrait augmenter de près de 80% à 2,2 millions de cas au cours des deux prochaines décennies et la plupart de cette augmentation se produira dans les régions les moins développées du monde (62%), le taux de ce cancer augmentant avec l'industrialisation et l'urbanisation [14]. Dans les pays industrialisés, le cancer du côlon est deux fois plus fréquent que celui du rectum. Environ 70% des patients atteints de cancer du côlon ont plus de 65 ans. Le cancer colorectal survient le plus souvent de façon sporadique et il est héréditaire ou familial dans seulement 5% des cas [15]. Les habitudes alimentaires, le mode de vie et d'autres facteurs non génétiques tels que le tabac, l'utilisation chronique d'anti-inflammatoires non-stéroïdiens sont considérés comme ayant un fort impact sur le risque de cancer colorectal. De même, certaines conditions physiopathologiques telles que le syndrome métabolique et les maladies inflammatoires de l'intestin (la MC et la rectocolite ulcéro-hémorragique) augmentent le risque de cancer du côlon [13, 16, 17].

De nombreuses études ont évalué l'association entre le cancer colorectal et les agents infectieux [18]. Des études utilisant des modèles animaux suggèrent que les micro-organismes sont impliqués dans le développement du cancer colorectal [19]. Bien que l'implication des CM de *E. coli* ait été suggérée [20-22], aucune étude n'a abordé la colonisation de *E. coli* produisant ces toxines chez les patients atteints de cancer colorectal. Afin d'étudier une possible association entre la production de CM de *E. coli* et le

développement du cancer colorectal, nous avons exploré la prévalence des gènes et l'expression fonctionnelle des CM chez des *E. coli* isolées de tumeurs colorectales, de diverticuloses humaines et de muqueuses iléales de patients atteints de MC.

## Matériel et Méthodes

### Souches bactériennes de référence

Les souches de *E. coli* de référence archétypes des facteurs de virulence étudiés et la souche *E. coli* K-12 DH5 $\alpha$  utilisés comme contrôles positifs et négatif respectivement sont présentées dans le **tableau 1** [23].

**Tableau 1.** Souches de *E. coli* archétypes utilisées dans cette étude.

Souche de <i>E. coli</i>	Caractéristique génétique	Origine	Source ou référence
28C	<i>hly, cnf1, cdt-IV, clbA, clbK, clbJ, clbQ</i>	Septicémie porcine	[24]
1404	<i>cnf2, cdt-III</i>	Septicémie bovine	[25]
IHE3034 $\Delta clbP$	<i>cdt-I, clb \Delta clbP</i>	Méningite de nouveau-né	[6]
DH5 $\alpha$ (pCP2123)	<i>cdt-II</i>	Souche de laboratoire	[26]
C48a	<i>cnf3, eae, cif</i>	Agneau sain	[9]
DH10 $\beta$ (pBACpks)	<i>clb</i>	Souche de laboratoire	[6]
E22	<i>cif, eae</i>	Lapin, EPEC	[10]
RS218	<i>chuA, yjaA, svg, TSPE4.C2, uidA</i> (Phylogroupe B2.1)	Méningite de nouveau-né	[27]
EDL933	<i>eae, stx1, stx2</i>	EHEC	[28]
DH5 $\alpha$		Souche de laboratoire	Novagen

### Patients

Les souches de *E. coli* incluses dans cette étude ont été isolées à partir d'échantillons de muqueuses de patients du Centre Hospitalier Universitaire de Clermont-Ferrand. Entre mars 2007 et novembre 2009, 77 échantillons ont été prélevés au niveau de la tumeur de patients présentant un adénocarcinome colorectal (CCR) et 30 échantillons ont été prélevés au niveau de diverticuloses. Des souches issues de lésions de l'iléon de patients atteints de MC ont également été incluses [29, 30] ainsi que 51 souches de *E. coli* commensales [23]. Tous les patients ont donné leur consentement éclairé et l'approbation du comité d'éthique local a été obtenue.

Soixante-dix-sept échantillons d'adénocarcinomes colorectaux ont été prélevés lors d'une intervention chirurgicale pour exérèse. L'âge médian des patients était de 69 ans (extrêmes : 22 à 95 ans) ; le sexe ratio de la cohorte était de 1,26 (**tableau 2**). La cohorte inclut 39 échantillons de cancers coliques (21 échantillons du côlon droit et 18 échantillons du colon gauche) et 38 échantillons de cancers rectaux. Tous les patients ont reçu de la céfoxitine (2 g par voie intraveineuse) au moment de l'incision.

Trente échantillons de diverticuloses ont été prélevés lors d'une intervention chirurgicale pour exérèse. L'âge médian des patients était de 58 ans (extrêmes : 34 à 81 ans) ; le sexe ratio de la cohorte était de 0,72 (**tableau 2**). Tous les patients ont reçu de la céfoxitine (2 g par voie intraveineuse) au moment de l'incision.

**Tableau 2.** Données démographiques des patients ayant eu une biopsie de tumeur ou de diverticulose.

n=	TR 38	TDG 39	TDGR 77	Divert. 30
âge médian	64,5	71	69	58
âge moyen	65	67	66	60
écarts d'âge	22-87	35-95	22-95	34-81
hommes	23	20	43	13
femmes	15	19	34	18
sexe ratio	1,53	1,05	1,26	0,72

TR, groupe tumeurs du rectum ; TDG, groupe tumeurs du côlon ; TDGR, groupe tumeurs du colon et du rectum ; Divert., groupe diverticuloses

Les résections iléales de lésions chroniques de 9 patients présentant une MC (MCC) qui avaient subi une iléocoléctomie pour la première fois ont été incluses. Les échantillons ont été prélevés au niveau de la muqueuse de l'iléon réséqué avec une pince endoscopique [29, 30]. Les biopsies iléales de lésions récurrentes précoces de MC (MCP) ont été obtenues à partir de l'iléon néoterminal (à partir de 10 cm de l'anastomose) par endoscopie chez 11 patients qui avaient subi une iléocoléctomie avec anastomose iléo-colique de bout en bout. Aucun des patients n'avait reçu d'antibiotiques dans les 4 semaines précédant le prélèvement [29, 30].

Les isolats fécaux commensaux avaient été recueillis par écouvillonnages anaux d'adultes dans la communauté [23]. Ces personnes ne présentaient aucun signe d'infection aiguë ou de troubles bénins ou malins du tractus gastro-intestinal et n'avaient pas reçu d'antibiotiques dans le mois précédant le prélèvement. Une colonie de *E. coli* choisie arbitrairement par échantillon avait été analysée. L'âge médian des 51 patients était de 46,5 ans (extrêmes : 23 à 62 ans) ; le sexe ratio de la cohorte était de 0,54.

## **Traitement des biopsies**

Les biopsies ont été prélevées dans un flacon stérile contenant une solution saline tamponnée au phosphate (PBS). Chaque échantillon a été divisé en deux parties et pesé (environ 5 à 10 mg chacune).

La première partie des échantillons a été lavée en PBS, puis broyée (Ultra-Turrax, IKA) et incubée 15 minutes en présence de Triton 0,1X. Des dilutions de raison 10 du lysat ont ensuite été réalisées et étalées sur milieu gélosé sélectif pour entérobactéries Drigalski (bioMérieux) et sur milieu gélosé chromogénique chromID® CPS3 (bioMérieux). Les milieux ont été incubés à 37°C pendant 48 heures en atmosphère aérobie. La seconde partie des échantillons a été incubée une heure à 37°C en milieu EMEM (PAA) additionnée de 10µg/mL de gentamicine (Euromedex) afin d'éliminer les bactéries extracellulaires adhérentes, puis broyée, diluée et ensemencée de manière similaire à la première partie.

Les colonies isolées ont été repiquées en plaques 96 puits en milieu Luria Bertani additionné de 15% de glycérol, incubées 24 heures à 37°C puis stockées à -80°C. Les souches de *E. coli* ont été identifiées à l'aide du milieu gélosé chromogénique chromID® CPS3 (bioMérieux, France) et du système automatisé Vitek II® (bioMérieux) en deuxième intention. Au moins 10 colonies par prélèvement ont été retenues pour les études suivantes.

## **Etude de la clonalité des isolats et groupage phylogénétique**

Les expérimentations ont été conduites comme indiqué précédemment [23]. Brièvement, l'ADN des isolats a été extrait par la méthode de l'eau bouillante. Les isolats de *E. coli* ont été classés d'après le système de la collection de référence de *E. coli* ECOR [31] dans les groupes phylogénétiques A, B1, B2 et D en fonction de la présence de trois gènes (*chuA*, *yjaA* et *svg*) et d'un fragment d'ADN (TSPE4.C2), en utilisant une technique de PCR multiplex incluant le gène *uidA* comme contrôle interne [32].

Deux techniques de typage génomique ont été utilisées : une « Enterobacterial Repetitive Intergenic Consensus sequence » (ERIC)-PCR utilisant l'amorce ERIC2 (5'-AAG TAA GTG ACT GGG GTG AGC G-3') et une « Random Amplified Polymorphic DNA » (RAPD) PCR utilisant l'amorce RAPD 1283 (5'-GCG ATC CCC A-3'). Ces amorces ont été utilisées comme indiqué précédemment [33, 34]. Les isolats représentatifs d'un clone ont été stockés à -80°C en milieu Luria-Bertani supplémenté avec 15% de glycérol.

## **Identification des gènes codant les cyclomodulines**

La détection des gènes codant la production de cyclomodulines a été réalisée par hybridation d'ADN selon la méthode du dot-blot. Chaque dépôt a été vérifié avec une sonde du gène codant l'ARNr 16S de *E. coli*. L'ilot *pks*, qui contient le groupe de gènes *clb* permettant la production de la colibactine a été détecté à l'aide d'une sonde chevauchant les gènes *clbK* et *clbJ*. Les 3 types de gènes *cnf* ont été détectés avec un mélange de sondes spécifiques de *cnf1*, *cnf2*, et *cnf3*. Les 5 types de gènes *cdtB* ont été détectés par deux expériences d'hybridation avec une sonde spécifique des gènes *cdtB-II*, *cdtB-III* et *cdtB-V*, et une sonde spécifique des gènes *cdtB-I* et *cdtB-IV*. Le gène *cif* a été détecté en utilisant une sonde interne spécifique. La sensibilité et la spécificité des sondes ont été vérifiées sur chaque membrane en déposant des extraits d'ADN de toutes les souches contrôles.

Les hybridations positives ont été soumises à confirmation par PCR en utilisant des amorces spécifiques des gènes *clbA* et *clbQ* de l'ilot *pks* par PCR multiplex, des gènes *cnf1*, *cnf2* et *cnf3* par PCR multiplex, des gènes *cdtB-I*, *cdtB-II*, *cdtB-III/cdtB-V*, et *cdtB-IV*, et du gène *cif* par PCR simplex, comme précédemment décrit [23].

Les gènes *cdt* et *cif* étant fréquemment associés aux gènes *stx1* et *stx2* des *E coli* entérohémorragiques (EHEC) et *eae* des *E coli* entéro-pathogènes (EPEC) respectivement, tous les isolats ont été testés par PCR pour les gènes de virulence *stx1*, *stx2*, et *eae*, comme décrit précédemment [35].

## **Détection de l'effet cytopathogène induit par les cyclomodulines**

Les effets cytopathiques des CNF et des CDT ont été étudiés dans toutes les souches, et ceux de la colibactine et de Cif ont été étudiés dans les souches non hémolytiques, comme décrit précédemment [6, 7]. Brièvement, les effets des CNF et des CDT ont été détectés par un test d'interaction lysat bactérien-culture cellulaire. Après une culture de 48 h à 37°C sous agitation dans un milieu liquide de Luria-Bertani, les cellules bactériennes ont été soniquées puis filtrées stérilement à l'aide de filtres de 0,22 µm. Des cultures cellulaires HeLa ont été traitées avec les lysats bactériens stériles jusqu'à l'analyse. Les effets de la colibactine et de Cif ont été détectés par un test d'interaction bactéries-culture cellulaire. Des cultures de *E. coli* d'une nuit en bouillon Luria-Bertani ont été diluées dans le milieu d'interaction, et des cultures de cellules HeLa ont été infectées à des multiplicités d'infection (nombre de bactéries par cellule au début de l'infection) de 100 et 200. Les cultures cellulaires ont été lavées 3 à 6 fois 4 heures après infection et ont été incubées dans un milieu de culture cellulaire contenant 200 µg/mL de gentamicine jusqu'à l'analyse. Après 72 h d'incubation à 37 °C sous une atmosphère à 5% de CO<sub>2</sub> le milieu a été

éliminé par plusieurs lavages des monocouches de cellules HeLa. Les modifications morphologiques caractéristiques des CNF, des CDT, de la colibactine et de Cif ont été observées après coloration par le Giemsa, comme décrit par ailleurs [6]. La détection de l'alpha-hémolysine a été effectuée pour toutes les souches étudiées par croissance d'une nuit à 37°C sur milieu gélosé Columbia contenant 5% de sang de mouton (Oxoid).

## **Analyses statistiques**

Un premier test du chi-carré a été réalisé afin d'évaluer l'hétérogénéité entre les groupes, et les comparaisons individuelles par paires ont été testées seulement si ce test donnait une valeur de  $P < 0,05$ . Les analyses statistiques ont été ensuite effectuées en utilisant le test exact de Fisher. Le seuil de signification statistique était  $P < 0,05$ .

## **Résultats et Discussion**

### ***Distribution des gènes codant les CM dans les phylogroupes de E. coli***

La prévalence des gènes codant les CM chez les *E. coli* isolés est hétérogène (**tableau 3**). Les gènes codant les CM les plus fréquents étaient l'îlot *pks* (18,3%) et *cnf1* (13,5%), ainsi le gène codant pour l'alpha-hémolysine (16,9%), trois traits fortement associés au phylogroupe B2.

Toutes les souches possédant l'îlot *pks* appartenaient au phylogroupe B2 ( $P < 0,001$  pour le groupe B2 *versus* les groupes A, B1 et D, combinés ou séparés). 50,6% des isolats B2 possédaient ce trait, qui était observé dans tous les sous-types B2 ; un isolat sur cinq B2.1, deux isolats sur quatre B2 TSPE4.C2-négatif et 35 isolats sur 66 B2 TSPE4.C2-positif. Ces résultats sont en accord avec les études précédentes [6, 23, 36, 37]

Les gènes *cnf* étaient de type 1 dans 13,5% des isolats et de type 2 dans 0,5% des isolats. Le seul isolat positif *cnf2* ne possédait pas le gène *cnf1*. Les souches de *E. coli* présentant *cnf2* et *cnf3* ont été principalement observées chez des souches d'origine animale [9, 38]. Toutes les souches possédant le gène *cnf1* appartenaient au phylogroupe B2, et correspondaient à 37,3% de celles-ci. Comme pour l'îlot *pks*, cette association était forte ( $P, < 0,001$  pour le groupe B2 *versus* les groupes A, B1 et D, combinés ou séparés) et 33,3% des isolats B2 possédaient l'îlot *pks* et le gène *cnf1* ( $P, < 0,001$  pour le groupe B2 *versus* les groupes A, B1 et D, combinés ou séparés), comme précédemment observé [37].

Toutes les souches portant *cnf1* sauf trois présentaient le phénotype alpha-hémolytique. Cette association est probablement due à la présence de l'îlot de pathogénicité

(PAI) II<sub>J96</sub>-like, qui porte le gène *cnf1* en aval de l'opéron *hlyCABD* de l'alpha-hémolysine HlyA [39]. Les trois isolats non hémolytiques portant *cnf1* possédaient néanmoins le gène *hlyC*, laissant supposer une anomalie du PAI II<sub>J96</sub>-like telle que des mutations ponctuelles ou des délétions empêchant la production de l'alpha-hémolysine. Ces deux isolats ont été considérés *cnf1*- et Hly-positifs pour la suite de l'analyse. Seulement 3,4% des isolats présentaient un phénotype alpha-hémolytique sans le gène *cnf1*. Parmi ces isolats, soit 20% des souches alpha-hémolytiques, aucune association n'a été observée de manière significative avec un groupe phylogénétique. Ces isolats possédaient probablement un autre PAI contenant l'opéron *hlyCABD*, analogue au PAI I<sub>CFT073</sub>-like [40].

Les gènes *cdtB* ont été observés dans 4,3% des souches, avec un seul sous-type *cdtB* par souche. Bien que 67% des souches *cdtB*-positives appartenait au phylogroupe B2, aucune association à un groupe phylogénétique n'a été observée de manière significative, même en considérant les différents sous-types de *cdtB*. Les gènes *cdtB-I/cdtB-IV* (n=7) était surreprésentés par rapport au groupe de gènes *cdtB-II/cdtB-III/cdtB-V* (n=2). Les gènes *cdtB-I* et *cdtB-IV* sont deux homologues proches alors que les gènes *cdtB-II*, *cdtB-III* et *cdtB-V* forment un autre sous-groupe [41]. L'unique isolat *cdtB-III*-positif possédait aussi le seul gène *cnf2* observé. Cette association de gènes est rapportée comme fréquente et portée par un plasmide nommé pVir essentiellement présent chez des isolats d'origine bovine [38, 42]. Les gènes *cdtB* ne possédaient pas d'association particulière avec l'îlot *pks* ou le gène *cnf1*. Les gènes *cdt* ayant été largement étudiés dans les souches de *E. coli* produisant des toxines Shiga (STEC) [43-45], les souches *cdtB*-positives ont été étudiées pour les gènes *eae* et *stx* [4, 25, 46, 49, 51-53]. Aucun gène *eae* ou *stx* n'a été détecté dans les isolats positifs *cdtB*.

Le gène *cif* a été détecté dans trois souches. Elles appartenait aux phylogroupes A (n=1) et B1 (n=2) et ne présentaient pas d'autres gènes codant pour les CM. Cif est une protéine effectrice du système de sécrétion de type 3 (SST3) codée par le locus de l'îlot de pathogénicité d'effacement des entérocytes. Le gène *cif* n'a été rapporté que dans les EPEC et EHEC [10, 46]. Les trois isolats de cette étude possédaient le gène *eae* mais pas *stx1* ou *stx2* et appartenait donc au pathotype EPEC.



**Tableau 3.** Distribution des gènes codant les cyclomodulines et l'alpha-hémolysine HlyA en fonction du phylogroupe dans 207 isolats de *E. coli* issus de diverticuloses, de tumeurs colorectales et de lésions iléales de Maladie de Crohn.

Facteurs de virulence	Nombre (%) d'isolats de <i>E. coli</i>					P <sup>a</sup>
	Total	Groupes				
		A	B1	B2	D	
(n= 207)	(n= 57)	(n= 28)	(n= 75)	(n= 47)		
<i>pks</i>	38 (18,3)			38 (50,6)		<0,001
<i>cnf1</i> -HlyA	28 (13,5)			28* (37,3)		<0,001
HlyA non associé à <i>cnf1</i>	7 (3,4)		2 (7,1)	4 (5,3)	1 (2,1)	
<i>cnf1</i> -HlyA, <i>pks</i>	25 (12,1)			25* (33,3)		<0,001
<i>cnf2-cdtB</i> -III	1 (0,5)		1 (3,6)			
<i>cdtB</i>	9 (4,3)	1 (1,7)	2 (7,1)	6 (8)		
<i>cdtB</i> -I	2 (1)			2 (2,6)		
<i>cdtB</i> -IV	5 (2,4)	1 (1,7)		4 (5,3)		
<i>cdtB</i> -V	1 (0,5)		1 (3,6)			
<i>cif</i>	3 (1,5)	1 (1,7)	2 (7,1)			

<sup>a</sup> les valeurs de P (par le test exact de Fischer) sont montrées lorsque P est <0,05.

\* 3 isolats *cnf1*+ non hémolytiques mais présentant le gène *hlyC*

### Détection phénotypique des cyclomodulines

La fonctionnalité d'une majorité de CM et de l'opéron codant l'alpha-hémolysine a été confirmé dans les isolats étudiés (**tableau 4**). Toutes les souches qui ont induit un effet cytopathogène possédaient un ou des gènes codant des CM connues pour induire un effet compatible avec celui observé. L'interprétation des différents phénotypes cytopathogènes liés aux CM ou à leur combinaison était aisée pour tous les isolats, sauf pour une souche possédant l'îlot *pks* et un gène *cdtB*-I, des gènes produisant des CM qui induisent des effets cytopathogènes similaires sur les tests cellulaires réalisés. Ainsi, nous n'avons pas été en mesure d'affirmer que l'îlot *pks* et les gènes *cdt* de cette souche étaient fonctionnels.

Comme précédemment rapporté [23], des isolats *pks*-positifs étaient non fonctionnels. Ils représentaient 31% des isolats possédant l'îlot *pks* et pouvant être explorés (n= 13, isolats non hémolytiques et ne présentant pas *cdtB*). L'absence de production de colibactine peut être due à une altération de l'îlot *pks* ou de gènes situés en dehors de l'îlot et participant à la synthèse de précurseurs ou à la sécrétion de la colibactine.

La souche positive *cnf2* a présenté un effet cytopathogène caractéristique (multinucléation cellulaire) des CNF. Seul un isolat hémolytique provenant d'un diverticule et un isolat positif *cdtB*-IV non hémolytique présentant le gène *hlyC* provenant d'une lésion MCC portaient un gène *cnf1* non fonctionnels, soit 7,1% des isolats positifs *cnf1*. Trois isolats ont présenté un opéron *hlyCABD* non fonctionnel, soit 10,7% des isolats positifs *cnf1*. Parmi

les souches *cdtB*-positives, 22,2% (n=2) n'ont pas présenté d'effet cytopathogène caractéristique. L'absence de cytotoxicité de ces isolats positifs pour *cnf1*, *hly* ou *cdtB* peut avoir pour origine des mutations ou des séquences interrompues par des séquences d'insertion dans les séquences codantes ou régulatrices de ces gènes, des défauts dans le système de sécrétion de HlyA ou l'absence de production des deux autres protéines du complexe hétérotrimérique formant la toxine CDT active, à savoir CdtA et CdtC, nécessaires à l'entrée de CdtB dans la cellule eucaryote [12]. Un effet cytopathogène caractéristique de Cif a été observé pour les trois isolats possédant le gène correspondant.

**Tableau 4.** Détection phénotypique des cyclomodulines et de l'alpha-hémolysine HlyA de 207 isolats de *E. coli* issus d'échantillons de diverticuloses et de tumeurs colorectales et de lésions iléales de Maladie de Crohn.

Facteurs de virulence	Nombre (%) d'isolats de <i>E. coli</i>				
	Total	F	NF	NA	I
<i>pks</i>	38	9 (23,7)	4 (10,5)	24 (63,1)	1 (2,6)
<i>cnf1</i>	28	26 (92,8)	2* (7,1)		
<i>cnf2</i>	1	1 (100)			
<i>hly</i>	35	32 (91,4)	3** (8,6)		
Hly associé à <i>cnf1</i>	28	25 (89,3)	3** (10,7)		
<i>cdtB</i>	9	7 (77,8)	2*** (22,2)		
<i>cif</i>	3	3 (100)			

F, fonctionnel ; NF, non fonctionnel ; NA, non applicable ; I, indéterminé

\* un isolat hémolytique provenant d'un diverticule et un isolat *cdtB*-IV-positif non hémolytique présentant le gène *hlyC* provenant d'une lésion CDC

\*\* trois isolats *cnf1*-positifs mais présentant le gène *hlyC* (MCC, n=2 ; tumeur rectale, n=1)

\*\*\* un isolat *cdtB*-V-positif provenant d'une tumeur rectale et l'isolat *cdtB*-IV-positif non hémolytique présentant le gène *hlyC* provenant d'une lésion CDC

### **Diversité des clones de *E. coli* issus de cancers colorectaux et de diverticuloses**

La diversité des clones de *E. coli* isolés de cancers colorectaux et de diverticuloses a été comparée (**tableau 5**). Le nombre d'échantillons ne présentant pas de *E. coli* était significativement supérieur dans la série diverticuloses comparé à la série tumeurs (19,3% versus 5,2%,  $p=0,031$ ). Ceci suggère que des isolats de *E. coli* seraient capables de coloniser plus facilement les tumeurs colorectales que les diverticules. Il a été rapporté que l'administration de bactéries anaérobies facultatives, comme *Salmonella typhimurium*, *Shigella flexneri* et *E. coli* à des souris porteuses de tumeurs hépatique et splénique conduisait à une accumulation préférentielle et la prolifération des micro-organismes dans les tumeurs solides [47]. De plus, Martin *et al* ont mis en évidence que les isolats de *E. coli* exprimaient des hémagglutinines dans environ 40% des souches issues de MC et de cancers colorectaux, mais seulement dans 4% des contrôles. Cette expression

d'hémagglutinines était significativement corrélée avec l'adhésion à des cultures de cellules intestinales embryonnaires (I407) et de cellules d'adénocarcinome du côlon (HT29) [48]. Swidsinski *et al* ont observé que les isolats de *E. coli* associés aux muqueuses sont fréquemment trouvés dans la muqueuse colique des patients présentant un adénocarcinome colorectal, mais rarement chez les témoins sains. De plus, la muqueuse colique des patients atteints de CCR, mais pas la muqueuse colique normale était colonisée par des *E. coli* intracellulaires [49]. Il est possible que la présence de ces *E. coli* lors de la MC et du cancer colorectal puisse être favorisée par des changements de la muqueuse associés à la maladie. Par ailleurs, Carvalho *et al.* ont récemment montré que les AIEC associés à la MC colonisent et induisent une inflammation intestinale forte chez les souris transgéniques exprimant CEACAM6, une molécule d'adhérence humaine [50]. CEACAM6 agit comme un récepteur pour les pili de type 1 produites par les AIEC [51], et sa surexpression est observée dans les polypes hyperplasiques et les adénomes précoces [52].

**Tableau 5.** Diversité quantitative des clones de *E. coli* issus de 77 échantillons de cancers colorectaux (côlon droit, colon gauche et rectum) et de 31 échantillons de diverticuloses.

Nombres de clones de <i>E. coli</i> / échantillon	Nombre (%) de patients						<i>P</i> <sup>a</sup> Divert. versus TDGR
	Divertic. (n=31)	TDGR (n=77)	TR (n=38)	TDG (n=39)	TD (n=21)	TG (n=18)	
0	6 (19,3)	4 (5,2)	2 (5,3)	2 (5,1)	1 (4,8)	1 (5,5)	0,031
1	8 (25,8)	29 (37,7)	13 (34,2)	16 (41)	10 (47,6)	6 (33,3)	
2	14 (45,2)	28 (36,4)	15 (39,5)	13 (33,3)	8 (38,1)	5 (27,8)	
3	2 (6,4)	12 (15,6)	6 (15,8)	6 (15,4)	2 (9,5)	4 (22,2)	
4	1 (3,2)	1 (1,3)	1 (2,6)				
5		2 (2,6)		2 (5,1)		2 (11,1)	
6		1 (1,3)	1 (2,6)				

Divert., groupe diverticuloses ; TDGR, groupe tumeurs du côlon et du rectum ; TDG, groupe tumeurs du côlon ; TD, groupe tumeurs du côlon droit ; TG, groupe tumeurs du côlon gauche ; TR, groupe tumeurs du rectum

<sup>a</sup> les valeurs de *P* (par le test exact de Fischer) sont montrées lorsque *P* est <0,05.

Les phylogroupes des *E. coli* issus de cancers colorectaux, de diverticuloses, de biopsies iléales lésées de MC et de *E. coli* fécaux commensaux ont été comparés (**tableau 6**). Aucune différence n'ayant été observée entre les séries de biopsies de tumeurs coliques droites et gauches, celles-ci ont été associées en une seule série. Bien qu'un seul clone ait été isolé des séries de MC et fécale, il faut noter que la prévalence d'un clone de *E. coli* du phylogroupe B2 est significativement plus élevée dans les biopsies de tumeurs coliques que dans les biopsies de diverticuloses, de tumeurs rectales, de lésions iléales de MC et dans

les fèces ( $p$ , <0,001 à 0,015). De plus, la prévalence d'un clone de *E. coli* du phylogroupe D est significativement plus élevée dans la série de tumeurs rectales que dans la série de souches fécales commensales ( $p=0,027$ ). 60% des isolats issus de biopsies de MC appartiennent aux phylogroupes B2+D ; bien que non significative, cette prédominance a été précédemment rapportée par Kotlowski *et al* [53].

**Tableau 6.** Diversité phylogénétique des clones de *E. coli* issus de 77 échantillons de cancers colorectaux (côlon droit, côlon gauche et rectum), de 31 échantillons de diverticuloses, de 20 isolats de *E. coli* issus d'échantillons d'iléums lésés de maladie de Crohn et de 51 isolats de *E. coli* fécaux commensaux.

Phylo- groupes de <i>E.</i> <i>coli</i>	Nombre (%) de patients présentant un clone de <i>E. coli</i>					$P^a$				$P^a$
	Divert. (n=31)	TDG (n=39)	TR (n=38)	Fecal (n=51)	Crohn (n=20)	TDG <i>versus</i>				TR <i>versus</i>
						Divert.	TR	Fecal	Crohn	Fecal
<b>A</b>	13 (41,9)	11 (28,2)	14 (36,8)	19 (37)	4 (20)					
<b>B1</b>	2 (6,4)	9 (23,1)	9 (23,7)	12 (24)	4 (20)					
<b>B2</b>	13 (41,9)	28 (71,8)	16 (42,1)	12 (24)	6 (30)	0,015	0,011	<0,001	0,005	
<b>D</b>	10 (32,3)	10 (25,6)	14 (36,8)	8 (16)	6 (30)					0,027

Divert., groupe diverticuloses ; TDG, groupe tumeurs du côlon ; TR, groupe tumeurs du rectum ; Fecal, groupe fécaux commensaux ; Crohn, groupe biopsies iléales de Crohn

<sup>a</sup> les valeurs de  $P$  (par le test exact de Fischer) sont montrées lorsque  $P$  est <0,05.

### ***Distribution des gènes codant les cyclomodulines selon l'origine physiopathologique des souches de E. coli***

La distribution des gènes codant les cyclomodulines et l'alpha-hémolysine a été analysée selon le contexte physiopathologique (**tableau 7**). Quatre facteurs de virulence, l'îlot *pks*, le gène *cnf1*, la protéine HlyA et les gènes *cdtB* ont été retrouvés préférentiellement et de manière significative chez les patients présentant une tumeur.

L'îlot *pks* était significativement plus prévalent dans la série de tumeurs coliques que dans les autres séries. Cuevas-Ramos *et al* ont montré que la colibactine produite par des souches de *E. coli* possédant l'îlot *pks* avait des propriétés génotoxiques *in vitro* et *in vivo* [54]. Ces souches sont mutagènes, induisent des dommages chromosomiques et une instabilité génétique *in vitro* de manière dose-dépendante. Cette instabilité génétique est retrouvée dans 85% des cancers colorectaux sporadiques [55]. Une infection de quelques heures d'entérocytes dans un modèle d'anse intestinale ligaturée de souris par une souche de *E. coli* produisant la colibactine induisait des cassures double-brin de l'ADN correspondant à une irradiation de 0,5 Grays. [54]. De plus, les lignées cellulaires humaines

infectées transitoirement par des *E. coli* produisant la colibactine montrent la formation de colonies cellulaires indépendantes de l'ancrage, un processus intervenant dans la diffusion des métastases du cancer colorectal [54]. Ces données suggèrent une activité mutagène et de transformation cellulaire des *E. coli* possédant l'îlot *pks*. Par son action génotoxique, la colibactine pourrait être impliquée dans les différentes étapes du développement tumoral colorectal.

La prévalence de l'opéron *hly-cnf1* était significativement plus élevée dans les tumeurs coliques et rectales que dans les fèces ( $p < 0,001$  et  $p = 0,019$ , respectivement). CNF est un promoteur de la prolifération cellulaire [8]. Cette toxine active les Rho-GTPases, déclenchant la transition de phase G1-S du cycle cellulaire et induisant la réplication de l'ADN sans cytokinèse. De plus, il a été montré que CNF1 inhibe l'apoptose et induit des ségrégations chromosomiques anormales [56, 57].

La prévalence de l'alpha-hémolysine était significativement supérieure dans les tumeurs coliques que dans les fèces, les diverticuloses et la MC ( $p < 0,001$ ,  $p = 0,019$  et  $p = 0,042$ , respectivement). L'îlot *pks* et l'opéron *hly-cnf1* étant presque exclusivement retrouvés dans les isolats du phylogroupe B2, on retrouve une prévalence significativement plus élevée de ces facteurs de virulence dans les tumeurs coliques et rectales que dans les fèces ( $p < 0,001$  et  $p = 0,039$ , respectivement). Il a été rapporté que l'alpha-hémolysine de *E. coli* augmente la perméabilité du côlon [58]. Ce processus représente une voie de translocation bactérienne facilitant la colonisation d'un tissu tumoral et la dissémination d'autres toxines bactériennes comme CNF1 et la colibactine. Néanmoins, l'alpha-hémolysine pourrait aussi être un facteur anti-tumoral en favorisant la lyse des cellules néoplasiques.

La prévalence du gène *cdtB* était significativement supérieure dans les tumeurs coliques que dans les diverticules ( $p = 0,030$ ). Ce gène était aussi présent chez des *E. coli* issus de tumeurs rectales, de lésions iléales de Crohn et fécaux. CdtB présente une activité de type DNase et induit des cassures simple et/ou double-brin de l'ADN des cellules eucaryotes [12]. La toxine CDT produite par *Helicobacter hepaticus*, une bactérie hépatocarcinogène pour la souris favorise le développement de dysplasie hépatique dans un modèle murin [59]. De plus, *H. hepaticus* a montré un pouvoir promoteur de la tumorigenèse colique dans différents modèles murins [60-62]. L'activité génotoxique de CDT favoriserait l'acquisition d'évènements mutationnels pouvant activer des gènes oncogènes et/ou inactiver des gènes suppresseurs de tumeurs. Comme pour la colibactine, CDT pourrait ainsi être impliquée, par son action génotoxique dans les différentes étapes du développement tumoral colorectal. Les gènes *cdt* sont présents chez diverses bactéries entériques pathogènes pour l'homme [59], comme les espèces de *Campylobacter* spp. qui sont une des

premières causes de gastro-entérites dans les pays industrialisés. En outre, *Campylobacter jejuni* a été associé à certains lymphomes de l'intestin grêle [63].

La présence de trois isolats présentant les gènes *cif* et *eae*, et donc à ce titre classés comme des EPEC a été observée dans la série de tumeurs coliques. Le gène *eae* des EPEC est codé par l'îlot de pathogénicité LEE, qui provoque des lésions d'attachement et d'effacement au niveau de la muqueuse intestinale [2]. Ainsi, les EPEC sont une cause majeure de diarrhée infantile dans les pays en développement [7]. Dans les pays développés l'incidence des diarrhées induites par les EPEC a diminué, et leur dépistage systématique n'est plus effectué [64]. Toutefois, des études récentes en Europe et en Australie ont identifié les EPEC chez 2,5 à 10,9% des enfants en bonne santé [65, 66], ce qui suggère qu'un pourcentage important de la population en bonne santé des pays développés sont porteurs de EPEC. Les travaux de Maddocks *et al* suggèrent que les EPEC sont plus fréquents dans les tissus d'adénocarcinomes coliques que dans des tissus coliques normaux appariés, avec une prévalence respective de 25% et 0% (n=20) [67]. Le locus LEE code un système de sécrétion bactérien de type III (SST3) capable d'injecter des effecteurs bactériens qui pourraient favoriser la cancérogenèse en agissant sur différents processus comme celui de la réparation des lésions de l'ADN [67, 68]. Un des effecteur pouvant intervenir est le produit du gène *cif*, qui altère le processus d'ubiquitination et donc la dégradation de protéines intervenant dans de nombreux processus cellulaires comme la régulation du cycle cellulaire ou du cytosquelette [10, 11]. Par ailleurs, *Citrobacter rodentium*, qui porte le locus LEE et provoque des lésions d'attachement et d'effacement, est responsable d'hyperplasies coliques transmissibles [69], réduit la période de latence de l'apparition de tumeurs induites chimiquement [70], et favorise la formation adénome colique chez la souris *Min* [71].

**Tableau 7.** Présence chez les patients des gènes codant les CM dans des isolats de *E. coli* issus de 77 échantillons de cancers colorectaux (côlon droit, côlon gauche et rectum), de 31 échantillons de diverticuloses, et dans 20 isolats de *E. coli* issus d'échantillons d'iléums lésés de maladie de Crohn et 51 isolats de *E. coli* fécaux commensaux.

Facteurs de virulence de <i>E. coli</i>	Nombre (%) de patients présentant un clone de <i>E. coli</i> possédant le facteur de virulence					$P^a$			
	Divert. (n=31)	TDG (n=39)	TR (n=38)	Fecal (n=51)	Crohn (n=20)	TDG versus			
						Divert	R	Fecal	Crohn
<i>pks</i>	6 (19,4)	20 (51,3)	8 (21)	6 (12)	2 (10)	0,007	0,009	<0,001	0,002
<i>cnf1</i> -Hly	4 (12,9)	14 (35,9)	7 (18,4)	1 (2)	2* (10)			<0,001	0,019
Hly non associé à <i>cnf1</i>	1 (3,2)	3 (7,7)	1 (2,6)	3 (6)	1 (5)				
Hly	5 (16,1)	17 (43,6)	8 (21,1)	4 (7,8)	3 (15)	0,019		<0,001	0,042
<i>cnf1</i> -Hly, <i>pks</i>	4 (12,9)	12 (30,8)	6 (15,8)	1 (2)	2* (10)			<0,001	0,039
<i>cnf2-cdtB</i> -III	0	1 (2,6)	0	0	0				
<i>cdtB</i>	0	6 (15,4)	2 (5,3)	6 (12)	1 (5)	0,030			
<i>cif</i>	0	3 (7,7)	0	0	0				

Divert., groupe diverticuloses ; TDG, groupe tumeurs du côlon ; TR, groupe tumeurs du rectum ; Fecal, groupe fécaux commensaux ; Crohn, groupe biopsies iléales de Crohn  
<sup>a</sup> les valeurs de  $P$  (par le test exact de Fischer) sont montrées lorsque  $P$  est <0,05.

Les bactéries, pathogènes ou commensales peuvent avoir un rôle dans la carcinogenèse, notamment colorectale [18, 21, 22]. Cependant, le développement tumoral est un processus long et multifactoriel. Notre étude épidémiologique montre que des facteurs de virulence de *E. coli* pourraient être impliqués dans la tumorigenèse colorectale. Néanmoins, la présence de ces facteurs dans la tumeur ne présume pas de leur présence et de leur implication dans le processus tumoral entamé ultérieurement. Clarifier le rôle possible de ces facteurs de virulence dans la carcinogenèse colorectale nécessiterait une étude observationnelle longitudinale de population, prenant en compte les nombreux facteurs de risques et facteurs confondants (génétiques et environnementaux) chez les patients. Se concentrer sur plusieurs sous-ensembles homogènes de la maladie pourrait permettre d'augmenter la sensibilité des études.

## Références

1. Gordon, D.M., et al., *Assigning Escherichia coli strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method*. Environ Microbiol, 2008. **10**(10): p. 2484-96.
2. Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli*. Nat Rev Microbiol, 2004. **2**(2): p. 123-40.
3. Darfeuille-Michaud, A., *Adherent-invasive Escherichia coli: a putative new E. coli pathotype associated with Crohn's disease*. Int J Med Microbiol, 2002. **292**(3-4): p. 185-93.
4. Touchon, M., et al., *Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths*. PLoS Genet, 2009. **5**(1): p. e1000344.
5. Oswald, E., et al., *Bacterial toxins that modulate host cell-cycle progression*. Curr Opin Microbiol, 2005. **8**(1): p. 83-91.
6. Nougayrede, J.P., et al., *Escherichia coli induces DNA double-strand breaks in eukaryotic cells*. Science, 2006. **313**(5788): p. 848-51.
7. Peres, S.Y., et al., *A new cytolethal distending toxin (CDT) from Escherichia coli producing CNF2 blocks HeLa cell division in G2/M phase*. Mol Microbiol, 1997. **24**(5): p. 1095-107.
8. Lemonnier, M., L. Landraud, and E. Lemichez, *Rho GTPase-activating bacterial toxins: from bacterial virulence regulation to eukaryotic cell biology*. FEMS Microbiol Rev, 2007. **31**(5): p. 515-34.
9. Orden, J.A., et al., *Necrotogenic Escherichia coli from sheep and goats produce a new type of cytotoxic necrotizing factor (CNF3) associated with the eae and ehxA genes*. Int Microbiol, 2007. **10**(1): p. 47-55.
10. Marches, O., et al., *Enteropathogenic and enterohaemorrhagic Escherichia coli deliver a novel effector called Cif, which blocks cell cycle G2/M transition*. Mol Microbiol, 2003. **50**(5): p. 1553-67.
11. Jubelin, G., et al., *Pathogenic bacteria target NEDD8-conjugated cullins to hijack host-cell signaling pathways*. PLoS Pathog, 2010. **6**(9).
12. Ge, Z., D.B. Schauer, and J.G. Fox, *In vivo virulence properties of bacterial cytolethal-distending toxin*. Cell Microbiol, 2008. **10**(8): p. 1599-607.
13. Labianca, R., et al., *Colon cancer*. Crit Rev Oncol Hematol, 2010. **74**(2): p. 106-33.
14. Karsa, L., et al., *The dimensions of the CRC problem*. Best Pract Res Clin Anaesthesiol, 2010. **24**(4): p. 381-96.
15. Kwak, E.L. and D.C. Chung, *Hereditary colorectal cancer syndromes: an overview*. Clin Colorectal Cancer, 2007. **6**(5): p. 340-4.
16. von Roon, A.C., et al., *The risk of cancer in patients with Crohn's disease*. Dis Colon Rectum, 2007. **50**(6): p. 839-55.
17. Eaden, J.A., K.R. Abrams, and J.F. Mayberry, *The risk of colorectal cancer in ulcerative colitis: a meta-analysis*. Gut, 2001. **48**(4): p. 526-35.
18. Burnett-Hartman, A.N., P.A. Newcomb, and J.D. Potter, *Infectious agents and colorectal cancer: a review of Helicobacter pylori, Streptococcus bovis, JC virus, and human papillomavirus*. Cancer Epidemiol Biomarkers Prev, 2008. **17**(11): p. 2970-9.
19. Uronis, J.M. and C. Jobin, *Microbes and colorectal cancer: is there a relationship?* Curr Oncol, 2009. **16**(4): p. 22-4.
20. Travaglione, S., A. Fabbri, and C. Fiorentini, *The Rho-activating CNF1 toxin from pathogenic E. coli: a risk factor for human cancer development?* Infect Agent Cancer, 2008. **3**: p. 4.
21. Lax, A.J., *New genotoxin shows diversity of bacterial attack mechanisms*. Trends Mol Med, 2007. **13**(3): p. 91-3.



22. Mager, D.L., *Bacteria and cancer: cause, coincidence or cure? A review*. J Transl Med, 2006. **4**: p. 14.
23. Dubois, D., et al., *Cyclomodulins in urosepsis strains of Escherichia coli*. J Clin Microbiol, 2010. **48**(6): p. 2122-9.
24. Dozois, C.M., et al., *Expression of P, S, and F1C adhesins by cytotoxic necrotizing factor 1-producing Escherichia coli from septicemic and diarrheic pigs*. FEMS Microbiol Lett, 1997. **152**(2): p. 307-12.
25. Oswald, E., et al., *Virulence factors associated with cytotoxic necrotizing factor type two in bovine diarrheic and septicemic strains of Escherichia coli*. J Clin Microbiol, 1991. **29**(11): p. 2522-7.
26. Pickett, C.L., et al., *Patterns of variations in Escherichia coli strains that produce cytolethal distending toxin*. Infect Immun, 2004. **72**(2): p. 684-90.
27. Silver, R.P., et al., *Comparative analysis of plasmids and some metabolic characteristics of Escherichia coli K1 from diseased and healthy individuals*. Infect Immun, 1980. **29**(1): p. 200-6.
28. Perna, N.T., et al., *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7*. Nature, 2001. **409**(6819): p. 529-33.
29. Darfeuille-Michaud, A., et al., *High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease*. Gastroenterology, 2004. **127**(2): p. 412-21.
30. Darfeuille-Michaud, A., et al., *Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease*. Gastroenterology, 1998. **115**(6): p. 1405-13.
31. Ochman, H. and R.K. Selander, *Standard reference strains of Escherichia coli from natural populations*. J Bacteriol, 1984. **157**(2): p. 690-3.
32. Bidet, P., et al., *Detection and identification by PCR of a highly virulent phylogenetic subgroup among extraintestinal pathogenic Escherichia coli B2 strains*. Appl Environ Microbiol, 2007. **73**(7): p. 2373-7.
33. Versalovic, J., T. Koeuth, and J.R. Lupski, *Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes*. Nucleic Acids Res, 1991. **19**(24): p. 6823-31.
34. Wang, G., et al., *RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains*. Nucleic Acids Res, 1993. **21**(25): p. 5930-3.
35. Chassagne, L., et al., *Detection of stx1, stx2, and eae genes of enterohemorrhagic Escherichia coli using SYBR Green in a real-time polymerase chain reaction*. Diagn Microbiol Infect Dis, 2009. **64**(1): p. 98-101.
36. Putze, J., et al., *Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae*. Infect Immun, 2009. **77**(11): p. 4696-703.
37. Johnson, J.R., et al., *Molecular epidemiology and phylogenetic distribution of the Escherichia coli pks genomic island*. J Clin Microbiol, 2008. **46**(12): p. 3906-11.
38. De Rycke, J., A. Milon, and E. Oswald, *Necrotoxic Escherichia coli (NTEC): two emerging categories of human and animal pathogens*. Vet Res, 1999. **30**(2-3): p. 221-33.
39. Blum, G., et al., *Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic Escherichia coli strain J96*. FEMS Microbiol Lett, 1995. **126**(2): p. 189-95.
40. Guyer, D.M., J.S. Kao, and H.L. Mobley, *Genomic analysis of a pathogenicity island in uropathogenic Escherichia coli CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and Catheter-associated bacteriuria and from fecal samples*. Infect Immun, 1998. **66**(9): p. 4411-7.
41. Toth, I., et al., *Cytolethal distending toxin type I and type IV genes are framed with lambdoid prophage genes in extraintestinal pathogenic Escherichia coli*. Infect Immun, 2009. **77**(1): p. 492-500.

42. Johnson, T.J., et al., *Pyrosequencing of the Vir plasmid of necrotoxicogenic Escherichia coli*. Vet Microbiol, 2010. **144**(1-2): p. 100-9.
43. Bielaszewska, M., et al., *Characterization of cytolethal distending toxin genes and expression in shiga toxin-producing Escherichia coli strains of non-O157 serogroups*. Infect Immun, 2004. **72**(3): p. 1812-6.
44. Orth, D., et al., *Cytolethal distending toxins in Shiga toxin-producing Escherichia coli: alleles, serotype distribution and biological effects*. J Med Microbiol, 2006. **55**(Pt 11): p. 1487-92.
45. Pandey, M., et al., *Association of cytolethal distending toxin locus cdtB with enteropathogenic Escherichia coli isolated from patients with acute diarrhea in Calcutta, India*. J Clin Microbiol, 2003. **41**(11): p. 5277-81.
46. Loukiadis, E., et al., *Distribution, functional expression, and genetic organization of Cif, a phage-encoded type III-secreted effector from enteropathogenic and enterohemorrhagic Escherichia coli*. J Bacteriol, 2008. **190**(1): p. 275-85.
47. Stritzker, J., et al., *Tumor-specific colonization, tissue distribution, and gene induction by probiotic Escherichia coli Nissle 1917 in live mice*. Int J Med Microbiol, 2007. **297**(3): p. 151-62.
48. Martin, H.M., et al., *Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer*. Gastroenterology, 2004. **127**(1): p. 80-93.
49. Swidsinski, A., et al., *Association between intraepithelial Escherichia coli and colorectal cancer*. Gastroenterology, 1998. **115**(2): p. 281-6.
50. Carvalho, F.A., et al., *Crohn's disease adherent-invasive Escherichia coli colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM*. J Exp Med, 2009. **206**(10): p. 2179-89.
51. Barnich, N., et al., *CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease*. J Clin Invest, 2007. **117**(6): p. 1566-74.
52. Scholzel, S., et al., *Carcinoembryonic antigen family members CEACAM6 and CEACAM7 are differentially expressed in normal tissues and oppositely deregulated in hyperplastic colorectal polyps and early adenomas*. Am J Pathol, 2000. **156**(2): p. 595-605.
53. Kotlowski, R., et al., *High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease*. Gut, 2007. **56**(5): p. 669-75.
54. Cuevas-Ramos, G., et al., *Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells*. Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11537-42.
55. Vogelstein, B., et al., *Genetic alterations during colorectal-tumor development*. N Engl J Med, 1988. **319**(9): p. 525-32.
56. Fiorentini, C., et al., *Toxin-induced activation of Rho GTP-binding protein increases Bcl-2 expression and influences mitochondrial homeostasis*. Exp Cell Res, 1998. **242**(1): p. 341-50.
57. Malorni, W. and C. Fiorentini, *Is the Rac GTPase-activating toxin CNF1 a smart hijacker of host cell fate?* Faseb J, 2006. **20**(6): p. 606-9.
58. Troeger, H., et al., *Escherichia coli alpha-haemolysin induces focal leaks in colonic epithelium: a novel mechanism of bacterial translocation*. Cell Microbiol, 2007. **9**(10): p. 2530-40.
59. Ge, Z., et al., *Bacterial cytolethal distending toxin promotes the development of dysplasia in a model of microbially induced hepatocarcinogenesis*. Cell Microbiol, 2007. **9**(8): p. 2070-80.
60. Nagamine, C.M., et al., *Helicobacter hepaticus promotes azoxymethane-initiated colon tumorigenesis in BALB/c-IL10-deficient mice*. Int J Cancer, 2008. **122**(4): p. 832-8.
61. Nagamine, C.M., et al., *Helicobacter hepaticus infection promotes colon tumorigenesis in the BALB/c-Rag2(-/-) Apc(Min/+) mouse*. Infect Immun, 2008. **76**(6): p. 2758-66.

62. Fox, J.G., et al., *Gut microbes define liver cancer risk in mice exposed to chemical and viral transgenic hepatocarcinogens*. *Gut*, 2010. **59**(1): p. 88-97.
63. Lecuit, M., et al., *Immunoproliferative small intestinal disease associated with Campylobacter jejuni*. *N Engl J Med*, 2004. **350**(3): p. 239-48.
64. Nataro, J.P. and J.B. Kaper, *Diarrheagenic Escherichia coli*. *Clin Microbiol Rev*, 1998. **11**(1): p. 142-201.
65. Afset, J.E., et al., *Association of atypical enteropathogenic Escherichia coli (EPEC) with prolonged diarrhoea*. *J Med Microbiol*, 2004. **53**(Pt 11): p. 1137-44.
66. Beutin, L., et al., *HEp-2 cell adherence, actin aggregation, and intimin types of attaching and effacing Escherichia coli strains isolated from healthy infants in Germany and Australia*. *Infect Immun*, 2003. **71**(7): p. 3995-4002.
67. Maddocks, O.D., et al., *Attaching and effacing Escherichia coli downregulate DNA mismatch repair protein in vitro and are associated with colorectal adenocarcinomas in humans*. *PLoS One*, 2009. **4**(5): p. e5517.
68. Vogelmann, R. and M.R. Amieva, *The role of bacterial pathogens in cancer*. *Curr Opin Microbiol*, 2007. **10**(1): p. 76-81.
69. Schauer, D.B., et al., *Genetic and biochemical characterization of Citrobacter rodentium sp. nov.* *J Clin Microbiol*, 1995. **33**(8): p. 2064-8.
70. Barthold, S.W. and A.M. Jonas, *Morphogenesis of early 1, 2-dimethylhydrazine-induced lesions and latent period reduction of colon carcinogenesis in mice by a variant of Citrobacter freundii*. *Cancer Res*, 1977. **37**(12): p. 4352-60.
71. Newman, J.V., et al., *Bacterial infection promotes colon tumorigenesis in Apc(Min/+) mice*. *J Infect Dis*, 2001. **184**(2): p. 227-30.

**Chapitre III**

**Discussion**

**et**

**Perspectives**



Les résultats présentés précédemment mettent en évidence l'originalité de la protéine ClbP et de ses homologues au sein de la famille des enzymes S12 avec notamment leur association aux ilots de PKS-NRPS comme *pks* et l'implication potentielle de ce dernier dans les relations hôtes-bactéries. D'une part, l'activité peptidase de ClbP a été montrée par l'analyse de la structure cristalline de son domaine enzymatique, sa caractérisation biochimique et des travaux de mutagenèse qui ont confirmé le rôle essentiel de cette activité pour la fonctionnalité de l'ilot *pks*. De plus, ClbP présente une promiscuité fonctionnelle avec des homologues présents dans d'autres ilots synthétisant des composés NRP bactériens. D'autre part, les études épidémiologiques ont révélé une plus forte prévalence de l'ilot *pks* chez des isolats de *E. coli* issus d'urosepsis comparés à des isolats de *E. coli* commensaux, et chez des isolats de *E. coli* issus de biopsies de tumeurs coliques comparés à des isolats de *E. coli* commensaux ou issus de biopsies de diverticuloses ou de tumeurs rectales.

## **I. Fonction biochimique de ClbP et de ses orthologues**

### **A. ClbP, une peptidase des composés NRP**

Notre travail suggère que la colibactine synthétisée par les enzymes de l'ilot *pks* est maturée par la peptidase ClbP dans le périplasme. L'orthologue ZmaM serait impliqué de même dans la synthèse de la Zwittermicine A (ZmA), un composé PK-NRP de structure connue (aminopolyol linéaire), qui est produit par un ilot génomique de *Bacillus* du groupe *cereus* (*B. cereus*, *B. thuringiensis* et *B. mycoides*) [382]. ZmA est biosynthétisée sous forme d'un métabolite secondaire de grande taille, qui serait clivé au niveau d'une liaison peptidique pour former le composé actif ZmA [381]. La seule enzyme codée par l'ilot de biosynthèse susceptible de catalyser cette réaction est la protéine ClbP-like ZmaM. Ces données confortent nos résultats qui suggèrent un rôle conservé des protéines ClbP-like dans la maturation de composés de type NRP ou PK-NRP. Nous proposons donc de désigner ces protéines ClbP-like sous le terme de « peptidases associées aux NRP » (PaNRP).

### **B. Recherche d'un motif clivé par ClbP et ses orthologues**

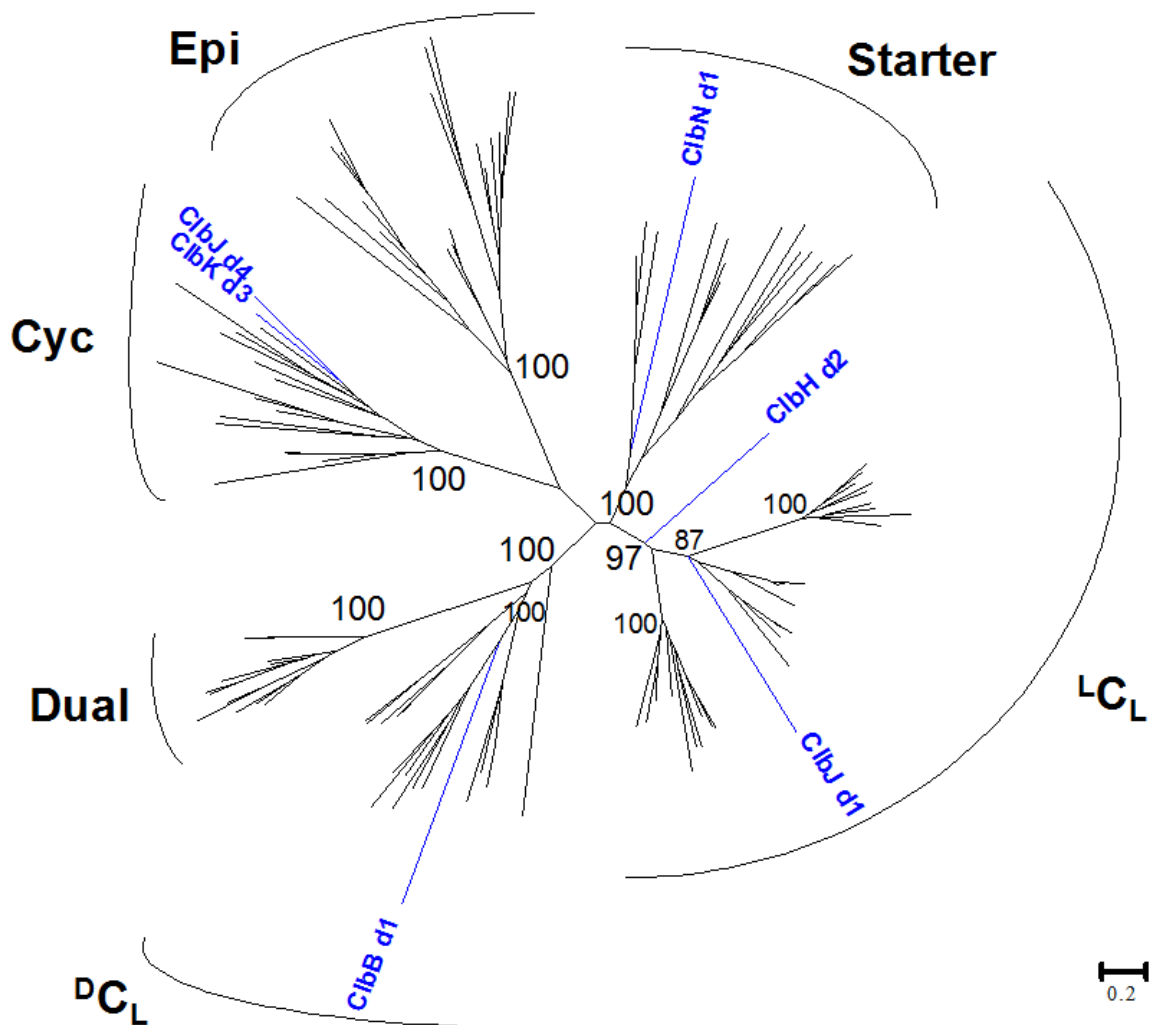
Le domaine PaNRP de ZmaM cliverait un métabolite « pré-Zma » entre une D-asparagine et une D-sérine, libérant un acide gras-D-asparagine et le composé ZmA portant le

résidu D-sérine dont la stéréospécificité est un déterminant important de l'activité biologique du composé [381, 383]. De manière intéressante, des analyses bioinformatiques de la spécificité du substrat montrent que les îlots génomiques produisant une PaNRP codent aussi pour des domaines d'adénylation A incorporant une asparagine ou un résidu apparenté, la glutamine. De plus, toutes les peptidases de la famille MEROPS S12 caractérisées biochimiquement comme les  $\beta$ -lactamases AmpC et la protéase Pab87 possèdent une spécificité pour les substrats porteurs d'une stéréospécificité D [379, 384-387]. Des résidus asparagine et glutamine de stéréospécificité D pourraient donc être un motif impliqué dans la reconnaissance du substrat par les PaNRP.

### **C. Localisation de ClbP, de ses orthologues et de la colibactine**

Au sein des clusters de biosynthèse, les PaNRP sont les seules protéines qui possèdent une séquence signal de translocation périplasmique et un domaine catalytique prédits en position périplasmique. Les PaNRP interviendraient donc dans le périplasma lors de la phase finale de synthèse, après l'assemblage du squelette carboné dans le cytoplasme. Cependant, l'isolement et la détermination de la structure de la colibactine ou de ZmA avant, après ou pendant l'action de ClbP ou d'une autre PaNRP comme ZmaM sont nécessaires pour avoir la preuve directe de cette fonction.

Le principal écueil à ce travail est l'isolement de la colibactine. A ce jour, elle n'a jamais été mise en évidence directement. Nous avons mené une analyse phylogénétique des domaines de condensation C des enzymes NRPS codées par l'îlot *pks*. Les résultats suggèrent que le domaine C de ClbN soit un domaine « starter C » habituellement observé dans des clusters synthétisant des lipopeptides (**figure 36**). Typiquement, il acyle le premier acide aminé intégré dans le composé avec un acide  $\beta$ -hydroxy-carboxylique ou acide gras  $\beta$ -hydroxylé et favorise ainsi l'ancrage du composé aux membranes [277]. La colibactine pourrait donc porter un acide gras responsable de sa rétention dans la membrane externe des bactéries, à la manière des lipoprotéines de Braun, du viscosinamide, ou du composé PDIM membranaire de *M. tuberculosis* [373, 388-390]. Cet aspect doit être intégré dans la stratégie d'extraction et de purification de la colibactine. Cette hypothèse expliquerait aussi la nécessité d'un contact étroit entre la bactérie produisant la colibactine et la cellule cible pour observer l'effet cytopathogène de la colibactine. Ainsi, la colibactine serait délivrée aux cellules eucaryotes par contact [24].



**Figure 36 - Arbre phylogénétique des sous-types de domaines de condensation C de NRPS (domaines <sup>L</sup>C<sub>L</sub>, <sup>D</sup>C<sub>L</sub>, Starter, Dual E/C, Epi : Epimerisation et Cyc : Hétérocyclisation). Modifié d'après Rausch *et al* [277].**

Les séquences peptidiques des domaines C de l'îlot *pks* ont été alignées à l'aide de COBALT un alignement de domaines C généré par Rausch *et al.*. La phylogénie a ensuite été reconstruite en utilisant PhyML avec le modèle de substitution WAG (Gamma et Sites invariables).

Les valeurs de bootstrap, générées à partir de 100 arbres, sont indiquées pour les principales branches. Les domaines C de l'îlot *pks* sont en bleu et désignés par le nom de la protéine puis le numéro de domaine sur la protéine.



ClbP est associée à la membrane cytoplasmique. Cette localisation membranaire est observée pour d'autres protéines intervenant dans la synthèse de composés PK, NRP, PK-NRP ou apparentés. Une protéine PKS intervenant dans la synthèse la myxovirescine A est localisée dans la membrane cytoplasmique de *Myxococcus xanthus* [391]. Les protéines de l'îlot de biosynthèse du bacillaene s'assemblent en un mégacomplexe de mégasynthases, formant un organite unique et juxtaposé à la membrane cellulaire, une organisation probablement maintenue par l'association des lignes d'assemblage à un domaine membranaire [392]. La synthèse des peptides lantibiotiques fait intervenir des complexes enzymatiques ancrés sur la face cytoplasmique de la membrane bactérienne [393, 394]. Un couplage entre la synthèse et l'exporteur du composé PDIM, un PK-acide gras a été rapporté chez *M. tuberculosis* : le domaine cytoplasmique de l'exporteur MmpL7 de la famille des RND (Résistance, Nodulation et Division cellulaire) perméases interagit avec l'enzyme PKS PpsE qui est requise pour l'étape finale de la biosynthèse du phthiocerol [395]. La localisation membranaire de ClbP est probablement due à ses trois hélices transmembranaires C-terminales putatives. Ces hélices sont conservées chez les protéines PaNRP. De plus, de nombreuses protéines PaNRP portent des domaines ABC exporteurs, connus pour être également membranaires. La localisation membranaire de ClbP est donc probablement conservée chez les orthologues. Cependant, l'importance de cette localisation et les éventuelles interactions associées, notamment avec les pompes d'efflux et la chaîne d'assemblage PK-NRP restent à explorer.

## II. Rôle biologique de ClbP et de ses orthologues

### A. ClbP et la synthèse d'un composé NRP bioactif

Retrouvés principalement chez des souches de Firmicutes (genres *Bacillus*, *Clostridium* et apparentés), les gènes codant les PaNRP sont souvent associés à des clusters de biosynthèse codant des NRPS. Cependant, certains gènes des PaNRP ne sont pas génétiquement liés à de tels clusters, alors qu'ils sont capables de restaurer l'activité biologique de l'îlot *pks* défectif pour le gène *clbP*. Cette observation pose le problème du rôle biologique des protéines PaNRP non liées génétiquement à un cluster de biosynthèse de NRP. Dans le génome de Streptomycètes, le nombre de voies de biosynthèse nécessitant une phosphopantéthéinylation dépassent celui des enzymes PPTases codées [396]. Une PPTase est donc utilisée par de multiples voies de biosynthèse et leurs gènes peuvent être codés dans une région distincte de l'îlot principal de biosynthèse. Les PaNRP non liées génétiquement à des clusters NRPS pourraient donc être codées à distance des principaux gènes de la chaîne d'assemblage du NRP. Cette hypothèse est soutenue par l'observation de un à trois îlots de biosynthèse de NRP putatifs dans 15 des 16 génomes possédant une PaNRP non liée génétiquement à des îlots de NRPS.

Par ailleurs, les gènes de ces PaNRP chez *Bacillus* sp. B14905 et *Lysinibacillus sphaericus* C3-41 sont dans le voisinage de gènes codant des homologues du système de quorum-sensing « *agr* », des enzymes de synthèse de lantibiotiques ou de protéines de germination de spores. Aussi nous ne pouvons pas éliminer la possibilité que ces PaNRP puissent être impliquées dans la biosynthèse d'autres types de molécules naturelles portant des liaisons peptidiques, comme des molécules de communication tels que les lantibiotiques et d'autres molécules du quorum-sensing [335].

La source majeure de diversité des composés NRP provient principalement des domaines d'adénylation A. Ces derniers sont responsables de la diversité des acides aminés incorporés par les modules NRPS [257]. Les enzymes PaNRP pourraient être utilisées comme un outil biotechnologique intéressant pour augmenter la diversité au sein des petites molécules synthétisées au laboratoire, comme cela a été proposé pour d'autres enzymes de finition intervenant la biosynthèse de métabolites secondaires [397, 398].

## **B. ClbP clive une séquence d'adressage ?**

Les bactériocines et autres peptides ribosomiaux antibiotiques tels que la microcine V de *E. coli* et les lantibiotiques des bactéries à Gram positif sont produits sous forme de pré-peptides incluant un peptide signal N-terminal. Ces pré-peptides maintiennent ces bactériocines dans une forme inactive dans la cellule productrice et facilitent les interactions avec les enzymes de modifications (hétérocyclases, oxydases) et les exporteurs. La séquence signal N-terminale est habituellement clivée pendant l'export par des cystéine protéases fusionnées à des transporteurs de type ABC ou après l'export par des sérine peptidases périplasmiques dédiées [338, 399, 400]. Par analogie, les PaNRP pourraient éliminer une partie du composé NRP impliqué dans les interactions avec des enzymes de synthèse et/ou des exporteurs.

## **C. ClbP et protection de la cellule productrice**

L'activité peptidase de ClbP est indispensable à l'activité biologique de l'îlot *pks*. De même, la maturation de ZmA par une peptidase, probablement la protéine PaNRP ZmaM permet de libérer le résidu D-Ser nécessaire à l'activité biologique du composé ZmA. L'analyse topologique de ces protéines *in silico* et les données expérimentales recueillies pour ClbP suggèrent que ces peptidases activent le composé NRP dans le périplasma. Par ailleurs, de nombreuses PaNRP comme ZmaM possèdent en C-terminal un domaine exporteur de type ABC. Cinq des sept exporteurs de type ABC fusionnés aux PaNRP appartiennent à la famille des exporteurs de peptides cycliques (accession InterProScan IPR005898). Ces exporteurs interviendraient dans le transport de composés NRP [401, 402]. ClbP et d'autres PaNRP liées génétiquement à des îlots de NRPS ne sont pas fusionnés à des domaines exporteurs. Cependant, les clusters de biosynthèse correspondants codent pour un exporteur de la famille MATE, comme ClbM dans l'îlot *pks*, ou d'autres familles de pompes d'efflux. La fusion d'une pompe d'efflux avec les PaNRP ne semble donc pas indispensable. De plus, le gène de la pompe ClbM n'est pas indispensable à l'îlot *pks* et peut probablement être suppléé par d'autres gènes du core-géome, les pompes d'efflux ayant souvent peu de spécificité pour les substrats. Après translocation d'un composé NRP inactif du cytoplasme vers le périplasma par des pompes d'efflux, la localisation périplasmique du domaine peptidase des PaNRP

pourrait permettre la maturation du composé NRP à distance d'une éventuelle cible cytoplasmique. L'association PaNRP-pompes d'efflux peut donc être considérée comme un mécanisme de protection à deux composants contre l'activité cytotoxique potentielle générée par certains ilots génomiques synthétisant des composés NRP.

#### **D. ClbP et protection contre des NRP exogènes**

La peptidase MlrB de *Sphingomonas* sp. ACM-3962 (44,3 kDa) est une peptidase de la famille MEROPS S12 qui intervient dans la dégradation de la microcystine LR, un NRP heptapeptidique toxique synthétisé par la cyanobactérie *Microcystis aeruginosa*. MlrB dégrade la microcystine LR en un térapeptide, en catalysant l'hydrolyse de la liaison peptidique D-alanyl-L-leucine [403]. Les PaNRP et notamment celles non liées génétiquement à des ilots de NRPS pourraient aussi avoir un rôle dans la dégradation de métabolites secondaires de type NRP ou proches, et ainsi permettre une résistance à des molécules antibiotiques produites par d'autres organismes. Par ailleurs, des centaines de bactéries du sol, phylogénétiquement diverses et souvent proches des pathogènes de l'homme ont la capacité de croître avec comme seule source de carbone des antibiotiques produits par des bactéries du sol ou d'origine synthétique [404]. Les PaNRP pourraient également se comporter comme des enzymes du métabolisme primaire en permettant à la bactérie d'utiliser les composés NRP comme source de carbone et d'azote.

### III. Distribution des gènes codant les PaNRP

Les PaNRP sont retrouvées chez les Firmicutes essentiellement et des  $\gamma$ -Protéobactéries ; *Hahella chejuensis* et, pour ClbP, certaines Entérobactéries comme *E. coli*, *K. pneumoniae* et *C. koseri* [246]. Les PaNRP sont présentes chez des pathogènes de mammifères (*E. coli*, *Clostridium botulinum*) et des pathogènes provoquant des intoxications alimentaires (*Bacillus pumilus* et groupe '*B. cereus*', synthétisant des NRP tels que la toxine émétisante céréulide) [260, 405]. Les PaNRP sont aussi retrouvées chez le pathogène des abeilles *Paenibacillus larvae* [406] et chez des souches étudiées dans le cadre de recherches appliquées : *H. chejuensis* [407], *Brevibacillus brevis* [408] et *Clostridium cellulolyticum* [409]. Cependant, le seul îlot génomique codant une PaNRP et conduisant à la production d'un composé NRP caractérisé est l'îlot de synthèse de la ZmA observé dans la souche de *B. cereus* UW85 et d'autres souches du groupe '*B. cereus*' [381].

La mobilité de l'îlot *pks* est supportée par la présence de gènes codant des intégrases, des transposases et l'analyse de l'environnement génétique chez *E. coli* et d'autres espèces d'Entérobactéries (cf revue bibliographique). Cette mobilité est une caractéristique commune à de nombreux clusters de biosynthèse de métabolites secondaires, notamment de composés NRPK. Ainsi, des îlots génomiques apparentés sont retrouvés dans des taxons microbiens qui ont divergé depuis des centaines de millions d'années. Par exemple, les îlots de synthèse des aminosides et des thiopeptides sont observés chez les Actinomycètes et les *Bacillus* spp. Les îlots de synthèse des céphalosporines sont retrouvés chez les Actinomycètes, les  $\gamma$ -Protéobactéries et les champignons avec des groupes de gènes de biosynthèse remarquablement similaires [410]. Ces données suggèrent le transfert en bloc de ces gènes par voie horizontale. Cette capacité à transférer pose le problème de l'origine de ces îlots génomiques et notamment de l'îlot *pks*. L'analyse de la distribution des gènes codant les PaNRP au sein des phylums bactériens fournit un indice. Ces derniers sont principalement observés dans les genres *Bacillus* et *Clostridium* (Firmicutes) et n'ont pas été retrouvés chez les Actinobactéries et les  $\beta$ -Protéobactéries, alors que les gènes des mégasynthèses sont principalement observés chez ces groupes bactériens [260]. Cette surreprésentation des gènes codant les PaNRP chez les genres *Bacillus* et *Clostridium* suggère une origine de ces gènes et donc de l'îlot *pks* chez les Firmicutes.

## IV. Fonctions biologiques de la colibactine

L'îlot *pks* de 54kb représente environ 1% du génome de *E. coli* et une seule copie de celui-ci code des mégasynthases et des protéines accessoires qui formeraient un complexe de 1,8 MDa, proche de la taille du ribosome (2.5 MDa). Bien que l'activité de la colibactine ne soit pas encore bien caractérisée, les ressources génétiques et métaboliques consacrées à sa production sous-tendent un rôle biologique important. Nougayrède *et al.* proposent que la colibactine puisse ralentir le renouvellement de l'épithélium intestinal chez les animaux et prolonger ainsi la colonisation de cet épithélium par la bactérie productrice en bloquant le cycle cellulaire par son action génotoxique [24]. D'autres fonctions pourraient lui être attribuées.

Comme certains lipopeptides, la colibactine pourrait être impliquée dans la motilité cellulaire et la formation de biofilm afin de favoriser le développement de colonies dans une niche écologique particulière [388]. La production de la colibactine comme celle des lipopeptides tel que le viscosinamide, lors de la phase de multiplication cellulaire jouerait probablement un rôle dans l'étape de colonisation d'une niche écologique. Les lipopeptides, dont la chaîne d'acide gras est importante pour leurs activités biologiques présentent tous à des degrés divers des propriétés de solubilité, biosurfactantes et d'affinité membranaire [388]. Cette propriété biosurfactante potentielle de la colibactine pourrait aider *E. coli* à coloniser les surfaces et à utiliser les substrats liés à ces surfaces, dans un environnement libre comme aux surfaces muqueuses intestinales des animaux et de l'homme.

La colibactine pourrait aussi posséder un rôle d'antagonisme contre d'autres microorganismes en perturbant des compétiteurs de la microflore, comme les macrolides inhibent le quorum-sensing de *P. aeruginosa* [411], ou présenter une action antimicrobienne [335]. *E. coli* peut vivre librement dans un environnement aquatique extérieur, y survivre longtemps et parcourir de longues distances : il est à ce titre un indicateur majeur de la contamination fécale de l'eau [412]. Certains environnements riches peuvent abriter de grandes populations de nématodes et de protozoaires, et par conséquent exercer une force sélective potentiellement puissante sur les microorganismes. La colibactine pourrait permettre à *E. coli* d'échapper à ces prédateurs, en les intoxiquant grâce à son action génotoxique ou par une possible action repulsive, comme cela a été mis en évidence pour certains lipopeptides tel que le viscosinamide [388, 413].

## V. Rôle de la colibactine dans le processus infectieux

Notre étude épidémiologique comparant une série de souches de *E. coli* issues d'urosepsis et une série de souches de *E. coli* souches isolées de fécès a mis en évidence une prévalence significativement supérieure de l'îlot *pks* parmi les souches urinaires. Ceci suggère que la colibactine serait un facteur de virulence dans le processus infectieux urinaire. Néanmoins la fonctionnalité de l'îlot a été vérifiée dans seulement 40 % de souches et un nombre non négligeable d'îlots était non fonctionnel dans les conditions expérimentales utilisées. De plus, si on compare les isolats appartenant au phylogroupe B2 entre les deux séries cliniques, la différence de prévalence n'est pas significative. Ainsi, d'un point de vue épidémiologique, la colibactine pourrait être tout aussi bien considérée comme un facteur de « fitness » ou de colonisation des épithélia intestinaux et urinaires.

Par ailleurs, la colibactine semble affecter le système immunitaire de l'hôte. Une lymphotoxicité de souches de *E. coli pks*-positives a été observée sur des cultures primaires mises en culture, avec des cassures double-brin de l'ADN, un arrêt du cycle cellulaire et une mort par apoptose [414]. De plus, il a été montré dans un modèle murin de sepsis avec injection dans le coussinet plantaire que la production de colibactine entraînait une aggravation de la lymphopénie et une apoptose massive des thymocytes [414]. L'exacerbation de la lymphopénie par des souches de *E. coli pks*-positives pourrait ainsi constituer un facteur de gravité lors d'un sepsis sévère. Néanmoins, la production de colibactine ne semblait pas influencer le taux de mortalité des souris. Il faut noter que le modèle animal de sepsis utilisé dans cette étude ne permet pas une analyse complète de la physiopathologie de l'infection urinaire et notamment la porte d'entrée usuelle de l'infection.

La colibactine favoriserait la colonisation de l'arbre urinaire en atténuant les réactions inflammatoires et immunitaires innées qui ne font pas intervenir les lymphocytes. Il serait ainsi intéressant d'évaluer l'action de souches de *E. coli pks*-positives sur le tractus urinaire - colonisation et altération de l'épithélium -, et les autres cellules du système immunitaire. Les cellules de l'immunité innée comme les polynucléaires neutrophiles et les monocytes-macrophages étant les premiers acteurs cellulaires face à l'infection urinaire, l'évaluation de la cytotoxicité de la colibactine sur ces cellules en culture pourrait être une approche dans l'étude de l'implication de la colibactine dans le processus infectieux. La colibactine pourrait permettre aux bactéries phagocytées d'entraîner une mort cellulaire rapide des polynucléaires neutrophiles et/ou des monocytes-macrophages. La colibactine serait alors une arme de

« résistance à la phagocytose » contre les cellules phagocytaires, permettant une survie et un réensemencement bactérien à partir de ces cellules immunitaires favorisant une infection sévère telle que le sepsis.

## VI. Rôle de la colibactine dans le cancer colorectal

Le cancer colorectal (CCR) est le troisième cancer le plus fréquent, chez l'homme comme chez la femme, et la quatrième cause de décès par cancer à travers le monde, avec 1,2 million de cas estimés et 609.000 décès estimés en 2008. En Europe, environ 250 000 nouveaux cas de tumeurs coliques sont diagnostiqués chaque année, comptant pour environ 9% de toutes les tumeurs malignes [415]. Basé sur les tendances démographiques, l'incidence annuelle devrait augmenter de près de 80% à 2,2 millions de cas au cours des deux prochaines décennies et la plupart de cette augmentation se produira dans les régions les moins développées du monde (62%), le taux de ce cancer augmentant avec l'industrialisation et l'urbanisation [416]. Dans les pays industrialisés, le cancer du côlon est deux fois plus fréquent que le cancer du rectum, alors que dans les pays en voie de développement, les taux sont généralement similaires. Environ 70% des patients atteints de cancer du côlon ont plus de 65 ans [415].

Les CCR surviennent le plus souvent de façon sporadique et sont héréditaires dans seulement 5% des cas [417]. Ces derniers sont dus à des altérations transmissibles du gène *APC* (*adenomatous polyposis coli*) ou de gènes de réparation de l'ADN. Divers facteurs de risque ont été identifiés pour les CCR sporadiques comme un régime alimentaire riche en viande rouge et en viandes de préparations industrielles, la consommation importante d'alcool (plus de 30g par jour), le tabagisme, l'utilisation chronique d'anti-inflammatoires non-stéroïdiens et certaines conditions physiopathologiques telles que le syndrome métabolique et les maladies inflammatoires chroniques de l'intestin (la maladie de Crohn et la recto-colite ulcéro-hémorragique) [415, 418, 419].

Différentes étapes du développement tumoral colorectal peuvent être observées. Cette séquence d'évènements bien définie – de la prolifération de cryptes aberrantes ou hyperplasies en adénomes bénins, puis en carcinomes *in situ* et finalement en carcinomes métastatiques – atteste de la progression par étape de ce cancer, qui se produit souvent sur de nombreuses années. Pendant ce développement séquentiel, des mutations somatiques se produisent dans un ordre prévisible, jouant également un rôle clé dans le développement du



CCR, notamment des mutations précoces du gène *APC*, souvent suivies de mutations dans les gènes de la famille *RAS* et des altération du gène *TP53* plus tardivement [420, 421].

Enfin l'inflammation chronique, d'origine infectieuse ou non, induit de nombreux processus associés à la cancérogenèse, notamment la production de monoxyde d'azote (NO), d'espèces réactives de l'oxygène (ROS), et l'élévation des niveaux de la cyclo-oxygénase inductible COX2. Ces processus entraînent notamment des dommages directs de l'ADN, l'inhibition de l'apoptose, la stimulation de la prolifération, l'augmentation de l'angiogenèse et une immunodépression [422].

Le lien entre infection et cancer fait l'objet de nombreuses études depuis les années quatre-vingts. Parmi les agents infectieux incriminés les plus étudiés, on trouve les papillomavirus qui induisent le cancer du col de l'utérus, le virus d'Epstein-Barr impliqué dans divers lymphomes, les virus des hépatites B et C qui sont responsables de cancers hépatiques, le parasite *Schistosoma haematobium* impliqué dans le cancer de la vessie, *Salmonella enterica* serovar Typhi liée au cancer des voies biliaires et *H. pylori* responsable de cancers et de lymphomes gastriques. Les mécanismes exacts déclenchant la formation du tissu malin ne sont pas toujours bien compris [423-425]. Cependant, l'inflammation chronique engendrée par la persistance des infections bactériennes [422], et la production de toxines et de métabolites secondaires par l'agent infectieux seraient responsables de dommages à l'ADN, de modifications du cycle et de la prolifération cellulaire et d'échappement tumoral au système immunitaire, qui conduisent au CCR [426]. Les agents infectieux seraient ainsi responsables d'environ 18% des cancers [423]. Une observation fascinante reliant les bactéries au développement du CCR a été faite il y a plus d'une décennie grâce des animaux gnotobiotiques<sup>11</sup>. Dove *et al.* ont observé que le développement des tumeurs chez des souris *Min* (mutées pour le gène *APC*) était réduit de moitié par rapport à celle des souris élevées dans des conditions exemptes de microbes [427].

Nos études épidémiologiques portaient principalement sur la prévalence des cyclomodulines et en particulier de l'îlot *pks* chez des souches de *E. coli* isolées de biopsies de tumeurs coliques. Cette étude a mis en évidence une association entre l'îlot *pks* et les cancers coliques : la prévalence de l'îlot *pks* était significativement supérieure dans la série de tumeurs coliques comparé aux autres origines cliniques ( $p$ , <0,001 à 0,009). Cependant notre étude comporte plusieurs limites. La taille de la cohorte était petite. Le développement du

---

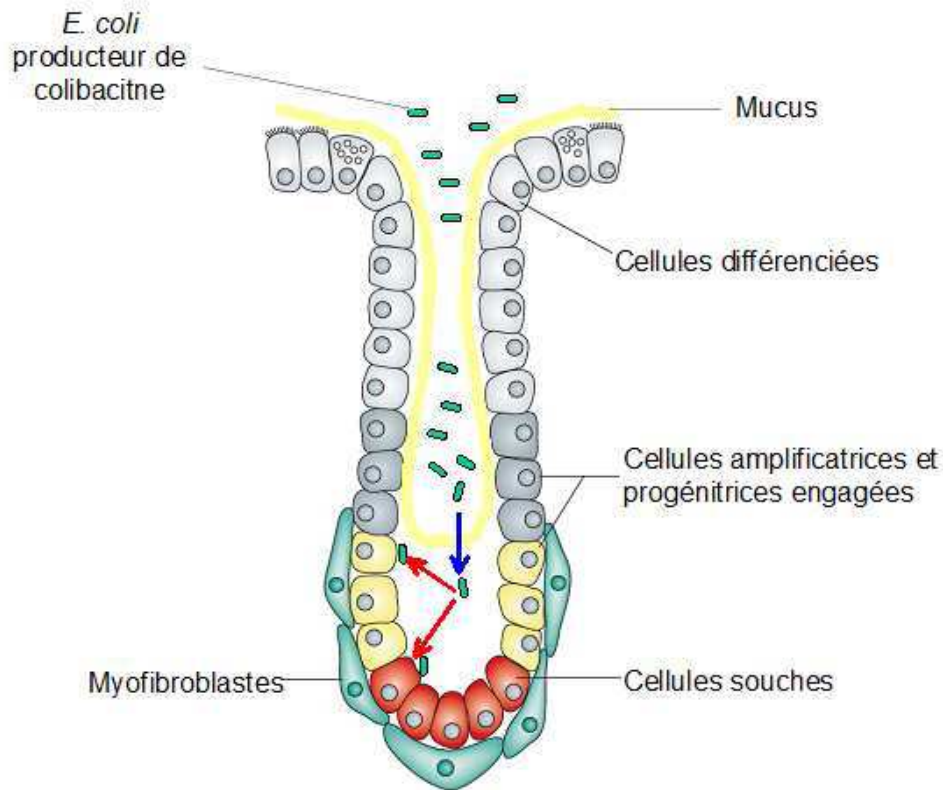
<sup>11</sup> Animaux nés et maintenus en l'absence de micro-organismes

CCR est un processus long. Le délai entre la colonisation bactérienne, et l'apparition du cancer peut demander des années, voire des dizaines d'années. La présence d'isolats de *E. coli* possédant l'îlot *pks* au moment du prélèvement au sein d'une biopsie tumorale ne présume donc pas de leur présence plusieurs années avant, ni de leur implication dans le processus tumoral. Clarifier le rôle possible des micro-organismes porteurs de l'îlot *pks* dans le CCR nécessite donc sur le plan épidémiologique une étude observationnelle et longitudinale de la population et la prise en compte des facteurs génétiques et environnementaux associés au risque de CCR. En outre, un îlot *pks* fonctionnel ayant été rapporté chez d'autres espèces d'Entérobactéries [246], notamment commensales, celui-ci devrait être recherché sans restriction d'espèce. Vu l'important réservoir microbien que constitue le côlon humain, une « association de malfaiteurs » correspondant à différentes toxines comme les cyclomodulines, à d'autres agents microbiens ou à une association microbe-agent chimique peut être impliquée dans la tumorigenèse.

L'îlot *pks* a une forte prévalence dans les souches de *E. coli* commensales (12 à 32%) et la colibactine est produite presque exclusivement par des *E. coli* du groupe phylogénétique B2 [24, 246, 247, 428], dont la proportion dans les flores intestinales humaines est en augmentation depuis ces 20 dernières années en Europe, au Japon et aux Etats-Unis [429], régions où la prévalence du CCR est élevée [415]. Les souches de *E. coli* du groupe B2 sont d'excellents colonisateurs et il a été observé qu'elles pouvaient persister plus durablement dans le côlon que d'autres souches de *E. coli* [430]. Cette prédominance du groupe B2 s'observe même dès la naissance puisque plus de 50% des souches de *E. coli* isolées à partir de selles d'enfants sont des B2 [431], et l'analyse du microbiome de 13 Japonais en bonne santé a montré que les gènes codant pour la colibactine sont présents dans près de 40% des selles des individus étudiés [432]. Il apparaît donc que les *E. coli* possédant l'îlot *pks* sont très présents dans les flores intestinales humaines et ont la capacité de coloniser l'homme dès la naissance puis de se maintenir durablement dans le côlon. Le fait d'héberger une souche de *E. coli* produisant la colibactine pourrait donc représenter une exposition génotoxique chronique au niveau de l'épithélium intestinal. Même à un faible niveau, cette génotoxicité pourrait être significative car une seule cassure double-brin de l'ADN peut déclencher des changements dégénératifs au niveau chromosomique [433]. Ainsi chez l'individu génétiquement susceptible, souffrant de problèmes intestinaux chroniques, immunodéficient ou ayant des habitudes alimentaires favorisant la flore colibacillaire, la colibactine pourrait jouer un rôle dans le développement du cancer.

Cuevas-Ramos *et al* ont montré que la colibactine produite par des souches de *E. coli* possédant l'îlot *pks* possédait des propriétés génotoxiques *in vitro* et *in vivo* [250]. Une infection de quelques heures d'entérocytes dans un modèle d'anse intestinale ligaturée de souris par une souche de *E. coli* produisant la colibactine induisait des cassures double-brin de l'ADN correspondant à une irradiation de 0,5 Grays [250]. Les lignées cellulaires humaines infectées transitoirement par des *E. coli* produisant la colibactine montrent aussi des lésions chromosomiques et la formation de colonies cellulaires indépendantes de l'ancrage, un processus intervenant dans le développement de métastases [250]. Ces souches sont donc mutagènes, induisent des dommages chromosomiques et une instabilité génomique *in vitro* de manière dose-dépendante. Cette instabilité génomique est retrouvée dans 85% des cancers colorectaux sporadiques [420]. L'instabilité génomique est proposée depuis longtemps comme une constante de la transformation cancéreuse.

Il faut néanmoins souligner que l'effet génotoxique de la colibactine en culture n'est observable que lorsque la bactérie productrice est en contact direct avec la cellule hôte [24]. Par ailleurs, le modèle de carcinogenèse colorectale communément admis propose que les cellules tumorales colorectales proviennent de mutations des cellules souches ou de leurs progéniteurs précoces [434, 435]. L'organisation histologique du côlon montre que les cellules souches et les cellules progénitrices en division rapide sont localisées à la base des cryptes, un site normalement stérile. Pour être un facteur initiateur du processus tumoral, la bactérie productrice de colibactine doit donc accéder à cette zone (**figure 37**). Par ailleurs, la colibactine pourrait aussi être un facteur promoteur de tumeur. Les cellules souches épithéliales ont besoin de temps pour accumuler des mutations et se transformer progressivement. Avec un dommage à l'ADN constant et répété, consécutif à la présence continue de bactéries génotoxiques, il est possible que ces cellules puissent acquérir les changements nécessaires à leur transformation.



**Figure 37 – Modèle hypothétique d’initiation de la tumorigenèse colique par la colibactine.**

Pour que la colibactine soit un facteur initiateur du processus tumoral colique, les souches de *E. coli* commensales du groupe phylogénétique B2 portant l’îlot *pks* devraient :(1) franchir la couche de mucus interne si celle-ci n’est pas préalablement altérée, et (2) par contact cellulaire, délivrer ensuite la colibactine au niveau des cellules souches ou des cellules progénitrices épithéliales des cryptes coliques.

Une lymphotoxicité de souches de *E. coli* produisant la colibactine a été observée sur des cultures primaires et dans un modèle murin de sepsis - aggravation de la lymphopénie et apoptose massive des thymocytes [414]. Ces souches de *E. coli* pourraient favoriser l’échappement des cellules tumorales à la surveillance du système immunitaire, les lymphocytes semblant plus sensibles à l’action de la colibactine que les cellules épithéliales tumorales.

Nos résultats épidémiologiques associés aux études expérimentales réalisées à ce jour ne permettent pas encore d’affirmer l’existence d’un rôle des souches mutagènes de *E. coli* dans la cancérogenèse. Il apparaît néanmoins plausible que des souches de *E. coli* possédant l’îlot *pks* puissent modifier le renouvellement et la physiologie de l’épithélium intestinal pour induire ou favoriser à long terme un cancer.

Des expérimentations animales pourraient aider à cerner le rôle potentiel de la colibactine dans la carcinogenèse colorectale. Les souris *Min* semblent intéressantes car la

colonisation de leur tube digestif par des *Bacteroides fragilis* entérotoxigéniques déclenche une colite inflammatoire avec une promotion rapide de tumeurs coliques distales, reproduisant le cancer colique humain [436]. L'évaluation de l'effet de la colibactine seule ou en association avec des *B. fragilis* entérotoxigéniques sur le tube digestif de ces souris peut apporter de nouveaux éléments de réponses. Elle pourrait favoriser l'émergence de cellules tumorales mais aussi favoriser l'élimination des cellules initiées ou transformées, car ces dernières sont en général plus sensibles aux dommages à l'ADN.

D'autres organes que le côlon sont aussi exposés aux ExPEC produisant la colibactine. Il serait intéressant d'étudier l'association et l'impact de ces bactéries génotoxiques dans les cancers de l'arbre urinaire. En effet, le principal microorganisme isolé dans les infections urinaires humaines (80 à 90%) est *E. coli*. La prévalence de souches de *E. coli* issues d'urosepsis portant l'îlot *pks* était de 32% dans une de nos études [428]. Les UPEC sont connues pour former des colonies bactériennes intracellulaires [437, 438]. Pour cette raison, elles sont difficiles à éradiquer et elles provoquent ainsi des infections récurrentes. Les individus ayant présenté plus de deux infections urinaires ont un risque plus important d'avoir un carcinome de la vessie, principalement chez les femmes [439]. *E. coli* peut induire une hyperplasie réactionnelle dans un modèle murin de prostatite chronique [440] et a été proposé comme agent étiologique de l'inflammation chronique carcinogène dans le cancer de la prostate [441]. La colibactine pourrait jouer un rôle chez les individus génétiquement susceptibles, où un polymorphisme génétique a été associé au cancer de la vessie [442]. Les cellules des individus génétiquement susceptibles sont en général affectées dans les systèmes de réparation de l'ADN et sont donc moins efficaces pour réparer des cassures double-brin. Face à des infections chroniques et/ou récurrentes à *E. coli* produisant la colibactine, l'épithélium du tractus urinaire pourrait initier ou accélérer un processus tumoral.

En conclusion, le cancer gastrique et *H. pylori* est la meilleure association étudiée entre une infection bactérienne et le cancer. Il est actuellement établi que d'autres bactéries pathogènes ou « commensales » peuvent avoir un rôle direct dans la carcinogénèse, notamment colorectale. Comprendre le lien entre cancer et bactérie(s) est un enjeu majeur pour la prévention de ces pathologies. Cependant, les mécanismes impliqués sont multifactoriels et font intervenir la réponse inflammatoire, des effecteurs bactériens, la susceptibilité de l'hôte et probablement des facteurs environnementaux dont l'exploration reste un défi pour les années à venir.

## VII. Autres perspectives relatives à l'étude de l'îlot *pks*

L'absence d'activité cytopathogène dans le surnageant des souches de *E. coli* ayant un îlot *pks* fonctionnel suggère que la colibactine présente une caractéristique physico-chimique particulière la retenant dans une structure bactérienne externe (membrane externe, capsule), et/ou qu'elle soit instable dans le milieu extérieur. Le lipopeptide bacillaene produit par *B. subtilis* est sensible à la lumière et à la température [306]. Les structures oléfines<sup>12</sup> de ce composé sont probablement responsables de son instabilité. Des éléments structuraux équivalents pourraient être générés dans la colibactine par les réductases de la mégasynthase ClbB et/ou du complexe enzymatique probable ClbD-G, et donc engendrer des problèmes de stabilité. Ces problèmes et la nature probablement lipopeptidique de la colibactine doivent être intégrés dans les approches visant à purifier la colibactine et à déterminer sa structure.

Chez *B. subtilis*, les mégasynthèses synthétisant la surfactine présentent une organisation diffuse dans le cytoplasme alors que celles produisant le bacillaene s'assemblent en une structure unique de 10 à 100 MDa associée à la membrane [392]. Cette organisation serait mieux adaptée à la synthèse de composés instables en favorisant leur protection au sein d'une structure qui s'apparente à un organite. La localisation et l'organisation de la chaîne d'assemblage de la colibactine reste à étudier.

La colibactine induit des cassures double-brin de l'ADN, qui peuvent résulter soit d'une action directe de la colibactine sur l'ADN comme pour de nombreux composés NRPK tels que la bléomycine ou la thiocoraline, soit d'une interférence avec les systèmes impliqués dans la synthèse ou la réparation de l'ADN. Le composé ZmA, qui est maturé par une enzyme proche de ClbP interagirait avec les sous-unités de l'ARN polymérase car des mutations dans les gènes correspondants induisent une résistance à son action antibiotique chez *E. coli* [443]. Le mécanisme d'action de la colibactine reste à explorer sur ce plan. Une lignée cellulaire eucaryote haploïde combinée à une mutagénèse aléatoire par transposons pourrait être utilisée pour identifier des facteurs de l'hôte nécessaires à l'action de la colibactine (récepteur, voie d'entrée dans la cellule eucaryote et cibles), comme déjà effectué pour la toxine diphtérique et la CM CDT de *E. coli* [444].

---

<sup>12</sup> Liaisons carbone-carbone insaturées, qui, lorsqu'elles sont conjuguées sont fréquemment responsables de réactions de dégradation et de destruction.

## VIII. Conclusion

L'îlot génomique *pks* de *Escherichia coli* et d'autres *Enterobacteriaceae* code des NRPK synthases qui permettent l'assemblage d'un composé hybride PK-NRP putatif nommé colibactine, qui induit des cassures double-brin de l'ADN des cellules eucaryotes.

Nous avons montré que la protéine ClbP, essentielle à l'activité génotoxique de l'îlot *pks*, possède une partie N-terminale catalytique et périplasmique, et une partie C-terminale associée à la membrane cytoplasmique. La structure cristalline de ClbP et des expériences de mutagenèse ont révélé un site actif à sérine et des caractéristiques structurales originales compatibles avec une activité peptidase, confirmée par des analyses biochimiques. Dix homologues de ClbP ont été identifiés *in silico* dans des îlots génomiques de synthases de peptides non ribosomiaux d'espèces bactériennes proches et éloignées. Tous les homologues testés ont présenté une promiscuité fonctionnelle avec ClbP. Nous avons proposé finalement de désigner cette nouvelle sous-famille de peptidases, probablement impliquée dans la maturation finale de composés NRP bioactifs chargés positivement, les peptidases associées aux NRP (PaNRP). Cette fonction devra être confirmée en isolant la colibactine et/ou son précurseur direct, ou en étudiant d'autres PaNRP codées dans d'autres îlots génomiques.

Par ailleurs, nous avons exploré le rôle de l'îlot *pks* au travers de deux études épidémiologiques des cyclomodulines de *E. coli*, dans deux contextes physiopathologiques, l'urosepsis et les cancers coliques et rectaux. L'îlot *pks* était significativement associé aux souches issues d'urosepsis comparé à des souches commensales, et aux souches issues de biopsies de tumeurs coliques comparé à des souches commensales et issues de biopsies de tumeurs rectales, de diverticuloses et de lésions iléales de maladie de Crohn. Le rôle de la colibactine devra néanmoins être investigué par d'autres études épidémiologiques et à l'aide de modèles expérimentaux animaux reproduisant au mieux les conditions physiopathologiques.

**Chapitre IV**

**Autres  
publications**





**I. Article 4 – CTX-M et virulence de *E. coli* K1**

**CTX-M  $\beta$ -lactamase production and virulence  
of *Escherichia coli* K1**

Damien Dubois, Nemani V. Prasadarao, Rahul Mittal, Laurent Bret, Marie Roujou-Gris et  
Richard Bonnet

*Emerging Infectious Diseases* (2009), Vol. **15** (12), p1988-90.

# CTX-M β-Lactamase Production and Virulence of *Escherichia coli* K1

Damien Dubois, Nemani V. Prasadarao,  
Rahul Mittal, Laurent Bret, Marie Roujou-Gris,  
and Richard Bonnet

We report a patient with neonatal meningitis caused by a CTX-M-1–producing *Escherichia coli* K1 strain. The influence of CTX-M production on virulence was investigated in cell culture and a newborn mouse model of meningitis. CTX-M production had no influence on virulence but was a major factor in clinical outcome.

*Escherichia coli* is the second most common cause of neonatal meningitis. Neonatal meningitis *E. coli* (NMEC) belong mainly to phylogenetic group B2 and harbor numerous virulence factors (1).

Since the beginning of the 21st century, an explosive spread of CTX-M–type extended-spectrum β-lactamases (ESBLs) in *E. coli* has occurred (2). These enzymes confer resistance to nearly all β-lactam antimicrobial drugs, including third-generation cephalosporins, the first-line treatment for patients with serious *E. coli* infections. However, CTX-M–type ESBLs have been observed mainly in *E. coli* strains with few virulence factors or in strains causing minor infections (3–5). In addition, bacterial resistance to antimicrobial drugs is frequently reported as difficult to reconcile with bacterial virulence (6). Highly pathogenic *E. coli* such as NMEC are therefore considered susceptible to antimicrobial drugs (7). We report a clinical case of neonatal meningitis caused by CTX-M–producing NMEC and the influence of CTX-M production on virulence.

## The Study

In April 2007, a 39-year-old pregnant woman with amniotic sac rupture was admitted to a hospital in Orleans, France, at 28 weeks and 4 days of gestation. Treatment

Author affiliations: University of Auvergne, Clermont-Ferrand, France (D. Dubois, R. Bonnet); Teaching Hospital, Clermont-Ferrand (D. Dubois, R. Bonnet); Childrens Hospital, Los Angeles, California, USA (N.V. Prasadarao, R. Mittal); University of Southern California, Los Angeles (N.V. Prasadarao); and La Source Hospital, Orléans, France (L. Bret, M. Roujou-Gris)

DOI: 10.3201/eid1512.090928

was started with betamethasone (12 mg 1×/d) for fetal lung maturation and amoxicillin (1g 3×/d) for 4 days. Because of a high serum level of C-reactive protein, antimicrobial drug therapy was switched to amoxicillin with clavulanic acid (1g 3×/d) for 1 day. A cesarean delivery was performed at 29 weeks and 2 days of gestation. A lumbar puncture sample of the low-weight (1,560 g) newborn female was tinged with blood. Cerebrospinal fluid (CSF) protein and glucose values were 4.00 g/L and 3.5 mmol/L, respectively. Results of CSF Gram staining were negative.

The infant was admitted to the neonatal critical care unit and received amoxicillin (150 mg), cefotaxime (120 mg), and netilmicin (8 mg) 2×/d for 2 days. Culture of placenta, maternal and infant blood, and infant gastric fluid yielded *E. coli*. The isolate was resistant to antimicrobial drugs, including third-generation cephalosporins. Imipenem/cilastatin (25 mg 4×/d) and amikacin (15 mg 2×/d) were given for 2 days. Treatment with imipenem/cilastatin was given for 15 days and then stopped because of the infant's clinical improvement and return of C-reactive protein to the reference level. Similar drug treatment was administered to the mother.

One week after drug treatment was discontinued, the infant showed signs of septicemia. A second lumbar puncture sample had protein and glucose levels of 4.56 g/L and 0.1 mmol/L, respectively, and a leukocyte count of 4,700 cells/μL (54% polymorphonuclear cells). *E. coli* were isolated from blood and CSF cultures and showed a resistance pattern identical to that of the previous isolate. Meningitis was a complication of the initial sepsis or a relapse of initial unapparent meningitis (8).

Treatment was started with imipenem/cilastatin (30 mg 4×/d) for 25 days and amikacin (15 mg 2×/d) for 5 days. Because the infant had a seizure, phenobarbital (22.5 mg) and ciprofloxacin (15 mg 2×/d) were prescribed for 5 additional days. Her condition gradually improved and blood and CSF values returned to reference levels. The infant was discharged from the hospital 1 month later and treatment with the anticonvulsant was discontinued. She showed normal psychomotor development at a regular follow-up pediatric visit.

*E. coli* strains isolated from the mother and infant were indistinguishable by enterobacterial repetitive intergenic consensus sequence 2 PCR, random amplified polymorphic DNA analysis, and typing with a MALDI BioTyper (Bruker Daltonique, Wissembourg, France) (9). Thus, the isolates corresponded to the same strain, designated Or1-1. PCR-based phylogenetic analysis and serotyping showed that the strain belonged to group B2 and serotype O7:K1:H7, a major O antigen encountered worldwide in NMEC (1).

Resistance of the Or1-1 strain (MICs 128 μg/mL for cefotaxime and 8 μg/mL for ceftazidime) was caused by

the gene encoding CTX-M-1. This strain was also resistant to tetracycline, trimethoprim, and sulfamethoxazole and susceptible to cefoxitin, imipenem, aminoglycosides, quinolones, chloramphenicol, and fosfomycin. It harbored the major *E. coli* genes associated with neonatal meningitis (Table 1) (1,10).

Plasmids from Orl-1 were used to transform *E. coli* K-12 DH5 $\alpha$ . Three resistance profiles that enabled detection of 3 plasmids were obtained. On the basis of screening of plasmid transformants and Orl-1, most virulence factors genes were presumably chromosomally encoded. Three virulence factors (*aer*, *iss*, and a second copy of *iroN*) were mediated by a tetracycline-resistant, large ( $\approx$ 180 kb), conjugative plasmid (pOrl-1-Te). The 2 other plasmids were pOrl-1-CTX-M-1, the CTX-M-1-encoding large ( $\approx$ 150 kb) conjugative plasmid carrying resistance to trimethoprim and sulfamethoxazole, and pOrl-1-TEM-1, a TEM-1-encoding small (<40 kb) plasmid.

A derivative strain that did not harbor the 3 plasmids (Orl-c) was obtained by plasmid elimination with ethidium bromide. Orl-1, Orl-c, and *E. coli* DH5 $\alpha$  harboring pOrl-1-CTX-M-1 were tested for invasiveness in human brain microvascular endothelial cells (10) and in a newborn mouse (C57BL/6 wild-type) model of meningitis (R. Mittal et al., unpub. data) to investigate the influence of CTX-M-1 production on virulence. *E. coli* strain E44, a rifampicin-resistant mutant of archetypical NMEC K1 strain RS218, was used as a positive control (10).

Orl-1 and Orl-c strains exhibited 3.5 $\times$  lower invasiveness than strain E44. However, their ability to invade human brain microvascular endothelial cells was 400 $\times$  higher than that of strain DH5 $\alpha$ -CTX-M1. In the mouse model, DH5 $\alpha$ -CTX-M1 did not cause bacteremia or meningitis. In contrast, Orl-1, Orl-c, and E44 induced meningitis (prevalences of 100%, 84%, and 85%, respectively) (Table 2).

The difference between Orl-1 and Orl-c in the mouse model may be explained by loss of plasmid pOrl-1-Te from Orl-c. Plasmid pOrl-1-Te is likely similar to pS88-related plasmids of NMEC because they share 3 virulence factor genes (*iss*, *aer*, and *iroN*) and are large and conjugative. These plasmids contribute to virulence of NMEC (11). Orl-1 and Orl-c showed similar behaviors, which suggested that the CTX-M-1-encoding plasmid pOrl-1-CTX-M-1 does not alter virulence of the strain.

Table 1. Virulence factors of Orl-1 *Escherichia coli* K1 strain, France

Integrative elements* (1)	Virulence genes	Present
PAI III <sub>536</sub> -like	<i>iroN</i>	Yes
	<i>sfa/foc</i>	Yes
GimA-like	<i>ibeA (gimA4)</i> , <i>ptnC (gimA1)</i>	Yes
PAI II <sub>J96</sub> -like	<i>hra</i>	Yes
	<i>hlyC</i> , <i>cnf1</i>	No
PAI I <sub>CF7073</sub> -like	<i>hlyC</i>	No
	<i>aer (iucC)</i>	Yes
HPI-like	<i>fyuA</i> , <i>irp-2</i>	Yes
GimB-like	<i>gimB</i>	Yes
<i>pks</i> island†	<i>clbA</i> , <i>clbK-J</i> , <i>clbP</i> , <i>clbQ</i>	Yes
Others	<i>chuA</i>	Yes
	<i>ompA</i>	Yes
	<i>hek</i>	Yes
	<i>iss</i>	Yes
	<i>malX</i>	Yes
	<i>cdtB-I</i> to <i>-V</i>	No

\*PAI, pathogenicity island; HPI, high-pathogenicity island.

†Positive cytopathogen effect with transient infection of HeLa cells.

Mice with E44- and Orl-1-induced neonatal meningitis were treated with the third-generation cephalosporin cefotaxime, as recommended for humans. Despite antimicrobial drug treatment, Orl-1, but not strain E44, caused meningitis, suggesting that drug resistance is a major factor in clinical outcomes.

### Conclusions

Studies have reported emergence of *E. coli* as the predominant organisms responsible for sepsis at any gestational age and for increased rates of drug-resistant *E. coli* caused by intrapartum drug prophylaxis (12). Spread of ESBLs in *E. coli* and intrapartum exposure to antimicrobial drugs may favor emergence of NMEC strains resistant to third-generation cephalosporins.

Two other well-characterized *E. coli* K1 strains producing ESBLs have been isolated from patients with neonatal meningitis in Algeria and France. The ESBL was identified as CTX-M-15 in both patients, and 1 infection was lethal (13,14). Other putative ESBL-producing *E. coli* K1 have been recently isolated, especially in developing countries (15).

Emergence of ESBL-producing *E. coli* strains, which are frequently resistant to fluoroquinolone (2), highlights the need for possible alternatives to third-generation ce-

Table 2. Incidence of meningitis in a newborn mouse model by *Escherichia coli* strain, France\*

Bacterial strain	No. animals	Mean $\pm$ SD bacteremia, log CFU/mL blood	No. positive CSF cultures (% meningitis)
E44	20	6.95 $\pm$ 0.6	17 (85)†
Orl-1	16	6.75 $\pm$ 0.8	16 (100)†
Orl-c	17	6.60 $\pm$ 0.5	14 (82)†
DH5 $\alpha$ -CTX-M1	10	0.10 $\pm$ 0.1	0

\*CSF, cerebrospinal fluid.

†p<0.005, significantly higher than the incidence of meningitis by DH5 $\alpha$ -CTX-M1 by  $\chi^2$  test.

phalosporins for treatment of patients with infected with NMEC. Carbapenems are usually recommended for treatment of infections with ESBL-producing *E. coli* (2). However, this case report shows the role of treatment duration and the need for additional pharmacokinetic and safety studies in neonates and for adjunctive therapies (8).

This characterization of a CTX-M-1–producing NMEC strain highlights the emergence of CTX-M–type ESBL in highly virulent *E. coli*. Because of worldwide spread of CTX-Ms, caution should be exercised in the management of patients with NMEC, and first-line treatment for neonatal meningitis may need to be reconsidered.

#### Acknowledgments

We thank Marlène Jan and Rolande Perroux for technical assistance.

This study was supported by le Ministère Français de l'Éducation Nationale, de la Recherche et de la Technologie grant JE2526, l'Institut National de la Recherche Agronomique grant USC-2018 to R.B., and National Institutes of Health grant AI 40567 to N.V.P.

Dr Dubois is a medical microbiologist at the Teaching Hospital in Clermont-Ferrand, France. His main research interests are pathogenesis and drug resistance of *E. coli*.

#### References

- Bonacorsi S, Clermont O, Houdouin V, Cordevant C, Brahimi N, Marecat A, et al. Molecular analysis and experimental virulence of French and North American *Escherichia coli* neonatal meningitis isolates: identification of a new virulent clone. *J Infect Dis*. 2003;187:1895–906. DOI: 10.1086/375347
- Pitout JD, Laupland KB. Extended-spectrum beta-lactamase-producing *Enterobacteriaceae*: an emerging public-health concern. *Lancet Infect Dis*. 2008;8:159–66. DOI: 10.1016/S1473-3099(08)70041-0
- Lavigne JP, Blanc-Potard AB, Bourg G, Moreau J, Chanal C, Bouziges N, et al. Virulence genotype and nematode-killing properties of extra-intestinal *Escherichia coli* producing CTX-M beta-lactamases. *Clin Microbiol Infect*. 2006;12:1199–206. DOI: 10.1111/j.1469-0691.2006.01536.x
- Pitout JD, Laupland KB, Church DL, Menard ML, Johnson JR. Virulence factors of *Escherichia coli* isolates that produce CTX-M-type extended-spectrum beta-lactamases. *Antimicrob Agents Chemother*. 2005;49:4667–70. DOI: 10.1128/AAC.49.11.4667-4670.2005
- Branger C, Zamfir O, Geoffroy S, Laurans G, Arlet G, Thien HV, et al. Genetic background of *Escherichia coli* and extended-spectrum beta-lactamase type. *Emerg Infect Dis*. 2005;11:54–61.
- Andersson DI. The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr Opin Microbiol*. 2006;9:461–5. DOI: 10.1016/j.mib.2006.07.002
- Houdouin V, Bonacorsi S, Bidet P, Bingen E. Antimicrobial susceptibility of 136 *Escherichia coli* isolates from cases of neonatal meningitis and relationship with virulence. *Clin Microbiol Infect*. 2007;13:1207–10. DOI: 10.1111/j.1469-0691.2007.01838.x
- Galiza EP, Heath PT. Improving the outcome of neonatal meningitis. *Curr Opin Infect Dis*. 2009;22:229–34. DOI: 10.1097/QCO.0b013e32832ad49e
- Jonas D, Spitzmuller B, Weist K, Ruden H, Daschner FD. Comparison of PCR-based methods for typing *Escherichia coli*. *Clin Microbiol Infect*. 2003;9:823–31. DOI: 10.1046/j.1469-0691.2003.00661.x
- Prasadarao NV, Wass CA, Weiser JN, Stins MF, Huang SH, Kim KS. Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. *Infect Immun*. 1996;64:146–53.
- Peigne C, Bidet P, Mahjoub-Messai F, Plainvert C, Barbe V, Mediague C, et al. The plasmid of neonatal meningitis *Escherichia coli* strain S88 (O45:K1:H7) is closely related to avian pathogenic *E. coli* plasmids and is associated with high level bacteremia in a neonatal rat meningitis model. *Infect Immun*. 2009;77:2272–84. DOI: 10.1128/IAI.01333-08
- Bizzarro MJ, Dembry LM, Baltimore RS, Gallagher PG. Changing patterns in neonatal *Escherichia coli* sepsis and ampicillin resistance in the era of intrapartum antibiotic prophylaxis. *Pediatrics*. 2008;121:689–96. DOI: 10.1542/peds.2007-2171
- Boyer-Mariotte S, Duboc P, Bonacorsi S, Lemeland JF, Bingen E, Pinquier D. CTX-M-15–producing *Escherichia coli* in fatal neonatal meningitis: failure of empirical chemotherapy. *J Antimicrob Chemother*. 2008;62:1472–4. DOI: 10.1093/jac/dkn362
- Ramdani-Bouguessa N, Mendonca N, Leitao J, Ferreira E, Tazir M, Canica M. CTX-M-3 and CTX-M-15 extended-spectrum beta-lactamases in isolates of *Escherichia coli* from a hospital in Algiers, Algeria. *J Clin Microbiol*. 2006;44:4584–6. DOI: 10.1128/JCM.01445-06
- Sehgal R, Gaiind R, Chellani H, Agarwal P. Extended-spectrum beta lactamase-producing gram-negative bacteria: clinical profile and outcome in a neonatal intensive care unit. *Ann Trop Paediatr*. 2007;27:45–54. DOI: 10.1179/146532807X170501

Address for correspondence: Richard Bonnet, Laboratoire de Bactériologie, Faculté de Médecine, Université d'Auvergne, 28 Place Henri Dunant, Clermont-Ferrand, F-63001, France; email: rbonnet@chu-clermontferrand.fr

All material published in Emerging Infectious Diseases is in the public domain and may be used and reprinted without special permission; proper citation, however, is required.

Search past issues of EID at [www.cdc.gov/eid](http://www.cdc.gov/eid)

**II. Article 5 – Identification des *Staphylococcus* par spectrométrie de masse**

**Identification of a variety of *Staphylococcus* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry**

Damien Dubois, David Leysse, Jean-Paul Chacornac, Markus Kostrzewa, Pierre-Olivier Schmit, Régine Talon, Richard Bonnet et Julien Delmas

*J Clin Microbiol* (2010) Mar, Vol. **48** (3), p941-5.

## Identification of a Variety of *Staphylococcus* Species by Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry<sup>∇</sup>

Damien Dubois,<sup>1,2</sup> David Leyssene,<sup>1,2</sup> Jean Paul Chacornac,<sup>3</sup> Markus Kostrzewa,<sup>4</sup>  
Pierre Olivier Schmit,<sup>5</sup> Régine Talon,<sup>3</sup> Richard Bonnet,<sup>1,2</sup> and Julien Delmas<sup>1,2\*</sup>

CHU Clermont-Ferrand, Centre de Biologie, Laboratoire de Bactériologie Clinique, Clermont-Ferrand F-63003, France<sup>1</sup>;  
Université d'Auvergne Clermont-1, Faculté de Médecine, Laboratoire de Bactériologie, JE2526, USC-INRA 2018,  
Clermont-Ferrand F-63001, France<sup>2</sup>; INRA, Centre de Clermont-Ferrand-Theix, UR 454, Microbiologie,  
63122 Saint-Genès Champanelle, France<sup>3</sup>; Bruker Daltonik GmbH, Leipzig, Germany<sup>4</sup>; and  
Bruker Daltonique, Wissembourg, France<sup>5</sup>

Received 25 February 2009/Returned for modification 5 August 2009/Accepted 11 December 2009

**Whole-cell fingerprinting by matrix-assisted laser desorption ionization–time-of-flight mass spectrometry (MALDI-TOF MS) in combination with a dedicated bioinformatic software tool (MALDI Biotyper 2.0) was used to identify 152 staphylococcal strains corresponding to 22 staphylococcal species. Spectra of the 152 isolates, previously identified at the species level using a *sodA* gene-based oligonucleotide array, were analyzed against the main spectra of 3,030 microorganisms. A total of 151 strains out of 152 (99.3%) were correctly identified at the species level; only one strain was identified at the genus level. The MALDI-TOF MS method revealed different clonal lineages of *Staphylococcus epidermidis* that were of either human or environmental origin, which suggests that the MALDI-TOF MS method could be useful in the profiling of staphylococcal strains. The topology of the dendrogram generated by the MALDI Biotyper 2.0 software from the spectra of 120 *Staphylococcus* reference strains (representing 36 species) was in general agreement with that inferred from the 16S rRNA gene-based analysis. Our findings indicate that the MALDI-TOF MS technology, associated with a broad-spectrum reference database, is an effective tool for the swift and reliable identification of *Staphylococci*.**

Most staphylococci are harmless and reside normally on the skin and mucous membranes of humans and other organisms (16, 22, 34). Staphylococcal strains are isolated from various food products in which they are involved in fermentation (18, 29). *Staphylococcus* species can cause a wide variety of diseases in humans and other animals (2, 22, 30–32, 35). *S. aureus* is a major pathogen in human infections (31). Several other *Staphylococcus* species have also been implicated in human infections, notably *S. saprophyticus*, *S. epidermidis*, *S. lugdunensis*, and *S. schleiferi* (4, 16, 31, 34). Coagulase-negative staphylococci (CoNS) have emerged as predominant pathogens in hospital-acquired infections (4, 16, 31, 34). One of the major challenges of daily diagnostic work is therefore to identify *Staphylococcus* species.

Several manual and automated methods based on phenotypic characteristics have been developed for the identification of *Staphylococci* (12, 24). Unfortunately, these systems have their limitations, mostly due to phenotypic differences between strains from the same species (6, 10, 19, 21). Over the last 10 years, many genotypic methods based on the analysis of selected DNA targets have been designed for species-level identification of most common isolated CoNS (20, 26, 33). The sequence polymorphism of the *sodA* gene has significant discriminatory power (20) and allows the development of assays based on DNA chip technologies (“Staph array”) (8). Recently, matrix-assisted laser desorption ionization–time of

flight mass spectrometry (MALDI-TOF MS) using protein “fingerprints” was used for the identification of microorganisms (1, 3, 5, 9, 11, 14, 25, 36). In the present study, we assessed the ability of the MALDI Biotyper system (Bruker Daltonique, Wissembourg, France) to identify *Staphylococcus* species of clinical and environmental origins previously identified by *sodA* gene-based oligonucleotide array (8).

### MATERIALS AND METHODS

**Bacterial strains.** A total of 152 blind-coded *Staphylococcus* strains from our collection, isolated from clinical ( $n = 60$ ) and food and plant samples ( $n = 92$ ) and representing 22 species, were studied (Table 1). All strains were previously identified as belonging to a specific species by the oligonucleotide “Staph array” (8). This system associates PCR amplification of the *sodA* gene with an oligonucleotide-based array to efficiently discriminate between the *Staphylococcus* species. Identifications found with this array were identical to those obtained by sequencing the internal fragment *sodA* (*sodA*<sub>int</sub>). In addition, the Vitek 2 system has been used with the Gram-positive (GP) identification card (bioMérieux, Marcy l’Etoile, France) to identify *Staphylococcus* strains. If the GP card indicated the possibility of two or three species, supplementary tests (pigmentation, hemolysis, or novobiocin resistance) were then performed to determine the identification.

**Sample preparation.** Sample preparation for MALDI-TOF MS was performed as previously described (14). Briefly, a few colonies of a fresh overnight culture grown on Columbia blood agar at 37°C under aerobic conditions were suspended in 300  $\mu$ l distilled water to which 900  $\mu$ l absolute ethanol was added. The mixture was centrifuged at 12,000  $\times g$  for 2 min, and the supernatant was discarded. Ten microliters of formic acid (70%) was added to the pellet and mixed thoroughly by pipetting before the addition of 10  $\mu$ l of acetonitrile to the mixture. The mixture was centrifuged at 12,000  $\times g$  for 2 min. One microliter of the supernatant was placed onto a steel target plate and air dried at room temperature. Each sample was overlaid with 2  $\mu$ l saturated solution of  $\alpha$ -cyano-4-hydroxycinnamic acid in 50% acetonitrile–2.5% trifluoroacetic acid and air dried at room temperature. MALDI-TOF MS measurements were performed with a Bruker Ultraflex II MALDI-TOF/TOF (tandem TOF) instrument equipped with 200-Hz Smartbeam laser technology. Spectra were recorded in the

\* Corresponding author. Mailing address: Laboratoire de Bactériologie, 28 place H. Dunant, 63001 Clermont-Ferrand, France. Phone: 33 4 73 75 49 20. Fax: 33 4 73 75 49 22. E-mail: julien.delmas@u-clermont1.fr.

<sup>∇</sup> Published ahead of print on 23 December 2009.



TABLE 1. Clinical and environmental staphylococcal strains used in this study

Species ("Staph array" identification)	No. of isolates		
	Clinical	Environmental	Total
<i>S. arlettae</i>	0	1	1
<i>S. aureus</i>	6	1	7
<i>S. capitis</i>	5	2	7
<i>S. caprae</i>	1	0	1
<i>S. carnosus</i>	0	7	7
<i>S. cohnii</i> subsp. <i>ureal</i>	2	2	4
<i>S. delphini</i>	1	0	1
<i>S. epidermidis</i>	11	8	19
<i>S. equorum</i>	0	13	13
<i>S. fleuretii</i>	0	2	2
<i>S. haemolyticus</i>	5	1	6
<i>S. hominis</i>	5	1	6
<i>S. lugdunensis</i>	5	0	5
<i>S. pasteurii</i>	0	8	8
<i>S. saprophyticus</i>	6	10	16
<i>S. schleiferi</i>	2	0	2
<i>S. sciuri</i>	1	2	3
<i>S. simulans</i>	4	0	4
<i>S. succinus</i>	0	9	9
<i>S. vitulinus</i>	0	5	5
<i>S. warneri</i>	6	6	12
<i>S. xylosum</i>	0	14	14
Total	60	92	152

positive linear mode within a mass range of 2,000 to 20,000 Da. Five hundred laser shots were recorded for each spectrum.

**Data analysis.** Raw spectra of the *Staphylococcus* strains were analyzed by MALDI Biotyper 2.0 software (Bruker Daltonique) with default settings. The whole process from MALDI-TOF MS measurement to identification was performed automatically without any user intervention. The peak lists generated were used for matches against the reference library directly using the integrated pattern-matching algorithm of the software. The software provides a log score, and the cutoff log score of 2 was used to validate identification at the species level, as recommended by the manufacturer. For strain classification, the creation of the dendrogram is based on cross-wise minimum spanning tree (MSP) matching. Similar MSPs result in a high matching score value. Each MSP is matched against all MSPs of the analyzed set. The list of score values is used to calculate normalized distance values between the analyzed species, resulting in a matrix of matching scores. The visualization of the respective relationship between the MSPs is displayed in a dendrogram using the standard settings of the MALDI Biotyper 2.0 software. Species with distance levels under 500 have been described as reliably classified (25).

## RESULTS AND DISCUSSION

**Classification of staphylococcal reference strains.** A score-oriented dendrogram was generated on the basis of the 120 reference strains in the Bruker MALDI Biotyper 2.0 database (Fig. 1). This dendrogram revealed, with the default critical distance level of 500, 4 of the 11 cluster groups defined by Poyart et al. from the phylogenetic analysis based on the *sodA* gene (20): *S. sciuri*, *S. warneri*, *S. lugdunensis*, and *S. auricularis* (Fig. 1). These cluster groups were also in accordance with those established by the phylogenetic analysis of the 16S rRNA gene (28). With a critical distance level of 750, the dendrogram revealed overall the cluster groups of 16S rRNA genes, with some variations (Fig. 1) (17, 28). The 16S rRNA gene-based cluster group *S. haemolyticus* including *S. haemolyticus* and *S. hominis* was divided into branches which linked at a distance level of 850. Unlike in other phylogenetic analyses, in which *S.*

*kloosii* belonged to *S. saprophyticus* cluster group, this species was clustered within the *S. simulans* group in the MALDI-TOF MS-based dendrogram (13, 20, 28). Inside the *S. saprophyticus* cluster group, the two strains *S. equorum* subsp. *equorum* DSM 20675 and DSM 20674 belonged to two distinct clusters. It is noteworthy that the *S. lutrae* and *S. schleiferi* species each constituted a well-defined clade inside the *S. hyicus-S. intermedius* cluster group, like *S. succinus*, *S. xylosum*, *S. cohnii*, and *S. arlettae* in the *S. saprophyticus* cluster group and *S. capitis* in the *S. epidermidis* cluster group.

Although some differences were observed, the topology of the MALDI-TOF MS-based phylogenetic tree was in more general agreement with that inferred from the analysis of the 16S rRNA gene sequences than that based on the *sodA* gene (20, 28). The proteins analyzed by MALDI Biotyper correspond to predominant proteins, such as ribosomal proteins. The probable coevolution of the ribosomal proteins and rRNAs may explain the similarities between 16S rRNA gene and MALDI-TOF-based dendrograms (23).

MALDI-TOF MS could thus be a valuable tool in phyloproteomics. It might serve as a technique for protein profiling, which is recommended as an additional test for the description of new staphylococcal species (7).

**Species identification of staphylococcal isolates.** Our dendrogram showed that the 36 reference *Staphylococcus* species were correctly distinguished. *S. vitulinus* and *S. pulvereri*, which have recently been shown to be one single species (27), were not differentiated by the MALDI-TOF MS technique. This technique seemed to be suitable for the differentiation of *Staphylococcus* isolates at the species level.

MALDI-TOF MS spectra were then obtained for all clinical and environmental staphylococcal isolates. A total of 151 strains out of 152 (99.3%) were identified at the species level by the MALDI Biotyper. One *S. saprophyticus* strain out of 16 was identified as *S. xylosum*, with a log score of 1.995. It was therefore identified at the genus level. The system identified both clinical (14 different species) and environmental (17 different species) strains without any complementary tests.

Carbannelle et al. used the MALDI-TOF technology to identify a clinical collection of *Staphylococcus* strains constituted by the four major clinically relevant *Staphylococcus* species (*S. aureus*, *S. epidermidis*, *S. warneri*, and *S. haemolyticus*) (3). Their reference database contained 22 staphylococci reference strains, including 17 staphylococcal species. In our study, we identified 22 *Staphylococcus* species from the MALDI Biotyper database, which comprised the spectra of 3,030 microorganisms, corresponding to Gram-negative and Gram-positive bacteria, whose 120 staphylococcal strains represented 36 staphylococcal species. *A fortiori*, the MALDI Biotyper did not require any initial assessment such as Gram staining or a catalase test, unlike the technology developed by Carbannelle et al., which requires a presumptive identification of *Staphylococcus* at the genus level.

The MALDI Biotyper database comprised 36 staphylococcal species. The *Staphylococcus* genus comprises 39 validly described species (J. P. Euzéby, List of bacterial names with standing in nomenclature, 5 November 2008, posting date; <http://www.bacterio.cict.fr/>). The three species *S. pettenkoferi*, *S. pseudintermedius*, and *S. gallinarum* were absent from the MALDI Biotyper database. The MALDI Biotyper database



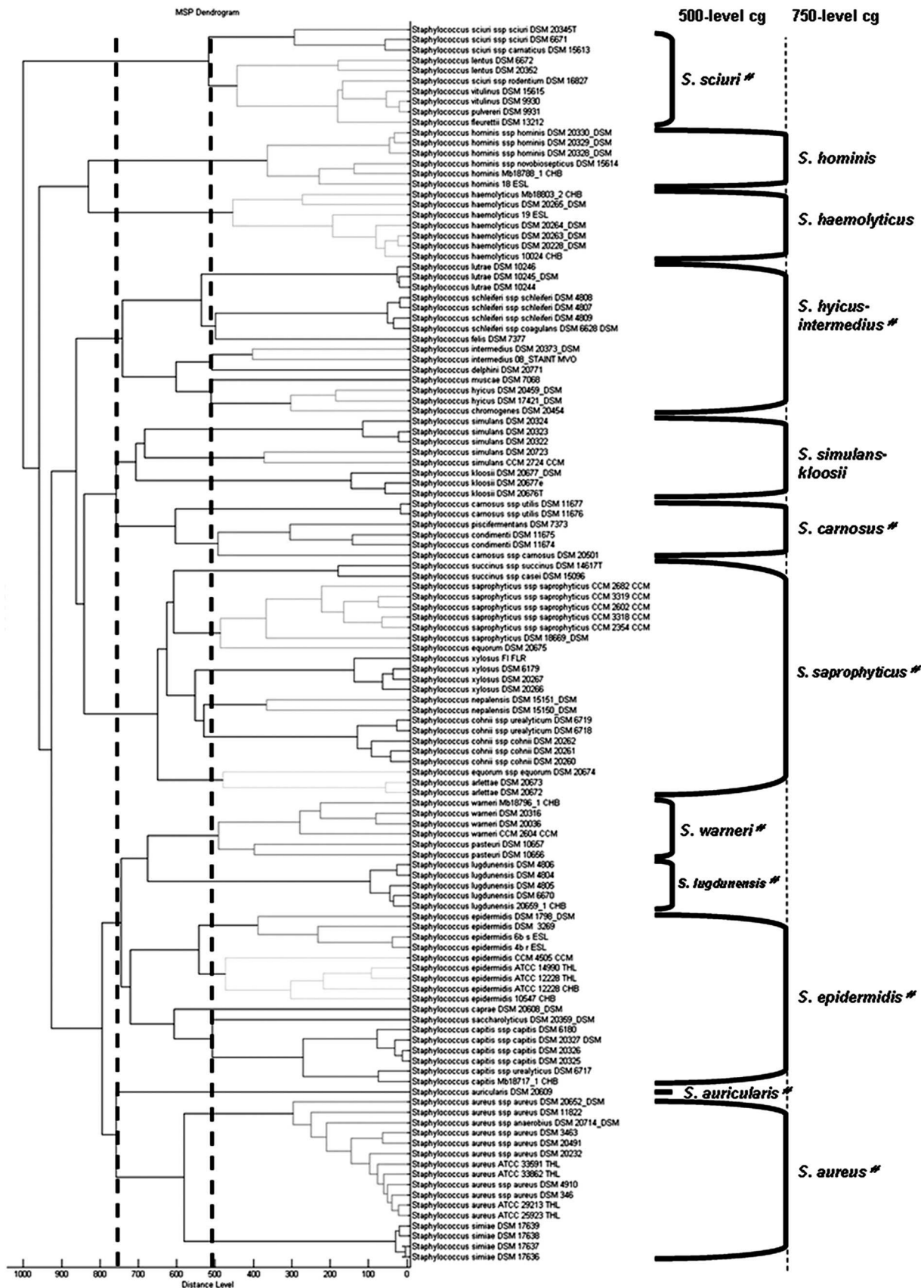


FIG. 1. Classification of staphylococcal reference strains. Shown is a score-oriented dendrogram of staphylococcal reference strains included in the database. The terms “500-level cg” and “750-level cg” define cluster groups (cg) based on the branching pattern using critical distance levels of 500 and 750, respectively. #, cluster groups as defined by the phylogenetic analysis of 16S rRNA genes.

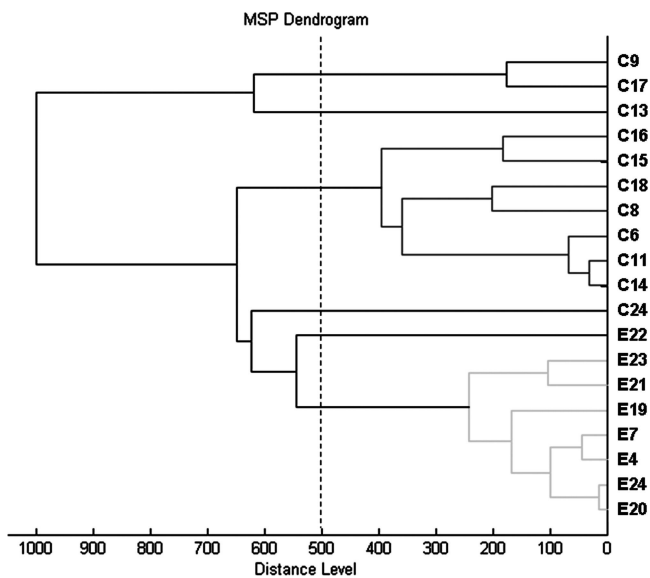


FIG. 2. Score-oriented dendrogram of clinical (C) and environmental (E) *S. epidermidis* isolates.

contains four species (*S. pasteurii*, *S. delphini*, *S. fleurettii*, and *S. succinus*) absent from the Vitek 2 Gram-positive (GP) card database (bioMérieux, La-Balme les Grottes, France). Moreover, according to the previously published data from our strain collection, the Vitek 2 system had misidentified or not identified 1 out of 7 *S. capitis*, 2 out of 7 *S. carnosus*, 11 out of 13 *S. equorum*, 1 out of 4 *S. simulans*, and 1 out of 5 *S. vitulinus* strains (6). This suggests that the MALDI-TOF MS technique is better fitted to the identification of staphylococcal strains that are rarely isolated in clinical bacteriology than automated biochemical methods. In addition, sample preparation and analysis with the MALDI-TOF MS technique are easier and more time-saving (approximately 20 min for one identification) than with genotypical methods, and result acquisition is faster than with biochemical methods.

**Biodiversity of *S. epidermidis* species.** The dendrogram of Fig. 2 shows protein “fingerprint” heterogeneity of clinical and environmental (salt meat originated) *S. epidermidis* strains. Clinical isolates are gathered within two cluster groups and two isolated strains, whereas meat origin isolates displayed one distinct cluster group and one isolated strain. Diversity within the *S. epidermidis* species has been observed for human origin *S. epidermidis* and animal origin *S. epidermidis* (15). The MALDI-TOF MS method might have enough discriminatory power to group isolates below the species level.

**Conclusion.** MALDI-TOF MS associated with MALDI Biotyper software appears a reliable and accurate tool for the identification of *Staphylococcus* species. However, further studies are required to test this technology with a large collection of staphylococci of diverse origins. The speed and the simplicity of sample preparation and result acquisition associated with minimal consumable costs make this method well suited for routine and high-throughput use. Hence, it is an excellent alternative to traditional methods in food processing and medical care laboratories. It may also be used for the analysis of clonal and/or taxonomic relationships.

## REFERENCES

- Barbuddhe, S. B., T. Maier, G. Schwarz, M. Kostrzewa, H. Hof, E. Domann, T. Chakraborty, and T. Hain. 2008. Rapid identification and typing of *Listeria* species using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* **74**:5402–5407.
- Bergonier, D., R. de Cremoux, R. Rupp, G. Lagriffoul, and X. Berthelot. 2003. Mastitis of dairy small ruminants. *Vet. Res.* **34**:689–716.
- Carbannelle, E., J. L. Beretti, S. Cottyn, G. Quesne, P. Berche, X. Nassif, and A. Ferroni. 2007. Rapid identification of staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* **45**:2156–2161.
- Choi, S. H., J. H. Woo, J. Y. Jeong, N. J. Kim, M. N. Kim, Y. S. Kim, and J. Ryu. 2006. Clinical significance of *Staphylococcus saprophyticus* identified on blood culture in a tertiary care hospital. *Diagn. Microbiol. Infect. Dis.* **56**:337–339.
- Claydon, M. A., S. N. Davey, V. Edwards-Jones, and D. B. Gordon. 1996. The rapid identification of intact microorganisms using mass spectrometry. *Nat. Biotechnol.* **14**:1584–1586.
- Delmas, J., J. P. Chacornac, F. Robin, P. Giammarinaro, R. Talon, and R. Bonnet. 2008. Evaluation of the Vitek 2 system with a variety of *Staphylococcus* species. *J. Clin. Microbiol.* **46**:311–313.
- Freney, J., W. E. Kloos, V. Hajek, J. A. Webster, M. Bes, Y. Brun, and C. Vernozy-Rozand. 1999. Recommended minimal standards for description of new staphylococcal species. Subcommittee on the taxonomy of staphylococci and streptococci of the International Committee on Systematic Bacteriology. *Int. J. Syst. Bacteriol.* **49**:489–502.
- Giammarinaro, P., S. Leroy, J. P. Chacornac, J. Delmas, and R. Talon. 2005. Development of a new oligonucleotide array to identify staphylococcal strains at species level. *J. Clin. Microbiol.* **43**:3673–3680.
- Grosse-Herrenthey, A., T. Maier, F. Gessler, R. Schaumann, H. Bohnel, M. Kostrzewa, and M. Kruger. 2008. Challenging the problem of clostridial identification with matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Anaerobe* **14**:242–249.
- Ieven, M., J. Verhoeven, S. R. Pattyn, and H. Goossens. 1995. Rapid and economical method for species identification of clinically significant coagulase-negative staphylococci. *J. Clin. Microbiol.* **33**:1060–1063.
- Keys, C. J., D. J. Dare, H. Sutton, G. Wells, M. Lunt, T. McKenna, M. McDowall, and H. N. Shah. 2004. Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect. Genet. Evol.* **4**:221–242.
- Kim, M., S. R. Heo, S. H. Choi, H. Kwon, J. S. Park, M. W. Seong, D. H. Lee, K. U. Park, J. Song, and E. C. Kim. 2008. Comparison of the MicroScan, VITEK 2, and Crystal GP with 16S rRNA sequencing and MicroSeq 500 v2.0 analysis for coagulase-negative staphylococci. *BMC Microbiol.* **8**:233.
- Kwok, A. Y., S. C. Su, R. P. Reynolds, S. J. Bay, Y. Av-Gay, N. J. Dovichi, and A. W. Chow. 1999. Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus *Staphylococcus*. *Int. J. Syst. Bacteriol.* **49**:1181–1192.
- Mellmann, A., J. Cloud, T. Maier, U. Keckevoet, I. Ramminger, P. Iwen, J. Dunn, G. Hall, D. Wilson, P. Lasala, M. Kostrzewa, and D. Harmsen. 2008. Evaluation of matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. *J. Clin. Microbiol.* **46**:1946–1954.
- Nagase, N., A. Sasaki, K. Yamashita, A. Shimizu, Y. Wakita, S. Kitai, and J. Kawano. 2002. Isolation and species distribution of staphylococci from animal and human skin. *J. Vet. Med. Sci.* **64**:245–250.
- Otto, M. 2009. *Staphylococcus epidermidis*—the ‘accidental’ pathogen. *Nat. Rev. Microbiol.* **7**:555–567.
- Pantucek, R., I. Sedlacek, P. Petras, D. Koukalova, P. Svec, V. Stetina, M. Vancanneyt, L. Chrastinova, J. Vokurkova, V. Ruzickova, J. Doskar, J. Swings, and V. Hajek. 2005. *Staphylococcus simiae* sp. nov., isolated from South American squirrel monkeys. *Int. J. Syst. Evol. Microbiol.* **55**:1953–1958.
- Papamanoli, E., P. Kotzekidou, N. Tzanetakis, and E. Litopoulou-Tzanetakis. 2002. Characterization of Micrococcaceae isolated from dry fermented sausage. *Food Microbiol.* **19**:441–449.
- Perl, T. M., P. R. Rhomberg, M. J. Bale, P. C. Fuchs, R. N. Jones, F. P. Koontz, and M. A. Pfaller. 1994. Comparison of identification systems for *Staphylococcus epidermidis* and other coagulase-negative *Staphylococcus* species. *Diagn. Microbiol. Infect. Dis.* **18**:151–155.
- Poyart, C., G. Quesne, C. Boumaila, and P. Trieu-Cuot. 2001. Rapid and accurate species-level identification of coagulase-negative staphylococci by using the *sodA* gene as a target. *J. Clin. Microbiol.* **39**:4296–4301.
- Rhoden, D. L., and J. M. Miller. 1995. Four-year prospective study of STAPH-IDENT system and conventional method for reference identification of *Staphylococcus*, *Stomatococcus*, and *Micrococcus* spp. *J. Clin. Microbiol.* **33**:96–98.
- Rich, M. 2005. Staphylococci in animals: prevalence, identification and antimicrobial susceptibility, with an emphasis on methicillin-resistant *Staphylococcus aureus*. *Br. J. Biomed. Sci.* **62**:98–105.

23. Ryzhov, V., and C. Fenselau. 2001. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Anal. Chem.* **73**:746–750.
24. Sampimon, O. C., R. N. Zadoks, S. De Vlieghe, K. Supre, F. Haesebrouck, H. W. Barkema, J. Sol, and T. J. Lam. 2009. Performance of API Staph ID 32 and Staph-Zym for identification of coagulase-negative staphylococci isolated from bovine milk samples. *Vet. Microbiol.* **136**:300–305.
25. Sauer, S., A. Freiwald, T. Maier, M. Kube, R. Reinhardt, M. Kostrzewa, and K. Geider. 2008. Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS One* **3**:e2843.
26. Sivadon, V., M. Rottman, S. Chaverot, J. C. Quincampoix, V. Avettand, P. de Mazancourt, L. Bernard, P. Trieu-Cuot, J. M. Feron, A. Lortat-Jacob, P. Piriou, T. Judet, and J. L. Gaillard. 2005. Use of genotypic identification by *sodA* sequencing in a prospective study to examine the distribution of coagulase-negative *Staphylococcus* species among strains recovered during septic orthopedic surgery and evaluate their significance. *J. Clin. Microbiol.* **43**:2952–2954.
27. Svec, P., M. Vancanneyt, I. Sedlacek, K. Engelbeen, V. Stetina, J. Swings, and P. Petras. 2004. Reclassification of *Staphylococcus pulvereri* Zakrzewska-Czerwinska et al. 1995 as a later synonym of *Staphylococcus vitulinus* Webster et al. *Int. J. Syst. Evol. Microbiol.* **54**:2213–2215.
28. Takahashi, T., I. Satoh, and N. Kikuchi. 1999. Phylogenetic relationships of 38 taxa of the genus *Staphylococcus* based on 16S rRNA gene sequence analysis. *Int. J. Syst. Bacteriol.* **49**:725–728.
29. Talon, R., S. Leroy-Setrin, and S. Fadda. 2002. Bacterial starters involved in the quality of fermented meat products, p. 175–191. *In* F. Toldra (ed.), *Research advances in quality of meat and meat products*. Research Signpost, Trivandrum, Kerala, India.
30. Taponen, S., and S. Pyorala. 2009. Coagulase-negative staphylococci as cause of bovine mastitis—not so different from *Staphylococcus aureus*? *Vet. Microbiol.* **134**:29–36.
31. Tenover, F. C., and R. P. Gayne. 2000. The epidemiology of *Staphylococcus* infections, p. 414–421. *In* J. I. Rood (ed.), *Gram-positive pathogens*. American Society for Microbiology, Washington, DC.
32. van Duijkeren, E., D. J. Houwers, A. Schoormans, M. J. Broekhuizen-Stins, R. Ikawaty, A. C. Fluit, and J. A. Wagenaar. 2008. Transmission of methicillin-resistant *Staphylococcus intermedius* between humans and animals. *Vet. Microbiol.* **128**:213–215.
33. Vliegen, I., J. A. Jacobs, E. Beuken, C. A. Bruggeman, and C. Vink. 2006. Rapid identification of bacteria by real-time amplification and sequencing of the 16S rRNA gene. *J. Microbiol. Methods* **66**:156–164.
34. von Eiff, C., G. Peters, and C. Heilmann. 2002. Pathogenesis of infections due to coagulase-negative staphylococci. *Lancet Infect. Dis.* **2**:677–685.
35. Weese, J. S., J. Rousseau, J. L. Traub-Dargatz, B. M. Willey, A. J. McGeer, and D. E. Low. 2005. Community-associated methicillin-resistant *Staphylococcus aureus* in horses and humans who work with horses. *J. Am. Vet. Med. Assoc.* **226**:580–583.
36. Winkler, M. A., J. Uher, and S. Cepa. 1999. Direct analysis and identification of *Helicobacter* and *Campylobacter* species by MALDI-TOF mass spectrometry. *Anal. Chem.* **71**:3416–3419.

**Chapitre V**

**Références  
bibliographiques**



1. Dethlefsen, L., et al., *The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing*. PLoS Biol, 2008. **6**(11): p. e280.
2. Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli*. Nat Rev Microbiol, 2004. **2**(2): p. 123-40.
3. Macdonald, T.T. and G. Monteleone, *Immunity, inflammation, and allergy in the gut*. Science, 2005. **307**(5717): p. 1920-5.
4. Hooper, L.V., T. Midtvedt, and J.I. Gordon, *How host-microbial interactions shape the nutrient environment of the mammalian intestine*. Annu Rev Nutr, 2002. **22**: p. 283-307.
5. Berg, R.D., *The indigenous gastrointestinal microflora*. Trends Microbiol, 1996. **4**(11): p. 430-5.
6. Hacker, J. and E. Carniel, *Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes*. EMBO Rep, 2001. **2**(5): p. 376-81.
7. Touchon, M., et al., *Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths*. PLoS Genet, 2009. **5**(1): p. e1000344.
8. Selander, R.K., D.A. Caugant, and T.S. Whittam, *Genetic structure and variation in natural populations of Escherichia coli*, in *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, F.C. Neidhardt, et al., Editors. 1987, American Society of Microbiology: Washington, DC. p. 1625–1648.
9. Ochman, H. and R.K. Selander, *Standard reference strains of Escherichia coli from natural populations*. J Bacteriol, 1984. **157**(2): p. 690-3.
10. Herzer, P.J., et al., *Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of Escherichia coli*. J Bacteriol, 1990. **172**(11): p. 6175-81.
11. Desjardins, P., et al., *Sex in Escherichia coli does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism*. J Mol Evol, 1995. **41**(4): p. 440-8.
12. Clermont, O., S. Bonacorsi, and E. Bingen, *Rapid and simple determination of the Escherichia coli phylogenetic group*. Appl Environ Microbiol, 2000. **66**(10): p. 4555-8.
13. Wirth, T., et al., *Sex and virulence in Escherichia coli: an evolutionary perspective*. Mol Microbiol, 2006. **60**(5): p. 1136-51.
14. Gordon, D.M., et al., *Assigning Escherichia coli strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method*. Environ Microbiol, 2008. **10**(10): p. 2484-96.
15. Escobar-Paramo, P., et al., *Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the Escherichia coli case study*. Mol Phylogenet Evol, 2004. **30**(1): p. 243-50.
16. Escobar-Paramo, P., et al., *A specific genetic background is required for acquisition and expression of virulence factors in Escherichia coli*. Mol Biol Evol, 2004. **21**(6): p. 1085-94.
17. Picard, B., et al., *The link between phylogeny and virulence in Escherichia coli extraintestinal infection*. Infect Immun, 1999. **67**(2): p. 546-53.
18. Johnson, J.R., et al., *Phylogenetic distribution of extraintestinal virulence-associated traits in Escherichia coli*. J Infect Dis, 2001. **183**(1): p. 78-88.
19. Boyd, E.F. and D.L. Hartl, *Chromosomal regions specific to pathogenic isolates of Escherichia coli have a phylogenetically clustered distribution*. J Bacteriol, 1998. **180**(5): p. 1159-65.

20. Darfeuille-Michaud, A., *Adherent-invasive Escherichia coli: a putative new E. coli pathotype associated with Crohn's disease*. Int J Med Microbiol, 2002. **292**(3-4): p. 185-93.
21. Wiles, T.J., R.R. Kulesus, and M.A. Mulvey, *Origins and virulence mechanisms of uropathogenic Escherichia coli*. Exp Mol Pathol, 2008. **85**(1): p. 11-9.
22. Russo, T.A. and J.R. Johnson, *Proposal for a new inclusive designation for extraintestinal pathogenic isolates of Escherichia coli: ExPEC*. J Infect Dis, 2000. **181**(5): p. 1753-4.
23. Nougayrede, J.P., et al., *Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle*. Trends Microbiol, 2005. **13**(3): p. 103-10.
24. Nougayrede, J.P., et al., *Escherichia coli induces DNA double-strand breaks in eukaryotic cells*. Science, 2006. **313**(5788): p. 848-51.
25. Fischbach, M.A. and C.T. Walsh, *Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms*. Chem Rev, 2006. **106**(8): p. 3468-96.
26. Pommier, Y. and K.W. Kohn, *[Cell cycle and checkpoints in oncology: new therapeutic targets]*. Med Sci (Paris), 2003. **19**(2): p. 173-86.
27. Norbury, C. and P. Nurse, *Animal cell cycles and their control*. Annu Rev Biochem, 1992. **61**: p. 441-70.
28. Malumbres, M. and M. Barbacid, *Mammalian cyclin-dependent kinases*. Trends Biochem Sci, 2005. **30**(11): p. 630-41.
29. Vermeulen, K., D.R. Van Bockstaele, and Z.N. Berneman, *The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer*. Cell Prolif, 2003. **36**(3): p. 131-49.
30. Morgan, D.O., *Principles of CDK regulation*. Nature, 1995. **374**(6518): p. 131-4.
31. Hohegger, H., S. Takeda, and T. Hunt, *Cyclin-dependent kinases and cell-cycle transitions: does one fit all?* Nat Rev Mol Cell Biol, 2008. **9**(11): p. 910-6.
32. Pines, J., *Cyclins: wheels within wheels*. Cell Growth Differ, 1991. **2**(6): p. 305-10.
33. Evans, T., et al., *Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division*. Cell, 1983. **33**(2): p. 389-96.
34. Glotzer, M., A.W. Murray, and M.W. Kirschner, *Cyclin is degraded by the ubiquitin pathway*. Nature, 1991. **349**(6305): p. 132-8.
35. Diehl, J.A., F. Zindy, and C.J. Sherr, *Inhibition of cyclin D1 phosphorylation on threonine-286 prevents its rapid degradation via the ubiquitin-proteasome pathway*. Genes Dev, 1997. **11**(8): p. 957-72.
36. Sherr, C.J., *G1 phase progression: cycling on cue*. Cell, 1994. **79**(4): p. 551-5.
37. King, R.W., et al., *A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B*. Cell, 1995. **81**(2): p. 279-88.
38. Lew, D.J. and S. Kornbluth, *Regulatory roles of cyclin dependent kinase phosphorylation in cell cycle control*. Curr Opin Cell Biol, 1996. **8**(6): p. 795-804.
39. Jeffrey, P.D., et al., *Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex*. Nature, 1995. **376**(6538): p. 313-20.
40. Fisher, R.P. and D.O. Morgan, *A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase*. Cell, 1994. **78**(4): p. 713-24.
41. Boutros, R., C. Dozier, and B. Ducommun, *The when and wheres of CDC25 phosphatases*. Curr Opin Cell Biol, 2006. **18**(2): p. 185-91.
42. Booher, R.N., P.S. Holman, and A. Fattaey, *Human Myt1 is a cell cycle-regulated kinase that inhibits Cdc2 but not Cdk2 activity*. J Biol Chem, 1997. **272**(35): p. 22300-6.



43. Parker, L.L. and H. Piwnicka-Worms, *Inactivation of the p34cdc2-cyclin B complex by the human WEE1 tyrosine kinase*. *Science*, 1992. **257**(5078): p. 1955-7.
44. Sherr, C.J. and J.M. Roberts, *Inhibitors of mammalian G1 cyclin-dependent kinases*. *Genes Dev*, 1995. **9**(10): p. 1149-63.
45. Carnero, A. and G.J. Hannon, *The INK4 family of CDK inhibitors*. *Curr Top Microbiol Immunol*, 1998. **227**: p. 43-55.
46. Polyak, K., et al., *Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals*. *Cell*, 1994. **78**(1): p. 59-66.
47. Harper, J.W., et al., *Inhibition of cyclin-dependent kinases by p21*. *Mol Biol Cell*, 1995. **6**(4): p. 387-400.
48. Lee, M.H., I. Reynisdottir, and J. Massague, *Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution*. *Genes Dev*, 1995. **9**(6): p. 639-49.
49. Hengst, L. and S.I. Reed, *Inhibitors of the Cip/Kip family*. *Curr Top Microbiol Immunol*, 1998. **227**: p. 25-41.
50. Waga, S., R. Li, and B. Stillman, *p53-induced p21 controls DNA replication*. *Leukemia*, 1997. **11 Suppl 3**: p. 321-3.
51. Toyoshima, F., et al., *Nuclear export of cyclin B1 and its possible role in the DNA damage-induced G2 checkpoint*. *Embo J*, 1998. **17**(10): p. 2728-35.
52. Heald, R., M. McLoughlin, and F. McKeon, *Human wee1 maintains mitotic timing by protecting the nucleus from cytoplasmically activated Cdc2 kinase*. *Cell*, 1993. **74**(3): p. 463-74.
53. Liu, F., et al., *The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex*. *Mol Cell Biol*, 1997. **17**(2): p. 571-83.
54. Peng, C.Y., et al., *Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216*. *Science*, 1997. **277**(5331): p. 1501-5.
55. Yang, J., et al., *Maintenance of G2 arrest in the Xenopus oocyte: a role for 14-3-3-mediated inhibition of Cdc25 nuclear import*. *Embo J*, 1999. **18**(8): p. 2174-83.
56. Kato, J., et al., *Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4*. *Genes Dev*, 1993. **7**(3): p. 331-42.
57. Buchkovich, K., L.A. Duffy, and E. Harlow, *The retinoblastoma protein is phosphorylated during specific phases of the cell cycle*. *Cell*, 1989. **58**(6): p. 1097-105.
58. Nevins, J.R., *Toward an understanding of the functional complexity of the E2F and retinoblastoma families*. *Cell Growth Differ*, 1998. **9**(8): p. 585-93.
59. Girard, F., et al., *Cyclin A is required for the onset of DNA replication in mammalian fibroblasts*. *Cell*, 1991. **67**(6): p. 1169-79.
60. Zindy, F., et al., *Cyclin A is required in S phase in normal epithelial cells*. *Biochem Biophys Res Commun*, 1992. **182**(3): p. 1144-54.
61. Bradbury, E.M., R.J. Inglis, and H.R. Matthews, *Control of cell division by very lysine rich histone (F1) phosphorylation*. *Nature*, 1974. **247**(439): p. 257-61.
62. Buttinelli, M., et al., *The role of histone H1 in chromatin condensation and transcriptional repression*. *Genetica*, 1999. **106**(1-2): p. 117-24.
63. Voitenleitner, C., E. Fanning, and H.P. Nasheuer, *Phosphorylation of DNA polymerase alpha-primase by cyclin A-dependent kinases regulates initiation of DNA replication in vitro*. *Oncogene*, 1997. **14**(13): p. 1611-5.
64. Ubersax, J.A., et al., *Targets of the cyclin-dependent kinase Cdk1*. *Nature*, 2003. **425**(6960): p. 859-64.



65. Sancar, A., et al., *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints*. Annu Rev Biochem, 2004. **73**: p. 39-85.
66. Zhou, B.B. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective*. Nature, 2000. **408**(6811): p. 433-9.
67. Jazayeri, A., et al., *ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks*. Nat Cell Biol, 2006. **8**(1): p. 37-45.
68. Kastan, M.B. and J. Bartek, *Cell-cycle checkpoints and cancer*. Nature, 2004. **432**(7015): p. 316-23.
69. Harrison, J.C. and J.E. Haber, *Surviving the breakup: the DNA damage checkpoint*. Annu Rev Genet, 2006. **40**: p. 209-35.
70. Lee, J.H. and T.T. Paull, *ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex*. Science, 2005. **308**(5721): p. 551-4.
71. Branzei, D. and M. Foiani, *Regulation of DNA repair throughout the cell cycle*. Nat Rev Mol Cell Biol, 2008. **9**(4): p. 297-308.
72. Fernandez-Capetillo, O., A. Celeste, and A. Nussenzweig, *Focusing on foci: H2AX and the recruitment of DNA-damage response factors*. Cell Cycle, 2003. **2**(5): p. 426-7.
73. Messick, T.E. and R.A. Greenberg, *The ubiquitin landscape at DNA double-strand breaks*. J Cell Biol, 2009. **187**(3): p. 319-26.
74. Taylor, W.R. and G.R. Stark, *Regulation of the G2/M transition by p53*. Oncogene, 2001. **20**(15): p. 1803-15.
75. Zeng, Y., et al., *Replication checkpoint requires phosphorylation of the phosphatase Cdc25 by Cds1 or Chk1*. Nature, 1998. **395**(6701): p. 507-10.
76. Falck, J., et al., *The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis*. Nature, 2001. **410**(6830): p. 842-7.
77. Bartek, J. and J. Lukas, *Mammalian G1- and S-phase checkpoints in response to DNA damage*. Curr Opin Cell Biol, 2001. **13**(6): p. 738-47.
78. Harper, J.W., et al., *The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases*. Cell, 1993. **75**(4): p. 805-16.
79. Hermeking, H., et al., *14-3-3 sigma is a p53-regulated inhibitor of G2/M progression*. Mol Cell, 1997. **1**(1): p. 3-11.
80. Bartek, J., C. Lukas, and J. Lukas, *Checking on DNA damage in S phase*. Nat Rev Mol Cell Biol, 2004. **5**(10): p. 792-804.
81. Musacchio, A. and E.D. Salmon, *The spindle-assembly checkpoint in space and time*. Nat Rev Mol Cell Biol, 2007. **8**(5): p. 379-93.
82. Schang, L.M., *The cell cycle, cyclin-dependent kinases, and viral infections: new horizons and unexpected connections*. Prog Cell Cycle Res, 2003. **5**: p. 103-24.
83. Caprioli, A., et al., *Partial purification and characterization of an escherichia coli toxic factor that induces morphological cell alterations*. Infect Immun, 1983. **39**(3): p. 1300-6.
84. Orden, J.A., et al., *Necrotoxic Escherichia coli from sheep and goats produce a new type of cytotoxic necrotizing factor (CNF3) associated with the eae and ehxA genes*. Int Microbiol, 2007. **10**(1): p. 47-55.
85. Lockman, H.A., et al., *Yersinia pseudotuberculosis produces a cytotoxic necrotizing factor*. Infect Immun, 2002. **70**(5): p. 2708-14.
86. De Rycke, J., A. Milon, and E. Oswald, *Necrotoxic Escherichia coli (NTEC): two emerging categories of human and animal pathogens*. Vet Res, 1999. **30**(2-3): p. 221-33.
87. Lemonnier, M., L. Landraud, and E. Lemichez, *Rho GTPase-activating bacterial toxins: from bacterial virulence regulation to eukaryotic cell biology*. FEMS Microbiol Rev, 2007. **31**(5): p. 515-34.

88. Lemichez, E., et al., *Molecular localization of the Escherichia coli cytotoxic necrotizing factor CNF1 cell-binding and catalytic domains*. Mol Microbiol, 1997. **24**(5): p. 1061-70.
89. Fabbri, A., M. Gauthier, and P. Boquet, *The 5' region of cnf1 harbours a translational regulatory mechanism for CNF1 synthesis and encodes the cell-binding domain of the toxin*. Mol Microbiol, 1999. **33**(1): p. 108-18.
90. Pei, S., A. Doye, and P. Boquet, *Mutation of specific acidic residues of the CNF1 T domain into lysine alters cell membrane translocation of the toxin*. Mol Microbiol, 2001. **41**(6): p. 1237-47.
91. Buetow, L., et al., *Structure of the Rho-activating domain of Escherichia coli cytotoxic necrotizing factor 1*. Nat Struct Biol, 2001. **8**(7): p. 584-8.
92. Schmidt, G., et al., *The Rho-deamidating cytotoxic necrotizing factor 1 from Escherichia coli possesses transglutaminase activity. Cysteine 866 and histidine 881 are essential for enzyme activity*. J Biol Chem, 1998. **273**(22): p. 13669-74.
93. Blum, G., et al., *Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic Escherichia coli strain J96*. FEMS Microbiol Lett, 1995. **126**(2): p. 189-95.
94. Landraud, L., et al., *Expression of cnf1 by Escherichia coli J96 involves a large upstream DNA region including the hlyCABD operon, and is regulated by the RfaH protein*. Mol Microbiol, 2003. **47**(6): p. 1653-67.
95. Johnson, T.J., et al., *Pyrosequencing of the Vir plasmid of necrotoxicogenic Escherichia coli*. Vet Microbiol, 2010. **144**(1-2): p. 100-9.
96. Leeds, J.A. and R.A. Welch, *RfaH enhances elongation of Escherichia coli hlyCABD mRNA*. J Bacteriol, 1996. **178**(7): p. 1850-7.
97. Kouokam, J.C., et al., *Active cytotoxic necrotizing factor 1 associated with outer membrane vesicles from uropathogenic Escherichia coli*. Infect Immun, 2006. **74**(4): p. 2022-30.
98. Yu, H. and K.S. Kim, *Ferredoxin is involved in secretion of cytotoxic necrotizing factor 1 across the cytoplasmic membrane in Escherichia coli K1*. Infect Immun, 2010. **78**(2): p. 838-44.
99. Davis, J.M., et al., *Cytotoxic necrotizing factor type 1 delivered by outer membrane vesicles of uropathogenic Escherichia coli attenuates polymorphonuclear leukocyte antimicrobial activity and chemotaxis*. Infect Immun, 2006. **74**(8): p. 4401-8.
100. Kim, K.J., J.W. Chung, and K.S. Kim, *67-kDa laminin receptor promotes internalization of cytotoxic necrotizing factor 1-expressing Escherichia coli K1 into human brain microvascular endothelial cells*. J Biol Chem, 2005. **280**(2): p. 1360-8.
101. Chung, J.W., et al., *37-kDa laminin receptor precursor modulates cytotoxic necrotizing factor 1-mediated RhoA activation and bacterial uptake*. J Biol Chem, 2003. **278**(19): p. 16857-62.
102. Contamin, S., et al., *The p21 Rho-activating toxin cytotoxic necrotizing factor 1 is endocytosed by a clathrin-independent mechanism and enters the cytosol by an acidic-dependent membrane translocation step*. Mol Biol Cell, 2000. **11**(5): p. 1775-87.
103. McNichol, B.A., et al., *Two domains of cytotoxic necrotizing factor type 1 bind the cellular receptor, laminin receptor precursor protein*. Infect Immun, 2007. **75**(11): p. 5095-104.
104. Grande, K.K., et al., *Cytotoxic necrotizing factor type 1-neutralizing monoclonal antibody NG8 recognizes three amino acids in a C-terminal region of the toxin and reduces toxin binding to HEp-2 cells*. Infect Immun, 2009. **77**(1): p. 170-9.

105. Blumenthal, B., et al., *The cytotoxic necrotizing factors from Yersinia pseudotuberculosis and from Escherichia coli bind to different cellular receptors but take the same route to the cytosol*. Infect Immun, 2007. **75**(7): p. 3344-53.
106. Stoll, T., et al., *A new member of a growing toxin family--Escherichia coli cytotoxic necrotizing factor 3 (CNF3)*. Toxicon, 2009. **54**(6): p. 745-53.
107. Knust, Z., et al., *Cleavage of Escherichia coli cytotoxic necrotizing factor 1 is required for full biologic activity*. Infect Immun, 2009. **77**(5): p. 1835-41.
108. Oswald, E., et al., *Cytotoxic necrotizing factor type 2 produced by virulent Escherichia coli modifies the small GTP-binding proteins Rho involved in assembly of actin stress fibers*. Proc Natl Acad Sci U S A, 1994. **91**(9): p. 3814-8.
109. Hoffmann, C., et al., *The Yersinia pseudotuberculosis cytotoxic necrotizing factor (CNFY) selectively activates RhoA*. J Biol Chem, 2004. **279**(16): p. 16026-32.
110. Etienne-Manneville, S. and A. Hall, *Rho GTPases in cell biology*. Nature, 2002. **420**(6916): p. 629-35.
111. Schmidt, G., et al., *Gln 63 of Rho is deamidated by Escherichia coli cytotoxic necrotizing factor-1*. Nature, 1997. **387**(6634): p. 725-9.
112. Flatau, G., et al., *Toxin-induced activation of the G protein p21 Rho by deamidation of glutamine*. Nature, 1997. **387**(6634): p. 729-33.
113. Lerm, M., et al., *Identification of the region of rho involved in substrate recognition by Escherichia coli cytotoxic necrotizing factor 1 (CNF1)*. J Biol Chem, 1999. **274**(41): p. 28999-9004.
114. Lerm, M., et al., *Deamidation of Cdc42 and Rac by Escherichia coli cytotoxic necrotizing factor 1: activation of c-Jun N-terminal kinase in HeLa cells*. Infect Immun, 1999. **67**(2): p. 496-503.
115. Lerm, M., et al., *Proteasomal degradation of cytotoxic necrotizing factor 1-activated rac*. Infect Immun, 2002. **70**(8): p. 4053-8.
116. Doye, A., et al., *Ubiquitin-mediated proteasomal degradation of Rho proteins by the CNF1 toxin*. Methods Enzymol, 2006. **406**: p. 447-56.
117. Doye, A., et al., *CNF1 exploits the ubiquitin-proteasome machinery to restrict Rho GTPase activation for bacterial host cell invasion*. Cell, 2002. **111**(4): p. 553-64.
118. Caprioli, A., et al., *A cell division-active protein from E. coli*. Biochem Biophys Res Commun, 1984. **118**(2): p. 587-93.
119. Fiorentini, C., et al., *Cytoskeletal changes induced in HEp-2 cells by the cytotoxic necrotizing factor of Escherichia coli*. Toxicon, 1988. **26**(11): p. 1047-56.
120. Fiorentini, C., et al., *Activation of rho GTPases by cytotoxic necrotizing factor 1 induces macropinocytosis and scavenging activity in epithelial cells*. Mol Biol Cell, 2001. **12**(7): p. 2061-73.
121. Falzano, L., et al., *Induction of phagocytic behaviour in human epithelial cells by Escherichia coli cytotoxic necrotizing factor type 1*. Mol Microbiol, 1993. **9**(6): p. 1247-54.
122. Falzano, L., et al., *Epithelial cells challenged with a Rac-activating E. coli cytotoxin acquire features of professional phagocytes*. Toxicol In Vitro, 2002. **16**(4): p. 421-5.
123. Malorni, W. and C. Fiorentini, *Is the Rac GTPase-activating toxin CNF1 a smart hijacker of host cell fate?* Faseb J, 2006. **20**(6): p. 606-9.
124. Lacerda, H.M., et al., *Cytotoxic necrotizing factor 1 from Escherichia coli and dermonecrotic toxin from Bordetella bronchiseptica induce p21(rho)-dependent tyrosine phosphorylation of focal adhesion kinase and paxillin in Swiss 3T3 cells*. J Biol Chem, 1997. **272**(14): p. 9587-96.
125. Falzano, L., et al., *Interaction of Escherichia coli cytotoxic necrotizing factor type 1 (CNF1) with cultured cells*. Cytotechnology, 1993. **11 Suppl 1**: p. S56-8.

126. Boyer, L., et al., *Rac GTPase instructs nuclear factor-kappaB activation by conveying the SCF complex and IκBα to the ruffling membranes*. Mol Biol Cell, 2004. **15**(3): p. 1124-33.
127. Johnson, J.R. and A.L. Stell, *Extended virulence genotypes of Escherichia coli strains from patients with urosepsis in relation to phylogeny and host compromise*. J Infect Dis, 2000. **181**(1): p. 261-72.
128. Johnson, J.R., et al., *Host-pathogen relationships among Escherichia coli isolates recovered from men with febrile urinary tract infection*. Clin Infect Dis, 2005. **40**(6): p. 813-22.
129. Bingen-Bidois, M., et al., *Phylogenetic analysis and prevalence of urosepsis strains of Escherichia coli bearing pathogenicity island-like domains*. Infect Immun, 2002. **70**(6): p. 3216-26.
130. Bonacorsi, S. and E. Bingen, *Molecular epidemiology of Escherichia coli causing neonatal meningitis*. Int J Med Microbiol, 2005. **295**(6-7): p. 373-81.
131. Kadhum, H.J., et al., *Occurrence and characteristics of cytotoxic necrotizing factors, cytolethal distending toxins and other virulence factors in Escherichia coli from human blood and faecal samples*. Epidemiol Infect, 2008. **136**(6): p. 752-60.
132. Mainil, J.G., et al., *Comparison of necrotoxicogenic Escherichia coli isolates from farm animals and from humans*. Vet Microbiol, 1999. **70**(1-2): p. 123-35.
133. Kavitha, K., et al., *Isolation of necrotoxicogenic Escherichia coli from paediatric patients with acute diarrhoea*. J Med Microbiol, 2010. **59**(Pt 4): p. 503-4.
134. Van Bost, S., et al., *Characteristics of necrotoxicogenic Escherichia coli isolated from septicemic and diarrheic calves between 1958 and 1970*. Vet Microbiol, 2001. **82**(4): p. 311-20.
135. Sannes, M.R., et al., *Virulence factor profiles and phylogenetic background of Escherichia coli isolates from veterans with bacteremia and uninfected control subjects*. J Infect Dis, 2004. **190**(12): p. 2121-8.
136. Andreu, A., et al., *Urovirulence determinants in Escherichia coli strains causing prostatitis*. J Infect Dis, 1997. **176**(2): p. 464-9.
137. Mitsumori, K., et al., *Virulence characteristics of Escherichia coli in acute bacterial prostatitis*. J Infect Dis, 1999. **180**(4): p. 1378-81.
138. Johnson, J.R., et al., *Phylogenetic and pathotypic comparison of concurrent urine and rectal Escherichia coli isolates from men with febrile urinary tract infection*. J Clin Microbiol, 2005. **43**(8): p. 3895-900.
139. Johnson, J.R. and T.A. Russo, *Molecular epidemiology of extraintestinal pathogenic (uropathogenic) Escherichia coli*. Int J Med Microbiol, 2005. **295**(6-7): p. 383-404.
140. Real, J.M., et al., *Specificity of immunomodulator secretion in urinary samples in response to infection by alpha-hemolysin and CNF1 bearing uropathogenic Escherichia coli*. Cytokine, 2007. **37**(1): p. 22-5.
141. Terai, A., et al., *Escherichia coli virulence factors and serotypes in acute bacterial prostatitis*. Int J Urol, 1997. **4**(3): p. 289-94.
142. Kanamaru, S., et al., *Distribution and genetic association of putative uropathogenic virulence factors iroN, iha, kpsMT, ompT and usp in Escherichia coli isolated from urinary tract infections in Japan*. J Urol, 2003. **170**(6 Pt 1): p. 2490-3.
143. Yamamoto, S., *Molecular epidemiology of uropathogenic Escherichia coli*. J Infect Chemother, 2007. **13**(2): p. 68-73.
144. Johnson, J.R., et al., *Extended virulence genotypes and phylogenetic background of Escherichia coli isolates from patients with cystitis, pyelonephritis, or prostatitis*. J Infect Dis, 2005. **191**(1): p. 46-50.

145. Johnson, J.R., et al., *Phylogenetic distribution of virulence-associated genes among Escherichia coli isolates associated with neonatal bacterial meningitis in the Netherlands*. J Infect Dis, 2002. **185**(6): p. 774-84.
146. Bonacorsi, S., et al., *Molecular analysis and experimental virulence of French and North American Escherichia coli neonatal meningitis isolates: identification of a new virulent clone*. J Infect Dis, 2003. **187**(12): p. 1895-906.
147. Okeke, I.N., et al., *Characterization of Escherichia coli strains from cases of childhood diarrhea in provincial southwestern Nigeria*. J Clin Microbiol, 2000. **38**(1): p. 7-12.
148. Elliott, S.J., et al., *Characterization of the roles of hemolysin and other toxins in enteropathy caused by alpha-hemolytic Escherichia coli linked to human diarrhea*. Infect Immun, 1998. **66**(5): p. 2040-51.
149. Paciorek, J., *Virulence properties of Escherichia coli faecal strains isolated in Poland from healthy children and strains belonging to serogroups O18, O26, O44, O86, O126 and O127 isolated from children with diarrhoea*. J Med Microbiol, 2002. **51**(7): p. 548-56.
150. Blanco, J., et al., *Enterotoxigenic and necrotizing Escherichia coli in human diarrhoea in Spain*. Eur J Epidemiol, 1992. **8**(4): p. 548-52.
151. Oswald, E., et al., *Virulence factors associated with cytotoxic necrotizing factor type two in bovine diarrheic and septicemic strains of Escherichia coli*. J Clin Microbiol, 1991. **29**(11): p. 2522-7.
152. Van Bost, S., et al., *Putative roles of the CNF2 and CDTIII toxins in experimental infections with necrotizing Escherichia coli type 2 (NTEC2) strains in calves*. Microbes Infect, 2003. **5**(13): p. 1189-93.
153. Khan, N.A., et al., *Outer membrane protein A and cytotoxic necrotizing factor-1 use diverse signaling mechanisms for Escherichia coli K1 invasion of human brain microvascular endothelial cells*. Microb Pathog, 2003. **35**(1): p. 35-42.
154. Mysorekar, I.U. and S.J. Hultgren, *Mechanisms of uropathogenic Escherichia coli persistence and eradication from the urinary tract*. Proc Natl Acad Sci U S A, 2006. **103**(38): p. 14170-5.
155. Mills, M., K.C. Meysick, and A.D. O'Brien, *Cytotoxic necrotizing factor type 1 of uropathogenic Escherichia coli kills cultured human uroepithelial 5637 cells by an apoptotic mechanism*. Infect Immun, 2000. **68**(10): p. 5869-80.
156. Falzano, L., et al., *Cytotoxic necrotizing factor 1 enhances reactive oxygen species-dependent transcription and secretion of proinflammatory cytokines in human uroepithelial cells*. Infect Immun, 2003. **71**(7): p. 4178-81.
157. Rippere-Lampe, K.E., et al., *Cytotoxic necrotizing factor type 1-positive Escherichia coli causes increased inflammation and tissue damage to the prostate in a rat prostatitis model*. Infect Immun, 2001. **69**(10): p. 6515-9.
158. Rippere-Lampe, K.E., et al., *Mutation of the gene encoding cytotoxic necrotizing factor type 1 (cnf(1)) attenuates the virulence of uropathogenic Escherichia coli*. Infect Immun, 2001. **69**(6): p. 3954-64.
159. Smith, Y.C., et al., *Hemolysin of uropathogenic Escherichia coli evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice*. Infect Immun, 2008. **76**(7): p. 2978-90.
160. Uhlen, P., et al., *Alpha-haemolysin of uropathogenic E. coli induces Ca<sup>2+</sup> oscillations in renal epithelial cells*. Nature, 2000. **405**(6787): p. 694-7.
161. Wullt, B., et al., *The role of P fimbriae for colonization and host response induction in the human urinary tract*. J Infect Dis, 2001. **183 Suppl 1**: p. S43-6.

162. Hofman, P., et al., *Escherichia coli* cytotoxic necrotizing factor 1 effaces microvilli and decreases transmigration of polymorphonuclear leukocytes in intestinal T84 epithelial cell monolayers. *Infect Immun*, 1998. **66**(6): p. 2494-500.
163. Davis, J.M., S.B. Rasmussen, and A.D. O'Brien, Cytotoxic necrotizing factor type 1 production by uropathogenic *Escherichia coli* modulates polymorphonuclear leukocyte function. *Infect Immun*, 2005. **73**(9): p. 5301-10.
164. Munro, P., et al., Activation and proteasomal degradation of rho GTPases by cytotoxic necrotizing factor-1 elicit a controlled inflammatory response. *J Biol Chem*, 2004. **279**(34): p. 35849-57.
165. Bos, J.L., *ras* oncogenes in human cancer: a review. *Cancer Res*, 1989. **49**(17): p. 4682-9.
166. Caygill, C.P., et al., The association between typhoid carriage, typhoid infection and subsequent cancer at a number of sites. *Eur J Cancer Prev*, 1995. **4**(2): p. 187-93.
167. Schlumberger, M.C. and W.D. Hardt, Triggered phagocytosis by *Salmonella*: bacterial molecular mimicry of RhoGTPase activation/deactivation. *Curr Top Microbiol Immunol*, 2005. **291**: p. 29-42.
168. Basseres, D.S. and A.S. Baldwin, Nuclear factor-kappaB and inhibitor of kappaB kinase pathways in oncogenic initiation and progression. *Oncogene*, 2006. **25**(51): p. 6817-30.
169. Thomas, W., et al., Cytotoxic necrotizing factor from *Escherichia coli* induces RhoA-dependent expression of the cyclooxygenase-2 Gene. *Infect Immun*, 2001. **69**(11): p. 6839-45.
170. Marx, J., *Cancer research. Inflammation and cancer: the link grows stronger*. *Science*, 2004. **306**(5698): p. 966-8.
171. Fiorentini, C., et al., Toxin-induced activation of Rho GTP-binding protein increases *Bcl-2* expression and influences mitochondrial homeostasis. *Exp Cell Res*, 1998. **242**(1): p. 341-50.
172. Fiorentini, C., et al., Rho-dependent cell spreading activated by *E.coli* cytotoxic necrotizing factor 1 hinders apoptosis in epithelial cells. *Cell Death Differ*, 1998. **5**(11): p. 921-9.
173. Miraglia, A.G., et al., Cytotoxic necrotizing factor 1 prevents apoptosis via the Akt/IkappaB kinase pathway: role of nuclear factor-kappaB and *Bcl-2*. *Mol Biol Cell*, 2007. **18**(7): p. 2735-44.
174. Hopkins, A.M., et al., Constitutive activation of Rho proteins by CNF-1 influences tight junction structure and epithelial barrier function. *J Cell Sci*, 2003. **116**(Pt 4): p. 725-42.
175. Min, C., et al., NF-kappaB and epithelial to mesenchymal transition of cancer. *J Cell Biochem*, 2008. **104**(3): p. 733-44.
176. Travaglione, S., A. Fabbri, and C. Fiorentini, The Rho-activating CNF1 toxin from pathogenic *E. coli*: A risk factor for human cancer development? *Infect Agent Cancer*, 2008. **3**: p. 4.
177. Peres, S.Y., et al., A new cytolethal distending toxin (CDT) from *Escherichia coli* producing CNF2 blocks HeLa cell division in G2/M phase. *Mol Microbiol*, 1997. **24**(5): p. 1095-107.
178. Ge, Z., D.B. Schauer, and J.G. Fox, *In vivo* virulence properties of bacterial cytolethal-distending toxin. *Cell Microbiol*, 2008. **10**(8): p. 1599-607.
179. Toth, I., et al., Production of cytolethal distending toxins by pathogenic *Escherichia coli* strains isolated from human and animal sources: establishment of the existence of a new *cdt* variant (Type IV). *J Clin Microbiol*, 2003. **41**(9): p. 4285-91.

180. Janka, A., et al., *Cytolethal distending toxin gene cluster in enterohemorrhagic Escherichia coli O157:H- and O157:H7: characterization and evolutionary considerations*. Infect Immun, 2003. **71**(6): p. 3634-8.
181. De Rycke, J. and E. Oswald, *Cytolethal distending toxin (CDT): a bacterial weapon to control host cell proliferation?* FEMS Microbiol Lett, 2001. **203**(2): p. 141-8.
182. Toth, I., et al., *Cytolethal distending toxin type I and type IV genes are framed with lambdoid prophage genes in extraintestinal pathogenic Escherichia coli*. Infect Immun, 2009. **77**(1): p. 492-500.
183. Friedrich, A.W., et al., *Cytolethal distending toxin in Escherichia coli O157:H7: spectrum of conservation, structure, and endothelial toxicity*. J Clin Microbiol, 2006. **44**(5): p. 1844-6.
184. Asakura, M., et al., *An inducible lambdoid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in enteropathogenic Escherichia coli*. Proc Natl Acad Sci U S A, 2007. **104**(36): p. 14483-8.
185. Lara-Tejero, M. and J.E. Galan, *CdtA, CdtB, and CdtC form a tripartite complex that is required for cytolethal distending toxin activity*. Infect Immun, 2001. **69**(7): p. 4358-65.
186. Nestic, D., Y. Hsu, and C.E. Stebbins, *Assembly and function of a bacterial genotoxin*. Nature, 2004. **429**(6990): p. 429-33.
187. Lee, R.B., et al., *Interactions of Campylobacter jejuni cytolethal distending toxin subunits CdtA and CdtC with HeLa cells*. Infect Immun, 2003. **71**(9): p. 4883-90.
188. Nestic, D. and C.E. Stebbins, *Mechanisms of assembly and cellular interactions for the bacterial genotoxin CDT*. PLoS Pathog, 2005. **1**(3): p. e28.
189. Elwell, C.A. and L.A. Dreyfus, *DNase I homologous residues in CdtB are critical for cytolethal distending toxin-mediated cell cycle arrest*. Mol Microbiol, 2000. **37**(4): p. 952-63.
190. Lara-Tejero, M. and J.E. Galan, *A bacterial toxin that controls cell cycle progression as a deoxyribonuclease I-like protein*. Science, 2000. **290**(5490): p. 354-7.
191. Scott, D.A. and J.B. Kaper, *Cloning and sequencing of the genes encoding Escherichia coli cytolethal distending toxin*. Infect Immun, 1994. **62**(1): p. 244-51.
192. Pickett, C.L., et al., *Cloning, sequencing, and expression of the Escherichia coli cytolethal distending toxin genes*. Infect Immun, 1994. **62**(3): p. 1046-51.
193. Ueno, Y., et al., *Biogenesis of the Actinobacillus actinomycetemcomitans cytolethal distending toxin holotoxin*. Infect Immun, 2006. **74**(6): p. 3480-7.
194. Lindmark, B., et al., *Outer membrane vesicle-mediated release of cytolethal distending toxin (CDT) from Campylobacter jejuni*. BMC Microbiol, 2009. **9**: p. 220.
195. McSweeney, L.A. and L.A. Dreyfus, *Carbohydrate-binding specificity of the Escherichia coli cytolethal distending toxin CdtA-II and CdtC-II subunits*. Infect Immun, 2005. **73**(4): p. 2051-60.
196. Mise, K., et al., *Involvement of ganglioside GM3 in G(2)/M cell cycle arrest of human monocytic cells induced by Actinobacillus actinomycetemcomitans cytolethal distending toxin*. Infect Immun, 2005. **73**(8): p. 4846-52.
197. Boesze-Battaglia, K., et al., *Cholesterol-rich membrane microdomains mediate cell cycle arrest induced by Actinobacillus actinomycetemcomitans cytolethal-distending toxin*. Cell Microbiol, 2006. **8**(5): p. 823-36.
198. Cortes-Bratti, X., et al., *Cellular internalization of cytolethal distending toxin from Haemophilus ducreyi*. Infect Immun, 2000. **68**(12): p. 6903-11.
199. McSweeney, L.A. and L.A. Dreyfus, *Nuclear localization of the Escherichia coli cytolethal distending toxin CdtB subunit*. Cell Microbiol, 2004. **6**(5): p. 447-58.

200. Guerra, L., et al., *Cellular internalization of cytolethal distending toxin: a new end to a known pathway*. Cell Microbiol, 2005. **7**(7): p. 921-34.
201. Nishikubo, S., et al., *An N-terminal segment of the active component of the bacterial genotoxin cytolethal distending toxin B (CDTB) directs CDTB into the nucleus*. J Biol Chem, 2003. **278**(50): p. 50671-81.
202. Haghjoo, E. and J.E. Galan, *Salmonella typhi encodes a functional cytolethal distending toxin that is delivered into host cells by a bacterial-internalization pathway*. Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4614-9.
203. Spano, S., J.E. Ugalde, and J.E. Galan, *Delivery of a Salmonella Typhi exotoxin from a host intracellular compartment*. Cell Host Microbe, 2008. **3**(1): p. 30-8.
204. Li, L., et al., *The Haemophilus ducreyi cytolethal distending toxin activates sensors of DNA damage and repair complexes in proliferating and non-proliferating cells*. Cell Microbiol, 2002. **4**(2): p. 87-99.
205. Hassane, D.C., R.B. Lee, and C.L. Pickett, *Campylobacter jejuni cytolethal distending toxin promotes DNA repair responses in normal human cells*. Infect Immun, 2003. **71**(1): p. 541-5.
206. Hassane, D.C., et al., *Cytolethal distending toxin demonstrates genotoxic activity in a yeast model*. Infect Immun, 2001. **69**(9): p. 5752-9.
207. Frisan, T., et al., *The Haemophilus ducreyi cytolethal distending toxin induces DNA double-strand breaks and promotes ATM-dependent activation of RhoA*. Cell Microbiol, 2003. **5**(10): p. 695-707.
208. Elwell, C., et al., *Escherichia coli CdtB mediates cytolethal distending toxin cell cycle arrest*. Infect Immun, 2001. **69**(5): p. 3418-22.
209. Shenker, B.J., et al., *A novel mode of action for a microbial-derived immunotoxin: the cytolethal distending toxin subunit B exhibits phosphatidylinositol 3,4,5-triphosphate phosphatase activity*. J Immunol, 2007. **178**(8): p. 5099-108.
210. Kitagawa, T., H. Hoshida, and R. Akada, *Genome-wide analysis of cellular response to bacterial genotoxin CdtB in yeast*. Infect Immun, 2007. **75**(3): p. 1393-402.
211. Alby, F., et al., *Study of the cytolethal distending toxin (CDT)-activated cell cycle checkpoint. Involvement of the CHK2 kinase*. FEBS Lett, 2001. **491**(3): p. 261-5.
212. Escalas, N., et al., *Study of the cytolethal distending toxin-induced cell cycle arrest in HeLa cells: involvement of the CDC25 phosphatase*. Exp Cell Res, 2000. **257**(1): p. 206-12.
213. Cortes-Bratti, X., et al., *The Haemophilus ducreyi cytolethal distending toxin induces cell cycle arrest and apoptosis via the DNA damage checkpoint pathways*. J Biol Chem, 2001. **276**(7): p. 5296-302.
214. Comayras, C., et al., *Escherichia coli cytolethal distending toxin blocks the HeLa cell cycle at the G2/M transition by preventing cdc2 protein kinase dephosphorylation and activation*. Infect Immun, 1997. **65**(12): p. 5088-95.
215. Smith, J.L. and D.O. Bayles, *The contribution of cytolethal distending toxin to bacterial pathogenesis*. Crit Rev Microbiol, 2006. **32**(4): p. 227-48.
216. Aragon, V., K. Chao, and L.A. Dreyfus, *Effect of cytolethal distending toxin on F-actin assembly and cell division in Chinese hamster ovary cells*. Infect Immun, 1997. **65**(9): p. 3774-80.
217. Johnson, J.R., et al., *Bacterial characteristics in relation to clinical source of Escherichia coli isolates from women with acute cystitis or pyelonephritis and uninfected women*. J Clin Microbiol, 2005. **43**(12): p. 6064-72.
218. Albert, M.J., et al., *Controlled study of cytolethal distending toxin-producing Escherichia coli infections in Bangladeshi children*. J Clin Microbiol, 1996. **34**(3): p. 717-9.



219. Ge, Z., et al., *Bacterial cytolethal distending toxin promotes the development of dysplasia in a model of microbially induced hepatocarcinogenesis*. Cell Microbiol, 2007. **9**(8): p. 2070-80.
220. Ge, Z., et al., *Cytolethal distending toxin is essential for Helicobacter hepaticus colonization in outbred Swiss Webster mice*. Infect Immun, 2005. **73**(6): p. 3559-67.
221. Young, V.B., et al., *In vitro and in vivo characterization of Helicobacter hepaticus cytolethal distending toxin mutants*. Infect Immun, 2004. **72**(5): p. 2521-7.
222. Pratt, J.S., et al., *Modulation of host immune responses by the cytolethal distending toxin of Helicobacter hepaticus*. Infect Immun, 2006. **74**(8): p. 4496-504.
223. McAuley, J.L., et al., *MUC1 cell surface mucin is a critical element of the mucosal barrier to infection*. J Clin Invest, 2007. **117**(8): p. 2313-24.
224. Fox, J.G., et al., *Gastroenteritis in NF-kappaB-deficient mice is produced with wild-type Campylobacter jejuni but not with C. jejuni lacking cytolethal distending toxin despite persistent colonization with both strains*. Infect Immun, 2004. **72**(2): p. 1116-25.
225. Jain, D., et al., *Differences in virulence attributes between cytolethal distending toxin positive and negative Campylobacter jejuni strains*. J Med Microbiol, 2008. **57**(Pt 3): p. 267-72.
226. Purdy, D., et al., *Characterisation of cytolethal distending toxin (CDT) mutants of Campylobacter jejuni*. J Med Microbiol, 2000. **49**(5): p. 473-9.
227. Wising, C., et al., *The cytolethal distending toxin of Haemophilus ducreyi aggravates dermal lesions in a rabbit model of chancroid*. Microbes Infect, 2005. **7**(5-6): p. 867-74.
228. Ohara, M., et al., *Caspase-2 and caspase-7 are involved in cytolethal distending toxin-induced apoptosis in Jurkat and MOLT-4 T-cell lines*. Infect Immun, 2004. **72**(2): p. 871-9.
229. Shenker, B.J., et al., *Actinobacillus actinomycetemcomitans immunosuppressive protein is a member of the family of cytolethal distending toxins capable of causing a G2 arrest in human T cells*. J Immunol, 1999. **162**(8): p. 4773-80.
230. Fernandes, K.P., et al., *Inhibition of interferon-gamma-induced nitric oxide production in endotoxin-activated macrophages by cytolethal distending toxin*. Oral Microbiol Immunol, 2008. **23**(5): p. 360-6.
231. Nougayrede, J.P., et al., *Type III secretion-dependent cell cycle block caused in HeLa cells by enteropathogenic Escherichia coli O103*. Infect Immun, 2001. **69**(11): p. 6785-95.
232. Marches, O., et al., *Enteropathogenic and enterohaemorrhagic Escherichia coli deliver a novel effector called Cif, which blocks cell cycle G2/M transition*. Mol Microbiol, 2003. **50**(5): p. 1553-67.
233. Loukiadis, E., et al., *Distribution, functional expression, and genetic organization of Cif, a phage-encoded type III-secreted effector from enteropathogenic and enterohemorrhagic Escherichia coli*. J Bacteriol, 2008. **190**(1): p. 275-85.
234. Jubelin, G., et al., *Cycle inhibiting factors (CIFs) are a growing family of functional cyclomodulins present in invertebrate and mammal bacterial pathogens*. PLoS One, 2009. **4**(3): p. e4855.
235. Yao, Q., et al., *A bacterial type III effector family uses the papain-like hydrolytic activity to arrest the host cell cycle*. Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3716-21.
236. Charpentier, X. and E. Oswald, *Identification of the secretion and translocation domain of the enteropathogenic and enterohemorrhagic Escherichia coli effector Cif*,

- using *TEM-1* beta-lactamase as a new fluorescence-based reporter. *J Bacteriol*, 2004. **186**(16): p. 5486-95.
237. Hsu, Y., et al., *Structure of the cyclomodulin Cif from pathogenic Escherichia coli*. *J Mol Biol*, 2008. **384**(2): p. 465-77.
  238. Crow, A., et al., *Crystal structures of Cif from bacterial pathogens Photorhabdus luminescens and Burkholderia pseudomallei*. *PLoS One*, 2009. **4**(5): p. e5582.
  239. Taieb, F., et al., *Escherichia coli cyclomodulin Cif induces G2 arrest of the host cell cycle without activation of the DNA-damage checkpoint-signalling pathway*. *Cell Microbiol*, 2006. **8**(12): p. 1910-21.
  240. Samba-Louaka, A., et al., *Bacterial cyclomodulin Cif blocks host cell cycle by stabilizing cyclin dependent kinase inhibitors p21(waf1) and p27(kip1)*. *Cell Microbiol*, 2008.
  241. Samba-Louaka, A., et al., *The enteropathogenic Escherichia coli effector Cif induces delayed apoptosis in epithelial cells*. *Infect Immun*, 2009. **77**(12): p. 5471-7.
  242. Cui, J., et al., *Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family*. *Science*, 2010. **329**(5996): p. 1215-8.
  243. Rabut, G. and M. Peter, *Function and regulation of protein neddylation. 'Protein modifications: beyond the usual suspects' review series*. *EMBO Rep*, 2008. **9**(10): p. 969-76.
  244. Oswald, E., et al., *Bacterial toxins that modulate host cell-cycle progression*. *Curr Opin Microbiol*, 2005. **8**(1): p. 83-91.
  245. Homburg, S., et al., *Expression analysis of the colibactin gene cluster coding for a novel polyketide in Escherichia coli*. *FEMS Microbiol Lett*, 2007. **275**(2): p. 255-62.
  246. Putze, J., et al., *Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae*. *Infect Immun*, 2009. **77**(11): p. 4696-703.
  247. Johnson, J.R., et al., *Molecular epidemiology and phylogenetic distribution of the Escherichia coli pks genomic island*. *J Clin Microbiol*, 2008. **46**(12): p. 3906-11.
  248. Hacker, J., U. Hentschel, and U. Dobrindt, *Prokaryotic chromosomes and disease*. *Science*, 2003. **301**(5634): p. 790-3.
  249. De Rycke, J., et al., *Mitotic block and delayed lethality in HeLa epithelial cells exposed to Escherichia coli BM2-1 producing cytotoxic necrotizing factor type 1*. *Infect Immun*, 1996. **64**(5): p. 1694-705.
  250. Cuevas-Ramos, G., et al., *Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells*. *Proc Natl Acad Sci U S A*, 2010. **107**(25): p. 11537-42.
  251. Hutter, R., *Overproduction of microbial metabolites*, in *Biotechnology, Microbial Products*, H.J. Rehm and J. Reed, Editors. 1986, Verlag chemie: Weinheim, Germany. p. 3-17.
  252. Clardy, J. and C. Walsh, *Lessons from natural molecules*. *Nature*, 2004. **432**(7019): p. 829-37.
  253. Fischbach, M.A., C.T. Walsh, and J. Clardy, *The evolution of gene collectives: How natural selection drives chemical innovation*. *Proc Natl Acad Sci U S A*, 2008. **105**(12): p. 4601-8.
  254. Nolan, E.M. and C.T. Walsh, *How nature morphs peptide scaffolds into antibiotics*. *ChemBiochem*, 2009. **10**(1): p. 34-53.
  255. Li, J.W. and J.C. Vederas, *Drug discovery and natural products: end of an era or an endless frontier?* *Science*, 2009. **325**(5937): p. 161-5.
  256. Harvey, A.L., *Natural products in drug discovery*. *Drug Discov Today*, 2008. **13**(19-20): p. 894-901.

257. Walsh, C.T. and M.A. Fischbach, *Natural products version 2.0: connecting genes to molecules*. J Am Chem Soc, 2010. **132**(8): p. 2469-93.
258. Flores-Sanchez, I.J. and R. Verpoorte, *Plant polyketide synthases: a fascinating group of enzymes*. Plant Physiol Biochem, 2009. **47**(3): p. 167-74.
259. Zaslaver, A., et al., *Just-in-time transcription program in metabolic pathways*. Nat Genet, 2004. **36**(5): p. 486-91.
260. Donadio, S., P. Monciardini, and M. Sosio, *Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics*. Nat Prod Rep, 2007. **24**(5): p. 1073-109.
261. Meier, J.L. and M.D. Burkart, *The chemical biology of modular biosynthetic enzymes*. Chem Soc Rev, 2009. **38**(7): p. 2012-45.
262. Keating, T.A. and C.T. Walsh, *Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis*. Curr Opin Chem Biol, 1999. **3**(5): p. 598-606.
263. Moore, B.S. and C. Hertweck, *Biosynthesis and attachment of novel bacterial polyketide synthase starter units*. Nat Prod Rep, 2002. **19**(1): p. 70-99.
264. Schwarzer, D., R. Finking, and M.A. Marahiel, *Nonribosomal peptides: from genes to products*. Nat Prod Rep, 2003. **20**(3): p. 275-87.
265. Samel, S.A., M.A. Marahiel, and L.O. Essen, *How to tailor non-ribosomal peptide products--new clues about the structures and mechanisms of modifying enzymes*. Mol Biosyst, 2008. **4**(5): p. 387-93.
266. Lambalot, R.H., et al., *A new enzyme superfamily - the phosphopantetheinyl transferases*. Chem Biol, 1996. **3**(11): p. 923-36.
267. Austin, M.B. and J.P. Noel, *The chalcone synthase superfamily of type III polyketide synthases*. Nat Prod Rep, 2003. **20**(1): p. 79-110.
268. Mootz, H.D., R. Finking, and M.A. Marahiel, *4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis*. J Biol Chem, 2001. **276**(40): p. 37289-98.
269. Tang, Y., S.C. Tsai, and C. Khosla, *Polyketide chain length control by chain length factor*. J Am Chem Soc, 2003. **125**(42): p. 12708-9.
270. Kohli, R.M. and C.T. Walsh, *Enzymology of acyl chain macrocyclization in natural product biosynthesis*. Chem Commun (Camb), 2003(3): p. 297-307.
271. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review*. Nat Prod Rep, 2001. **18**(4): p. 380-416.
272. Aron, Z.D., et al., *Characterization of a new tailoring domain in polyketide biogenesis: the amine transferase domain of MycA in the mycosubtilin gene cluster*. J Am Chem Soc, 2005. **127**(43): p. 14986-7.
273. Tang, L., et al., *Cloning and heterologous expression of the epothilone gene cluster*. Science, 2000. **287**(5453): p. 640-2.
274. Smith, S. and S.C. Tsai, *The type I fatty acid and polyketide synthases: a tale of two megasynthases*. Nat Prod Rep, 2007. **24**(5): p. 1041-72.
275. Kao, C.M., L. Katz, and C. Khosla, *Engineered biosynthesis of a complete macrolactone in a heterologous host*. Science, 1994. **265**(5171): p. 509-12.
276. Felnagle, E.A., et al., *Nonribosomal peptide synthetases involved in the production of medically relevant natural products*. Mol Pharm, 2008. **5**(2): p. 191-211.
277. Rausch, C., et al., *Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution*. BMC Evol Biol, 2007. **7**: p. 78.
278. Chooi, Y.H. and Y. Tang, *Adding the lipo to lipopeptides: do more with less*. Chem Biol, 2010. **17**(8): p. 791-3.

279. Keating, T.A., et al., *Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis*. *Chembiochem*, 2001. **2**(2): p. 99-107.
280. Sieber, S.A. and M.A. Marahiel, *Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides*. *J Bacteriol*, 2003. **185**(24): p. 7036-43.
281. Gaitatzis, N., B. Kunze, and R. Muller, *In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases*. *Proc Natl Acad Sci U S A*, 2001. **98**(20): p. 11136-41.
282. Stachelhaus, T. and C.T. Walsh, *Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase*. *Biochemistry*, 2000. **39**(19): p. 5775-87.
283. Clugston, S.L., et al., *Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst*. *Biochemistry*, 2003. **42**(41): p. 12095-104.
284. Balibar, C.J., F.H. Vaillancourt, and C.T. Walsh, *Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains*. *Chem Biol*, 2005. **12**(11): p. 1189-200.
285. Hoffmann, K., et al., *Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis*. *J Biol Chem*, 1994. **269**(17): p. 12710-4.
286. Gehring, A.M., et al., *The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of Yersinia pestis*. *Biochemistry*, 1998. **37**(33): p. 11637-50.
287. Du, L., et al., *An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in Streptomyces verticillus ATCC15003*. *FEMS Microbiol Lett*, 2000. **189**(2): p. 171-5.
288. Schneider, T.L., B. Shen, and C.T. Walsh, *Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation*. *Biochemistry*, 2003. **42**(32): p. 9722-30.
289. Reimann, C., et al., *Essential PchG-dependent reduction in pyochelin biosynthesis of Pseudomonas aeruginosa*. *J Bacteriol*, 2001. **183**(3): p. 813-20.
290. Miller, D.A., C.T. Walsh, and L. Luo, *C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase*. *J Am Chem Soc*, 2001. **123**(34): p. 8434-5.
291. Hacker, C., et al., *Mutational analysis of the N-methyltransferase domain of the multifunctional enzyme enniatin synthetase*. *J Biol Chem*, 2000. **275**(40): p. 30826-32.
292. Du, L., et al., *The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase*. *Chem Biol*, 2000. **7**(8): p. 623-42.
293. Tang, G.L., Y.Q. Cheng, and B. Shen, *Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase*. *Chem Biol*, 2004. **11**(1): p. 33-45.
294. Wu, K., et al., *The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units*. *Gene*, 2000. **251**(1): p. 81-90.

295. Miller, D.A., et al., *Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis*. Chem Biol, 2002. **9**(3): p. 333-44.
296. Walsh, C.T., *The chemical versatility of natural-product assembly lines*. Acc Chem Res, 2008. **41**(1): p. 4-10.
297. Neumann, C.S., D.G. Fujimori, and C.T. Walsh, *Halogenation strategies in natural product biosynthesis*. Chem Biol, 2008. **15**(2): p. 99-109.
298. Walsh, C.T., H.C. Losey, and C.L. Freel Meyers, *Antibiotic glycosyltransferases*. Biochem Soc Trans, 2003. **31**(Pt 3): p. 487-92.
299. Kim, B.S., et al., *Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin*. J Biol Chem, 2002. **277**(50): p. 48028-34.
300. Schwarzer, D., et al., *Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases*. Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14083-8.
301. Wenzel, S.C. and R. Muller, *Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic*. Curr Opin Chem Biol, 2005. **9**(5): p. 447-58.
302. Keating, T.A., et al., *Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain*. Biochemistry, 2000. **39**(9): p. 2297-306.
303. Piel, J., *Biosynthesis of polyketides by trans-AT polyketide synthases*. Nat Prod Rep, 2010. **27**(7): p. 996-1047.
304. Cheng, Y.Q., G.L. Tang, and B. Shen, *Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis*. Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3149-54.
305. Carvalho, R., et al., *The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation*. Gene, 2005. **359**: p. 91-8.
306. Butcher, R.A., et al., *The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis*. Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1506-9.
307. Beck, B.J., et al., *The hidden steps of domain skipping: macrolactone ring size determination in the pikromycin modular polyketide synthase*. Chem Biol, 2002. **9**(5): p. 575-83.
308. Wenzel, S.C., et al., *Structure and biosynthesis of myxochromides S1-3 in Stigmatella aurantiaca: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis*. ChemBiochem, 2005. **6**(2): p. 375-85.
309. He, J. and C. Hertweck, *Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster*. Chem Biol, 2003. **10**(12): p. 1225-32.
310. Gaitatzis, N., et al., *The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase*. J Biol Chem, 2002. **277**(15): p. 13082-90.
311. Olano, C., et al., *Evidence from engineered gene fusions for the repeated use of a module in a modular polyketide synthase*. Chem Commun (Camb), 2003(22): p. 2780-2.
312. Ullrich, M. and C.L. Bender, *The biosynthetic gene cluster for coronamic acid, an ethylcyclopropyl amino acid, contains genes homologous to amino acid-activating enzymes and thioesterases*. J Bacteriol, 1994. **176**(24): p. 7574-86.
313. Chen, H., et al., *Aminoacyl-S-enzyme intermediates in beta-hydroxylations and alpha,beta-desaturations of amino acids in peptide antibiotics*. Biochemistry, 2001. **40**(39): p. 11651-9.

314. Kleinkauf, H., J. Dittmann, and A. Lawen, *Cell-free biosynthesis of cyclosporin A and analogues*. Biomed Biochim Acta, 1991. **50**(10-11): p. S219-24.
315. Spivey, H.O. and J. Ovadi, *Substrate channeling*. Methods, 1999. **19**(2): p. 306-21.
316. Lai, J.R., A. Koglin, and C.T. Walsh, *Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis*. Biochemistry, 2006. **45**(50): p. 14869-79.
317. Weissman, K.J. and R. Muller, *Protein-protein interactions in multienzyme megasynthetases*. Chembiochem, 2008. **9**(6): p. 826-48.
318. Lomakin, I.B., Y. Xiong, and T.A. Steitz, *The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together*. Cell, 2007. **129**(2): p. 319-32.
319. Asturias, F.J., et al., *Structure and molecular organization of mammalian fatty acid synthase*. Nat Struct Mol Biol, 2005. **12**(3): p. 225-32.
320. Jenni, S., et al., *Architecture of a fungal fatty acid synthase at 5 Å resolution*. Science, 2006. **311**(5765): p. 1263-7.
321. Khosla, C., et al., *Structure and mechanism of the 6-deoxyerythronolide B synthase*. Annu Rev Biochem, 2007. **76**: p. 195-221.
322. Tanovic, A., et al., *Crystal structure of the termination module of a nonribosomal peptide synthetase*. Science, 2008. **321**(5889): p. 659-63.
323. Sieber, S.A., et al., *Evidence for a monomeric structure of nonribosomal Peptide synthetases*. Chem Biol, 2002. **9**(9): p. 997-1008.
324. Hillson, N.J. and C.T. Walsh, *Dimeric structure of the six-domain VibF subunit of vibriobactin synthetase: mutant domain activity regain and ultracentrifugation studies*. Biochemistry, 2003. **42**(3): p. 766-75.
325. Broadhurst, R.W., et al., *The structure of docking domains in modular polyketide synthases*. Chem Biol, 2003. **10**(8): p. 723-31.
326. Wu, N., D.E. Cane, and C. Khosla, *Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases*. Biochemistry, 2002. **41**(15): p. 5056-66.
327. Hahn, M. and T. Stachelhaus, *Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains*. Proc Natl Acad Sci U S A, 2004. **101**(44): p. 15585-90.
328. Richter, C.D., et al., *Multienzyme docking in hybrid megasynthetases*. Nat Chem Biol, 2008. **4**(1): p. 75-81.
329. Koglin, A., et al., *Conformational switches modulate protein interactions in peptide antibiotic synthetases*. Science, 2006. **312**(5771): p. 273-6.
330. Weissman, K.J., et al., *Evidence for a protein-protein interaction motif on an acyl carrier protein domain from a modular polyketide synthase*. Chem Biol, 2006. **13**(6): p. 625-36.
331. Waksman, S.A. and H.B. Woodruff, *The Soil as a Source of Microorganisms Antagonistic to Disease-Producing Bacteria*. J Bacteriol, 1940. **40**(4): p. 581-600.
332. Martinez, J.L., *Antibiotics and antibiotic resistance genes in natural environments*. Science, 2008. **321**(5887): p. 365-7.
333. Goh, E.B., et al., *Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics*. Proc Natl Acad Sci U S A, 2002. **99**(26): p. 17025-30.
334. Linares, J.F., et al., *Antibiotics as intermicrobial signaling agents instead of weapons*. Proc Natl Acad Sci U S A, 2006. **103**(51): p. 19484-9.

335. Fajardo, A. and J.L. Martinez, *Antibiotics as signals that trigger specific bacterial responses*. *Curr Opin Microbiol*, 2008. **11**(2): p. 161-7.
336. Keller, L. and M.G. Surette, *Communication in bacteria: an ecological and evolutionary perspective*. *Nat Rev Microbiol*, 2006. **4**(4): p. 249-58.
337. Kaufmann, G.F., et al., *Revisiting quorum sensing: Discovery of additional chemical and biological functions for 3-oxo-N-acylhomoserine lactones*. *Proc Natl Acad Sci U S A*, 2005. **102**(2): p. 309-14.
338. Cotter, P.D., C. Hill, and R.P. Ross, *Bacteriocins: developing innate immunity for food*. *Nat Rev Microbiol*, 2005. **3**(10): p. 777-88.
339. Pradel, E., et al., *Detection and avoidance of a natural product from the pathogenic bacterium *Serratia marcescens* by *Caenorhabditis elegans**. *Proc Natl Acad Sci U S A*, 2007. **104**(7): p. 2295-300.
340. Mazzola, M., et al., *Protozoan-induced regulation of cyclic lipopeptide biosynthesis is an effective predation defense mechanism for *Pseudomonas fluorescens**. *Appl Environ Microbiol*, 2009. **75**(21): p. 6804-11.
341. Matz, C., et al., *Impact of violacein-producing bacteria on survival and feeding of bacterivorous nanoflagellates*. *Appl Environ Microbiol*, 2004. **70**(3): p. 1593-9.
342. Raaijmakers, J.M., et al., *Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics*. *FEMS Microbiol Rev*, 2010.
343. Ron, E.Z. and E. Rosenberg, *Natural roles of biosurfactants*. *Environ Microbiol*, 2001. **3**(4): p. 229-36.
344. Tran, H., et al., *Role of the cyclic lipopeptide massetolide A in biological control of *Phytophthora infestans* and in colonization of tomato plants by *Pseudomonas fluorescens**. *New Phytol*, 2007. **175**(4): p. 731-42.
345. Nielsen, T.H., et al., *Genes involved in cyclic lipopeptide production are important for seed and straw colonization by *Pseudomonas* sp. strain DSS73*. *Appl Environ Microbiol*, 2005. **71**(7): p. 4112-6.
346. Braun, P.G., et al., *Evidence and characterization of a gene cluster required for the production of viscosin, a lipopeptide biosurfactant, by a strain of *Pseudomonas fluorescens**. *Can J Microbiol*, 2001. **47**(4): p. 294-301.
347. Julkowska, D., et al., *Comparative analysis of the development of swarming communities of *Bacillus subtilis* 168 and a natural wild type: critical effects of surfactin and the composition of the medium*. *J Bacteriol*, 2005. **187**(1): p. 65-76.
348. Neu, T.R., *Significance of bacterial surface-active compounds in interaction of bacteria with interfaces*. *Microbiol Rev* 1996. **60**: p. 151-66.
349. Hall-Stoodley, L., J.W. Costerton, and P. Stoodley, *Bacterial biofilms: from the natural environment to infectious diseases*. *Nat Rev Microbiol*, 2004. **2**(2): p. 95-108.
350. Lopez, D., et al., *Cannibalism enhances biofilm development in *Bacillus subtilis**. *Mol Microbiol*, 2009. **74**(3): p. 609-18.
351. Lopez, D., et al., *Paracrine signaling in a bacterium*. *Genes Dev*, 2009. **23**(14): p. 1631-8.
352. Nitschke, M., et al., *Surfactin reduces the adhesion of food-borne pathogenic bacteria to solid surfaces*. *Lett Appl Microbiol*, 2009. **49**(2): p. 241-7.
353. Mireles, J.R., 2nd, A. Toguchi, and R.M. Harshey, *Salmonella enterica serovar typhimurium swarming mutants with altered biofilm-forming abilities: surfactin inhibits biofilm formation*. *J Bacteriol*, 2001. **183**(20): p. 5848-54.
354. Straight, P.D., J.M. Willey, and R. Kolter, *Interactions between *Streptomyces coelicolor* and *Bacillus subtilis*: Role of surfactants in raising aerial structures*. *J Bacteriol*, 2006. **188**(13): p. 4918-25.

355. Hider, R.C. and X. Kong, *Chemistry and biology of siderophores*. Nat Prod Rep, 2010. **27**(5): p. 637-57.
356. Fischbach, M.A., et al., *How pathogenic bacteria evade mammalian sabotage in the battle for iron*. Nat Chem Biol, 2006. **2**(3): p. 132-8.
357. Flo, T.H., et al., *Lipocalin 2 mediates an innate immune response to bacterial infection by sequestering iron*. Nature, 2004. **432**(7019): p. 917-21.
358. Fischbach, M.A., et al., *The pathogen-associated iroA gene cluster mediates bacterial evasion of lipocalin 2*. Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16502-7.
359. Bais, H.P., R. Fall, and J.M. Vivanco, *Biocontrol of Bacillus subtilis against infection of Arabidopsis roots by Pseudomonas syringae is facilitated by biofilm formation and surfactin production*. Plant Physiol, 2004. **134**(1): p. 307-19.
360. Thrane, C., et al., *Viscosinamide-producing Pseudomonas fluorescens DR54 exerts a biocontrol effect on Pythium ultimum in sugar beet rhizosphere*. FEMS Microbiol Ecol, 2000. **33**(2): p. 139-146.
361. Ongena, M., et al., *Surfactin and fengycin lipopeptides of Bacillus subtilis as elicitors of induced systemic resistance in plants*. Environ Microbiol, 2007. **9**(4): p. 1084-90.
362. Partida-Martinez, L.P. and C. Hertweck, *Pathogenic fungus harbours endosymbiotic bacteria for toxin production*. Nature, 2005. **437**(7060): p. 884-8.
363. Partida-Martinez, L.P., et al., *Endosymbiont-dependent host reproduction maintains bacterial-fungal mutualism*. Curr Biol, 2007. **17**(9): p. 773-7.
364. George, K.M., et al., *Mycolactone: a polyketide toxin from Mycobacterium ulcerans required for virulence*. Science, 1999. **283**(5403): p. 854-7.
365. Pahlevan, A.A., et al., *The inhibitory action of Mycobacterium ulcerans soluble factor on monocyte/T cell cytokine production and NF-kappa B function*. J Immunol, 1999. **163**(7): p. 3928-35.
366. Coutanceau, E., et al., *Selective suppression of dendritic cell functions by Mycobacterium ulcerans toxin mycolactone*. J Exp Med, 2007. **204**(6): p. 1395-403.
367. Gooding, T.M., et al., *Immune response to infection with Mycobacterium ulcerans*. Infect Immun, 2001. **69**(3): p. 1704-7.
368. Marsollier, L., et al., *Aquatic insects as a vector for Mycobacterium ulcerans*. Appl Environ Microbiol, 2002. **68**(9): p. 4623-8.
369. Johnson, P.D., et al., *Mycobacterium ulcerans in mosquitoes captured during outbreak of Buruli ulcer, southeastern Australia*. Emerg Infect Dis, 2007. **13**(11): p. 1653-60.
370. Marsollier, L., et al., *Colonization of the salivary glands of Naucoris cimicoides by Mycobacterium ulcerans requires host plasmatocytes and a macrolide toxin, mycolactone*. Cell Microbiol, 2005. **7**(7): p. 935-43.
371. Hong, H., et al., *Mycolactones: immunosuppressive and cytotoxic polyketides produced by aquatic mycobacteria*. Nat Prod Rep, 2008. **25**(3): p. 447-54.
372. Daffe, M. and M.A. Laneelle, *Distribution of phthiocerol diester, phenolic mycosides and related compounds in mycobacteria*. J Gen Microbiol, 1988. **134**(7): p. 2049-55.
373. Brennan, P.J. and H. Nikaido, *The envelope of mycobacteria*. Annu Rev Biochem, 1995. **64**: p. 29-63.
374. Camacho, L.R., et al., *Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis*. Mol Microbiol, 1999. **34**(2): p. 257-67.
375. Cox, J.S., et al., *Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice*. Nature, 1999. **402**(6757): p. 79-83.
376. Astarie-Dequeker, C., et al., *Phthiocerol dimycocerosates of M. tuberculosis participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids*. PLoS Pathog, 2009. **5**(2): p. e1000289.



377. Rousseau, C., et al., *Production of phthiocerol dimycocerosates protects Mycobacterium tuberculosis from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection*. Cell Microbiol, 2004. **6**(3): p. 277-87.
378. Scholz-Schroeder, B.K., et al., *The contribution of syringopeptin and syringomycin to virulence of Pseudomonas syringae pv. syringae strain B301D on the basis of sypA and syrB1 biosynthesis mutant analysis*. Mol Plant Microbe Interact, 2001. **14**(3): p. 336-48.
379. Delfosse, V., et al., *Structure of the archaeal pab87 peptidase reveals a novel self-compartmentalizing protease family*. PLoS One, 2009. **4**(3): p. e4712.
380. Beadle, B.M., et al., *Structural milestones in the reaction pathway of an amide hydrolase: substrate, acyl, and product complexes of cephalothin with AmpC beta-lactamase*. Structure, 2002. **10**(3): p. 413-24.
381. Kevany, B.M., D.A. Rasko, and M.G. Thomas, *Characterization of the complete zwittermicin A biosynthesis gene cluster from Bacillus cereus*. Appl Environ Microbiol, 2009. **75**(4): p. 1144-55.
382. Zhou, Y., et al., *Novel roles of Bacillus thuringiensis to control plant diseases*. Appl Microbiol Biotechnol, 2008. **80**(4): p. 563-72.
383. Rogers, E.W., D.S. Dalisay, and T.F. Molinski, *(+)-Zwittermicin A: assignment of its complete configuration by total synthesis of the enantiomer and implication of D-serine in its biosynthesis*. Angew Chem Int Ed Engl, 2008. **47**(42): p. 8086-9.
384. Komeda, H. and Y. Asano, *Gene cloning, nucleotide sequencing, and purification and characterization of the D-stereospecific amino-acid amidase from Ochrobactrum anthropi SV3*. Eur J Biochem, 2000. **267**(7): p. 2028-35.
385. Asano, Y., et al., *Structural similarity of D-aminopeptidase to carboxypeptidase DD and beta-lactamases*. Biochemistry, 1992. **31**(8): p. 2316-28.
386. Asano, Y., et al., *An alkaline D-stereospecific endopeptidase with beta-lactamase activity from Bacillus cereus*. J Biol Chem, 1996. **271**(47): p. 30256-62.
387. Goffin, C. and J.M. Ghuysen, *Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs*. Microbiol Mol Biol Rev, 1998. **62**(4): p. 1079-93.
388. Raaijmakers, J.M., et al., *Natural functions of lipopeptides from Bacillus and Pseudomonas: more than surfactants and antibiotics*. FEMS Microbiol Rev, 2010. **34**(6): p. 1037-62.
389. Nielsen, T.H., et al., *Viscosinamide, a new cyclic depsipeptide with surfactant and antifungal properties produced by Pseudomonas fluorescens DR54*. J Appl Microbiol, 1999. **87**(1): p. 80-90.
390. Braun, V. and K. Hantke, *Biochemistry of bacterial cell envelopes*. Annu Rev Biochem, 1974. **43**(0): p. 89-121.
391. Simunovic, V., F.C. Gherardini, and L.J. Shimkets, *Membrane localization of motility, signaling, and polyketide synthetase proteins in Myxococcus xanthus*. J Bacteriol, 2003. **185**(17): p. 5066-75.
392. Straight, P.D., et al., *A singular enzymatic megacomplex from Bacillus subtilis*. Proc Natl Acad Sci U S A, 2007. **104**(1): p. 305-10.
393. Kiesau, P., et al., *Evidence for a multimeric subtilin synthetase complex*. J Bacteriol, 1997. **179**(5): p. 1475-81.
394. Engelke, G., et al., *Biosynthesis of the lantibiotic nisin: genomic organization and membrane localization of the NisB protein*. Appl Environ Microbiol, 1992. **58**(11): p. 3730-43.

395. Jain, M. and J.S. Cox, *Interaction between polyketide synthase and transporter suggests coupled synthesis and export of virulence lipid in M. tuberculosis*. PLoS Pathog, 2005. **1**(1): p. e2.
396. Lu, Y.W., A.K. San Roman, and A.M. Gehring, *Role of phosphopantetheinyl transferase genes in antibiotic production by Streptomyces coelicolor*. J Bacteriol, 2008. **190**(20): p. 6903-8.
397. Watanabe, K., *Exploring the biosynthesis of natural products and their inherent suitability for the rational design of desirable compounds through genetic engineering*. Biosci Biotechnol Biochem, 2008. **72**(10): p. 2491-506.
398. Challis, G.L., *Engineering Escherichia coli to produce nonribosomal peptide antibiotics*. Nat Chem Biol, 2006. **2**(8): p. 398-400.
399. Duquesne, S., et al., *Microcins, gene-encoded antibacterial peptides from enterobacteria*. Nat Prod Rep, 2007. **24**(4): p. 708-34.
400. Havarstein, L.S., D.B. Diep, and I.F. Nes, *A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export*. Mol Microbiol, 1995. **16**(2): p. 229-40.
401. Davidson, A.L., et al., *Structure, function, and evolution of bacterial ATP-binding cassette systems*. Microbiol Mol Biol Rev, 2008. **72**(2): p. 317-64, table of contents.
402. Saurin, W., M. Hofnung, and E. Dassa, *Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters*. J Mol Evol, 1999. **48**(1): p. 22-41.
403. Bourne, D.G., et al., *Characterisation of a gene cluster involved in bacterial degradation of the cyanobacterial toxin microcystin LR*. Environ Toxicol, 2001. **16**(6): p. 523-34.
404. Dantas, G., et al., *Bacteria subsisting on antibiotics*. Science, 2008. **320**(5872): p. 100-3.
405. Lapidus, A., et al., *Extending the Bacillus cereus group genomics to putative food-borne pathogens of different toxicity*. Chem Biol Interact, 2008. **171**(2): p. 236-49.
406. Qin, X., et al., *Genome sequences of the honey bee pathogens Paenibacillus larvae and Ascosphaera apis*. Insect Mol Biol, 2006. **15**(5): p. 715-8.
407. Jeong, H., et al., *Genomic blueprint of Hahella chejuensis, a marine microbe producing an algicidal agent*. Nucleic Acids Res, 2005. **33**(22): p. 7066-73.
408. Nakai, T., D. Yamauchi, and K. Kubota, *Enhancement of linear gramicidin expression from Bacillus brevis ATCC 8185 by casein peptide*. Biosci Biotechnol Biochem, 2005. **69**(4): p. 700-4.
409. Desvaux, M., *Clostridium cellulolyticum: model organism of mesophilic cellulolytic clostridia*. FEMS Microbiol Rev, 2005. **29**(4): p. 741-64.
410. Fischbach, M.A., *Antibiotics from microbes: converging to kill*. Curr Opin Microbiol, 2009. **12**(5): p. 520-7.
411. Giamarellos-Bourboulis, E.J., *Macrolides beyond the conventional antimicrobials: a class of potent immunomodulators*. Int J Antimicrob Agents, 2008. **31**(1): p. 12-20.
412. Foppen, J.W. and J.F. Schijven, *Evaluation of data from the literature on the transport and survival of Escherichia coli and thermotolerant coliforms in aquifers under saturated conditions*. Water Res, 2006. **40**(3): p. 401-26.
413. Schulenburg, H. and J.J. Ewbank, *The genetics of pathogen avoidance in Caenorhabditis elegans*. Mol Microbiol, 2007. **66**(3): p. 563-70.
414. Marcq, I., et al., *Génotoxicité in vitro et in vivo des souches de E. coli produisant la colibactine*, in 8ème Congrès National de la Société Française de Microbiologie. 2010: Marseille.
415. Labianca, R., et al., *Colon cancer*. Crit Rev Oncol Hematol, 2010. **74**(2): p. 106-33.

416. Karsa, L., et al., *The dimensions of the CRC problem*. Best Pract Res Clin Anaesthesiol, 2010. **24**(4): p. 381-96.
417. Kwak, E.L. and D.C. Chung, *Hereditary colorectal cancer syndromes: an overview*. Clin Colorectal Cancer, 2007. **6**(5): p. 340-4.
418. von Roon, A.C., et al., *The risk of cancer in patients with Crohn's disease*. Dis Colon Rectum, 2007. **50**(6): p. 839-55.
419. Eaden, J.A., K.R. Abrams, and J.F. Mayberry, *The risk of colorectal cancer in ulcerative colitis: a meta-analysis*. Gut, 2001. **48**(4): p. 526-35.
420. Lievre, A. and P. Laurent-Puig, [*Molecular biology in clinical cancer research: the example of digestive cancers*]. Rev Epidemiol Sante Publique, 2005. **53**(3): p. 267-82.
421. de la Chapelle, A., *Genetic predisposition to colorectal cancer*. Nat Rev Cancer, 2004. **4**(10): p. 769-80.
422. Karin, M. and F.R. Greten, *NF-kappaB: linking inflammation and immunity to cancer development and progression*. Nat Rev Immunol, 2005. **5**(10): p. 749-59.
423. Parkin, D.M., *The global health burden of infection-associated cancers in the year 2002*. Int J Cancer, 2006. **118**(12): p. 3030-44.
424. Herrera, L.A., et al., *Role of infectious diseases in human carcinogenesis*. Environ Mol Mutagen, 2005. **45**(2-3): p. 284-303.
425. Blaser, M.J., *Understanding microbe-induced cancers*. Cancer Prev Res (Phila), 2008. **1**(1): p. 15-20.
426. Lax, A.J., *Opinion: Bacterial toxins and cancer--a case to answer?* Nat Rev Microbiol, 2005. **3**(4): p. 343-9.
427. Dove, W.F., et al., *Intestinal neoplasia in the ApcMin mouse: independence from the microbial and natural killer (beige locus) status*. Cancer Res, 1997. **57**(5): p. 812-4.
428. Dubois, D., et al., *Cyclomodulins in urosepsis strains of Escherichia coli*. J Clin Microbiol, 2010. **48**(6): p. 2122-9.
429. Escobar-Paramo, P., et al., *Large-scale population structure of human commensal Escherichia coli isolates*. Appl Environ Microbiol, 2004. **70**(9): p. 5698-700.
430. Nowrouzian, F.L., A.E. Wold, and I. Adlerberth, *Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants*. J Infect Dis, 2005. **191**(7): p. 1078-83.
431. Gordon, D.M., S.E. Stern, and P.J. Collignon, *Influence of the age and sex of human hosts on the distribution of Escherichia coli ECOR groups and virulence traits*. Microbiology, 2005. **151**(Pt 1): p. 15-23.
432. Kurokawa, K., et al., *Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes*. DNA Res, 2007. **14**(4): p. 169-81.
433. Jeggo, P.A. and M. Lobrich, *DNA double-strand breaks: their cellular and clinical impact?* Oncogene, 2007. **26**(56): p. 7717-9.
434. Barker, N., et al., *Crypt stem cells as the cells-of-origin of intestinal cancer*. Nature, 2009. **457**(7229): p. 608-11.
435. Zhu, L., et al., *Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation*. Nature, 2009. **457**(7229): p. 603-7.
436. Wu, S., et al., *A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses*. Nat Med, 2009. **15**(9): p. 1016-22.
437. Anderson, G.G., et al., *Intracellular bacterial biofilm-like pods in urinary tract infections*. Science, 2003. **301**(5629): p. 105-7.
438. Rosen, D.A., et al., *Detection of intracellular bacterial communities in human urinary tract infection*. PLoS Med, 2007. **4**(12): p. e329.

439. Anderson, G.G., et al., *Intracellular bacterial communities of uropathogenic Escherichia coli in urinary tract pathogenesis*. Trends Microbiol, 2004. **12**(9): p. 424-30.
440. Elkahwaji, J.E., et al., *Chronic bacterial infection and inflammation incite reactive hyperplasia in a mouse model of chronic prostatitis*. Prostate, 2007. **67**(1): p. 14-21.
441. De Marzo, A.M., et al., *Inflammation in prostate carcinogenesis*. Nat Rev Cancer, 2007. **7**(4): p. 256-69.
442. Michiels, S., et al., *Genetic polymorphisms in 85 DNA repair genes and bladder cancer risk*. Carcinogenesis, 2009. **30**(5): p. 763-8.
443. Stabb, E.V. and J. Handelsman, *Genetic analysis of zwittermicin A resistance in Escherichia coli: effects on membrane potential and RNA polymerase*. Mol Microbiol, 1998. **27**(2): p. 311-22.
444. Carette, J.E., et al., *Haploid genetic screens in human cells identify host factors used by pathogens*. Science, 2009. **326**(5957): p. 1231-5.