Université de Montpellier - Ecole Doctorale GAIA

Habilitation à Diriger des Recherches
présentée par

Stéphane COCIANCICH
Chercheur au Cirad - UMR BGPI

About the added-value of genetics for the discovery and characterization of new natural products synthesized by the non-ribosomal peptide synthesis and the polyketide synthesis pathways in plant-associated bacteria

Soutenance le 13 mars 2019, devant le jury composé de :

Philippe BULET	Directeur de Recherche, CNRS, Biopark Archamps	Examinateur
Nicole COTTE-PATTAT	Directeur de Recherche, CNRS, UMR MAP, Villeurbanne	Examinateur
Valérie LECLÈRE	Professeur, Université de Lille	Examinateur
Eric OSWALD	Professeur, Université Paul Sabatier, Toulouse	Rapporteur
Jean-Luc PERNODET	Directeur de Recherche, CNRS, I2BC, Orsay	Rapporteur
Monique ROYER	Chercheur, CIRAD, UMR BGPI, Montpellier	Examinateur
Sabine SCHORR-GALINDO	Professeur, Université de Montpellier	Rapporteur
Roderich SÜSSMUTH	Professeur, Université Technique de Berlin	Examinateur

DEDICATION

To Anne, my beloved wife, for always being at my side

To Antoine and Mathilde, my children, for making me a proud father

To my parents and my grand-mother, for all they gave me

To the friendly memory of Shahab

Thanks to Kyeen M, Marion FLS and Alex C

ACKNOWLEDGEMENTS

I warmly thank Nicole COTTE-PATTAT, Valérie LECLÈRE, Monique ROYER, Sabine SCHORR-GALINDO, Philippe BULET, Eric OSWALD, Jean-Luc PERNODET and Roderich SÜSSMUTH for accepting to judge this work. I am very proud of this jury, and I want you to be sure that your presence means a lot to me, whether I know you since ages or since more recent times.

I am endlessly grateful to my mentors and team leaders for all they taught me: Philippe BULET, Charles HETRU and Jules HOFFMANN in Strasbourg, Dominique VIDAL and Max GOYFFON in Grenoble, Mohammed SHAHABUDDIN and Louis MILLER in Bethesda, Yvan RAHBÉ in Lyon, Monique ROYER and Philippe ROTT in Montpellier.

I want to acknowledge the full support of Roderich SÜSSMUTH and many of his co-workers in Berlin during our 13-year long (and still on-going) fruitful collaboration. Vielen Dank!

Special thanks to Jules HOFFMANN who trusted me at the first place, to Philippe BULET for being more than a supervisor, to Monique ROYER for her countless help and support, and to Stéphane BIROT for being brave enough to share my office since 2004.

Since 1990, I have worked alongside many researchers, technicians, post-docs, secretaries and, of course, students. I cannot cite them all but following is a shortlist of the closest colleagues from my former and collaborative teams as well as from the scientific environment (excluding the team leaders already listed above!). They made me feel like I was part of the family. Our interactions eventually resulted in ideas, projects, experiments and results (some of them being presented in this document). For some of these colleagues, it even turned into friendship. I hope I could give them back as much as they gave to me. A big thank you to all of them for their help and for turning worktime into nice moments: Serguey CHERNYSH, Pascale FEHLBAUM, Danièle HOFFMANNN, Jean LAMBERT, Marie-Ève MORITZ and Martine SCHNEIDER in Strasbourg, Alexandre COLS, Yves CROZET, Bertrand DUBOIS, Laurent LUTTENAUER and Dominique ROVARC'H in Grenoble, Catherine BARREAU, Yasmine BELKAÏD, Paul DOCHERTY, Robert GWADZ, André LAUGHINGHOUSE, Maria del Pilar VALENCIA, Jesus VALENZUELA and Helge "Dude" ZIELER in Bethesda, Jill TROYER in Baltimore, Kyeen MESESAN in Fort-Collins, Gérard DELOBEL and Abdelaziz HEDDI in Lyon, Tam DANG, Benjamin HEMPEL, Manuela HÜGELLAND, Martin JASYK, Andi MAINZ, Jonny NACHTIGALL, Alexander PESIC, Daniel PETRAS, Kathrin SCHNEIDER, Ginka VÖLLER, Leonard von ECKARDSTEIN and Kati WINTER in Berlin, Matthieu ARLAT in Toulouse, Marion FISCHER-LE SAUX and Marie-Agnès JACQUES in Angers, Delphine DESTOUMIEUX-GARZON, Alain GIVAUDAN and Ralf KOEBNIK in Montpellier.

Of course, UMR BGPI requires a paragraph on its own. Although it is impossible to list all those who work(ed) in building K since 2004 and who made this place so enjoyable to work at, I would like to particularly thank a few of them with whom links are tighter: Julie ARASTÉ, Florence BARTHOD, Stéphane BIROT, Katia BONNEMAYRE, Sonia BORRON, Geneviève BOURELLY, Marie-Line CARUANA, Matthieu CHABANNES, Patrice CHAMPOISEAU, Sandrine DUPLAN, Daniel GARGANI, Jean-Claude GIRARD, Thomas KROJ, Nathalie LABOUREAU, Dominique LAGRENÉE, Cathy LYONNET, Cyril MAGNO, Marie-Carmen MARTINEZ, Mélanie MARGUERETTAZ, Isabelle MEUSNIER, Jean-Benoît MOREL, Alexandre MORISSET, Xavier MOURICHON, Claire NEEMA, Julie NOËLL, Jean-Loup NOTTÉGHEM, Isabelle PIERETTI, Adeline RENIER, Philippe ROTT, Monique ROYER, Souhir SABRI, Véronique ROUSSEL, Didier THARREAU and Eric VIVIEN.

Finally, because research is a collective work, I want to address a warm thank you to all those who, as listed alphabetically below, have been, for once or more regularly, my 136 co-authors for publications or communications in meetings:

Julie ARASTÉ
Abdu AZAD
Rainer BISCHOFF
Stéphanie BOLOT
Tristan BOUREAU
Chrystelle BRIN
Sophie CESBRON
Serguey CHERNYSH
Armelle DARRASSE
Nicolas DENANCÉ
Sandrine DUPLAN
Henry FALES
Laura FLEITES
Dean GABRIEL
Alexandre GHAZI
Camila GOMEZ
Stefan GRÄTZ
Gilbert HÉGY
Danièle HOFFMANN
Manuela HÜGELLAND
Marie-Agnès JACQUES
Ralf KOEBNIK
Jean LAMBERT
Pierre LEFEUVRE
Elena LEVASHINA
Charles MANCEAU
Sandrine MÉGESSIER
Marc MÜLLER
Bruce NODEN
Serge OHRESSER
Xavier PERRIER
Isabelle PIERETTI
Perrine PORTIER
Jérôme PUIG
Jean-Marc REICHHART
Adrien RIEUX
Philippe ROTT
Ferd SAUBER
Béatrice SÉGURENS
Laurana SERRES-GIARDI
Boris SZUREK
Jill TROYER
Marie-Anne Van SLUYS
Dominique VIDAL
Ginka VÖLLER
Helge ZIELER

Matthieu ARLAT
Valérie BARBE
Sophie BONNEAU
Françoise BOUET
Jean-Paul BRIAND
Philippe BULET
Patrice CHAMPOISEAU
Arnaud COULOUX
Marie-Josée DARROUSSAT
Jean-Luc DIMARCQ
Annick DUPONT
David FIDDOCK
Stéphanie FOUTEAU
Lionel GAGNEVIN
Françoise GIGNOUX
Jérôme GOUZY
Fabien GUÉRIN
Benjamin HEMPEL
Jules HOFFMANN
Nicole HUGOUVIEUX-COTTE-PATTAT
Sebastian KEMPER
Simone KOENIG
René LANOT
Lucienne LETELLIER
Cyril MAGNO
Sophie MANGENOT
André MÉNEZ
Alejandra MUNOZ
Laurent NOËL
Soon PARK
Alexander PESIC
Paulo PIMENTA
Stéphane POUSSIER
Suzana RADULOVIC
Adeline RENIER
Isabelle ROBÈNE
Monique ROYER
Kathrin SCHNEIDER
Maria SEIDEL
Mohammed SHAHABUDDIN
François THIBAULT
Stefanie UHLMANN
Valérie VERDIER
Laura VIEWEG
Leonard Von ECKARDSTEIN

Mario ARRIETA-ORTIZ
Adriana BERNAL
François BONTEMS
Alice BOULANGER
Martial BRIAND
Sébastien CARRÈRE
Nicolas CHEN
Tam DANG
Jean-Heinrich DAUGROIS
Pieter DORRENSTEIN
Karine DURAND
Marion FISCHER-LE SAUX
Roger FRUTOS
Philippe GAYRAL
Jean-Claude GIRARD
Max GOYFFON
Endrick GUY
Charles HETRU
François HOLDER
Arnaud INDIANA
Dennis KERWAT
Julian KRETZ
Emmanuelle LAUBER
Philippe LETOURMY
Andi MAINZ
Mélanie MARGUERETTAZ
Alexandre MORISSET
Jonny NACHTIGALL
Julie NOËLL
Jean-Charles PAUCOD
Daniel PETRAS
Delphine PITORRE
Olivier PRUVOST
Saskia RAUSCH
Michel REULAND
Guillaume ROBIN
Souhir SABRI
Vivien SCHUBERT
Siamak SEMSARY
Roderich SÜSSMUTH
Isabelle THIBAULT
Alain Van DORSSELAER
Christian VERNIÈRE
Eric VIVIEN
John WESTON

TABLE OF CONTENTS

ABBREVIATIONS

aa..... aminoacid / acide aminé
AMP AntiMicrobial Peptide / peptide antimicrobien
BGC..... Biosynthetic Genes Cluster / groupement de gènes de biosynthèse
CFE Cell-Free Extracts / extraits dépourvus de cellules
Da..... dalton
Dab..... 2,4-diaminobutyric acid / acide 2,4-diaminobutyrique
Dhb..... 2,3-didehydroaminobutyric acid / acide 2,3 didéhydroaminobutyrique
Dpg..... 3,5-dihydroxyphenylglycine
GC Gas Chromatography / chromatographie en phase gazeuse
HPLC..... High Performance Liquid Chromatography / chromatographie liquide à haute performance
HSL HomoSerine Lactone
Hyn.....hydroxy-asparagine
kb..... kilobase
MS.....Mass Spectrometry / spectrométrie de masse
NMR.....Nuclear Magnetic Resonance / résonance magnétique nucléaire
NRP..... Non Ribosomal Peptide / peptide non ribosomique
NRPS..... Non Ribosomal Peptide Synthetase / synthétase peptidique non ribosomique
*p*ABA.....*para*-Amino Benzoic Acid / acide *para*-aminobenzoïque
PCDWE.....Plant Cell-Wall Degrading Enzyme / enzyme de dégradation de la paroi cellulaire des plantes
PK Polyketide / polycétide
PKS.....PolyKetide Synthase / synthase polycétidique
PPTase..... 4'-PhosphoPantetheinyl-Transferase
QS Quorum Sensing
TUB.....Technical University of Berlin / Université Technique de Berlin
v-ATPase.... vesicular ATPase / ATPase vésiculaire
WT..... wild-type / sauvage

CURRICULUM VITAE

Stéphane COCIANCICH

Né le 18 août 1969 à Strasbourg (67)
Marié, 2 enfants

Adresse professionnelle : Cirad – UMR BGPI
TA A-54/K
Campus International de Baillarguet
34398 MONTPELLIER Cedex 5
Tél : 04 99 62 48 15
E-mail: cociancich@cirad.fr

2000/… Chercheur en CDI au Cirad (ADR 2013)

Centre de Coopération Internationale en Recherche Agronomique pour le Développement
Département BIOS - UMR BGPI (Biologie et Génétique des Interactions Plante-Parasite) - (depuis 01/2003)
Campus International de Baillarguet - TA A-54/K - 34398 MONTPELLIER Cedex 5
Dir : Claire NEEMA (depuis 2013) - Philippe ROTT (2011 - 2013) - Jean-Loup NOTTEGHEM (2003 - 2010)
Equipe METABOL : Antibiotiques et autres métabolites secondaires chez les micro-organismes phytopathogènes

Chercheur spécialisé dans l'étude de métabolites secondaires chez les microorganismes associés aux plantes et/ou utilisés en bio-contrôle

Encadrement/co-encadrement de plusieurs étudiants (BTS, M1, M2, thèse, post-doc)

Affectation précédente (02/2000 – 12/2002):
Département AMIS - Unité Commune de Recherche et Développement CIRAD/Bayer CropScience
Dir : Roger FRUTOS

Autres fonctions au sein de l'UMR BGPI:
Correspondant Sécurité : depuis 01/2003 (depuis 06/2001 pour le Cirad)
Personne Compétente en Radioprotection : 06/2002 – 12/2007
Pilote du processus « Equipements » (Qualité : certification ISO9001) : 03/2005 – 03/2010
Membre du Conseil d'Unité (membre élu, représentant chercheurs) : 04/2005 – 12/2010
Membre du Conseil d'Unité (membre nommé, représentant sécurité) : 03/2011 – 12/2014
Membre du Conseil d'Unité (membre nommé, représentant équipe METABOL) : depuis 01/2015
Sauveteur Secouriste du Travail : depuis 04/2011

1999 Stage post-doctoral 2 (10 mois)

INRA/INSA - Laboratoire de Biologie Appliquée - VILLEURBANNE (69)
Dir : Gérard FEBVAY
Financement 1 : Région Rhône-Alpes "Contrat Emergence"
Financement 2 : Société de Secours des Amis des Sciences

Etude du mode d'action d'une protéine entomotoxique de la graine de pois

1995/98 Stage post-doctoral 1 (36 mois)

NIH - NIAID - Medical Entomology Section - Laboratory of Parasitic Diseases - BETHESDA, MD (USA)
Dir : Louis H MILLER
Financement : Fogarty International Center / NIH / John T & Catherine D McArthur Foundation

Etude du développement du parasite du paludisme chez le moustique

Encadrement/co-encadrement de plusieurs étudiants (L3, post-doc)

1994/95 Service National (Scientifique du Contingent)

Centre de Recherche du Service de Santé des Armées - Unité de Microbiologie - GRENOBLE (38)
Dir : Dominique VIDAL

Purification d'une exoprotéase chez la bactérie pathogène *Burkholderia pseudomallei*

1986/94 Etudes Universitaires (Université Louis Pasteur, Strasbourg 1)

1991/94 Doctorat de l'Université Louis Pasteur, Strasbourg 1, en biologie moléculaire

CNRS - Réponse Immunitaire et Développement chez les Insectes - IBMC - STRASBOURG (67)
Dir : Jules HOFFMANN
Financement : Ministère de la Recherche et de la Technologie

Isolement et caractérisation de peptides antibactériens inductibles chez les insectes

Soutenance le 23/09/1994 : Thèse obtenue avec la mention "Très honorable, avec les félicitations du jury"

Jury……	**Guy OURISSON**	*(Université Louis Pasteur, Strasbourg, France)*
	Fotis KAFATOS	*(European Molecular Biology Laboratory, Heidelberg, Allemagne)*
	Georges FREYSSINET	*(Rhône-Poulenc SA, Courbevoie, France)*
	Marius PTAK	*(Centre de Biophysique Moléculaire, Orléans, France)*
	Charles HETRU	*(CNRS, Strasbourg, France)*
	Louis MILLER	*(National Institutes of Health, Bethesda, MD, USA) (invité)*
	Jules HOFFMANN	*(CNRS, Strasbourg, France) (invité)*

Co-encadrement de plusieurs étudiants (DEA, thèse)

1990/91 DEA de Biologie Moléculaire et Cellulaire : *Option: Biologie Moléculaire du Développement (Mention AB)*

1989/90 Maitrise de Biochimie
UV Biochimie 3 (Mention AB) - UV Biochimie 4 (Mention AB) - UV Immunologie (Mention AB)
UV Génétique (Mention AB) - UV Microbiologie (Mention AB)

1988/89 Licence de Biologie Cellulaire et Physiologie
UV Biochimie 1 (Mention AB) - UV Biochimie 2 - UV Physiologie Animale 1

1986/88 DEUG Sciences de la Vie et de la Terre

1986 Baccalauréat (Serie D : Mathématiques / Sciences de la Vie et de la Terre)

LEGT Jean Monnet - STRASBOURG (67)

Encadrement (UMR BGPI, 2003-2018)

03-2018 / 08-2018 **Agathe Ropars** : Master 2 et diplôme d'ingénieur: Sur-expression d'un régulateur transcriptionnel de la famille AraC (M2 Microbiologie Fondamentale et Appliquée (Université de Rennes 1 / Université de Bretagne Occidentale) / (Ecole Supérieure d'Ingénieurs en Agroalimentaire de Bretagne Atlantique [ESIAB])

12-2017 / 04-2018 **Baptiste Durand** : Master 1: Recherche de loci codant de nouveaux métabolites secondaires chez les bactéries du genre *Streptomyces* (Master Biotin, Université de Nîmes, Mines d'Alès)

04-2016 / 05-2016 **Touhfa Fettouhi** : Licence 3: Etude de peptides non-ribosomiques produits par des bactéries (Université de Haute-Alsace).

03-2014 / 08-2014 **Alexandre Morisset** : Master 1: Etude du rôle du polymorphisme du locus LPS de *Xanthomonas albilineans* dans l'évolution et la pathogénie de cette bactérie responsable de l'échaudure des feuilles de la canne à sucre (Université Montpellier 2)

02-2013 / 07-2016 **Souhir Sabri** : Thèse (co-direction ADR) soutenue en juillet 2016: La biosynthèse non-ribosomique chez les bactéries du genre *Xanthomonas* (Montpellier SupAgro / ED GAIA)

01-2012 / 12-2015 **Isabelle Pieretti** : Thèse (co-encadrement; membre du comité de thèse): Génomique comparative et évolutive de *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre (Montpellier SupAgro / ED Sibaghe)

03-2012 / 06-2012 **Cyril Magno** : Master 1: Etude de trois loci NRPS identifiés dans le génome de *Xanthomonas albilineans*, bactérie pathogène de la canne à sucre (M1 Biologie des plantes tropicales / Université de Montpellier 2 / Montpellier SupAgro)

10-2011 / 12-2011 **Julie Arasté** : BTS 2: Purification des formes alpha et gamma de l'albicidine (BTS BioAnalyses et Contrôles / Lycée Jean Mermoz, Montpellier)

05-2011 / 07-2011 **Julie Arasté** : BTS 1: Purification des formes alpha de l'albicidine (BTS BioAnalyses et Contrôles / Lycée Jean Mermoz, Montpellier)

10-2010 / 11-2013 **Imène Mensi** : Thèse (co-encadrement): Localisation *in planta* de *Xanthomonas albilineans* et identification de déterminants moléculaires impliqués dans la colonisation épiphyte de sa plante hôte, la canne à sucre (Université de Montpellier II / ED Sibaghe)

01-2008 / 06-2008 **Jérôme Puig** : Master 2: Etude du système de sécrétion de type III de *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre (M2 Sciences et Technologies / Université de Montpellier II)

11-2007 / 12-2010 **Mélanie Marguerettaz** : Thèse (co-encadrement; membre du comité de thèse): Rôle du système de sécrétion de type III SPI-1 et des mégaenzymes NRPS dans le cycle de vie de *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre (Montpellier SupAgro / ED Sibaghe)

01-2004 / 05/2006 **Patrice Champoiseau** : Thèse (co-encadrement): *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre: caractérisation et variabilité des bases génétiques du pouvoir pathogène, en Guadeloupe et dans le monde (Université des Antilles et de la Guyane)

03-2004 / 06/2004 **Adeline Renier** : DEA: Diversité génétique de *Xanthomonas albilineans* et variabilité des enzymes de biosynthèse de l'albicidine, pathotoxine produite par l'agent causal de l'échaudure des feuilles de la canne à sucre (DEA Ressources Phytogénétiques et Interactions Biologiques / Agro Montpellier / Université de Montpellier II)

01-2003 / 10/2005 **Eric Vivien** : Thèse (co-encadrement; membre du comité de thèse): Biosynthèse de l'albicidine, une molécule antibiotique et une pathotoxine de *Xanthomonas albilineans*. (Université de Montpellier II)

Enseignement

Université de Montpellier - Master 2 « Biologie Agrosciences » - Spécialité IMHE « Interactions Micro-organismes, Hôtes et Environnement » - Unité d'Enseignement HMBA317 « Interactions des micro-organismes avec l'homme, les animaux et les plantes » (2018) **LA VOIE NON-RIBOSOMALE DE BIOSYNTHESE PEPTIDIQUE CHEZ LES BACTERIES PATHOGENES DE PLANTES.** 1h30, 01/10/2018

Université de Montpellier - Master 2 « Biologie Agrosciences » - Spécialité IMHE « Interactions Micro-organismes, Hôtes et Environnement » - Unité d'Enseignement HMBA317 « Interactions des micro-organismes avec l'homme, les animaux et les plantes » (2017) **LA VOIE NON-RIBOSOMALE DE BIOSYNTHESE PEPTIDIQUE CHEZ LES BACTERIES PATHOGENES DE PLANTES.** 1h15, 02/10/2017

Université de Montpellier - Master 2 « Biologie Agrosciences » - Spécialité IMHE « Interactions Micro-organismes, Hôtes et Environnement » - Unité d'Enseignement HMBA317 « Interactions des micro-organismes avec l'homme, les animaux et les plantes » (2016) **LA VOIE NON-RIBOSOMALE DE BIOSYNTHESE PEPTIDIQUE CHEZ LES BACTERIES PATHOGENES DE PLANTES.** 1h30, 03/10/2016

Université de Montpellier - Master 2 BPMBB « Biologie des Plantes et des Micro-organismes, Biotechnologies, Bioprocédés » - Spécialité IMHE « Interactions Micro-organismes, Hôtes et Environnement » - Unité d'Enseignement FMOM301 « Interactions des micro-organismes avec l'homme, les animaux et les plantes » (2014) **LA VOIE NON-RIBOSOMALE DE BIOSYNTHESE PEPTIDIQUE CHEZ LES BACTERIES PATHOGENES DE PLANTES.** 1h30, 29/09/2014

Relecture d'articles

Reviewer pour les revues ou éditeurs suivants : Bentham Science Publishers, Scientific Reports, Microbiological Research, Physiological and Molecular Plant Pathology, FEMS Microbiology Letters

Participation à des jurys

2018 BTS "Bioanalyses et Contrôles" - Lycée Jean Mermoz, Montpellier, France
2017 BTS "Bioanalyses et Contrôles" - Lycée Jean Mermoz, Montpellier, France
2016 BTS "Bioanalyses et Contrôles" - Lycée Jean Mermoz, Montpellier, France
2014 BTS "Bioanalyses et Contrôles" - Lycée Jean Mermoz, Montpellier, France
2012 BTS "Bioanalyses et Contrôles" - Lycée Jean Mermoz, Montpellier, France

Projets financés

Project submitted in 2019: ANR / DFG PRCI (Acronym: DickeyaVFM): Characterization and structure elucidation of a new quorum sensing system in plant pathogenic bacteria: 503 k€ (Coordinateur français)

2017-2019: Agropolis Foundation - OpenScience call (Acronym: CONTACT; 1605-052): Mycotoxins and fungal pathogens control by actinomycetes: 126 k€ (Coordinateur WP3)

2013-2016: ANR program "Investissements d'avenir" (Acronym: XANTHOMONAPEPTIDES; ANR-10-LABX-0001-01, project n° 1202-013): Deciphering a new family of small molecules produced by *Xanthomonas oryzae, Xanthomonas translucens* and *Xanthomonas albilineans*: 130 k€

2009-2012: ANR / DFG PCRI (Acronyme: ALBILINEANS; ANR-09-BLAN-0413-01; DFG SU 239/11-1) : Etude structurale et biologique de petites molécules impliquées dans la pathogénie de *Xanthomonas albilineans*, une bactérie pathogène de la canne à sucre : 213 k€

Organisation de congrès

1. **MEMBRE DU COMITE SCIENTIFIQUE.** AMP2016. 5th International Symposium on Antimicrobial Peptides. MONTPELLIER (France), 06-08/06/2016.

Langues

Anglais : Lu, parlé, écrit
Allemand : Lu, parlé, écrit

Prix

2010 Robert Antoine Award for Best Paper (Biology - XXVII ISSCT Congress, Veracruz, Mexique)
1999 NIH Fellows Award for Research Excellence

PUBLICATIONS LIST

Web of Science citations : 1468 (as of October 2, 2018) - h index : 17

Articles dans des revues internationales ou nationales avec comité de lecture

1. von Eckardstein L, Petras D, Dang T, Cociancich S, Sabri S, Grätz S, Kerwat D, Seidel M, Pesic A, Dorrenstein PC, Royer M, Weston JB & Süssmuth RD (2017) **TOTAL SYNTHESIS AND BIOLOGICAL ASSESSMENT OF NOVEL ALBICIDINS DISCOVERED BY MASS SPECTROMETRIC NETWORKING.** *Chemistry - A European Journal* **23**, 15316-15321.
IF 2017 : 5.160 - Web of Science citations : 2 - Doi: 10.1002/chem.201704074

2. Jacques M-A, Arlat M, Boulanger A, Boureau T, Carrère S, Cesbron S, Chen NWG, Cociancich S, Darrasse A, Denancé N, Fischer-Le Saux M, Gagnevin L, Koebnik R, Lauber E, Noël LD, Pieretti I, Portier P, Pruvost O, Rieux A, Robène I, Royer M, Szurek B, Verdier V & Vernière C (2016) **USING ECOLOGY, PHYSIOLOGY, AND GENOMICS TO UNDERSTAND HOST SPECIFICITY IN *XANTHOMONAS*.** *Annual Review of Phytopathology* **54**, 163-187. - Review
IF 2017 : 9.528 - Web of Science citations : 21 - Doi: 10.1146/annurev-phyto-080615-100147

3. Petras D, Kerwat D, Pesic A, Hempel B-F, von Eckardstein L, Semsary S, Arasté J, Marguerettaz M, Royer M, Cociancich S & Süssmuth RD (2016) **THE *O*-CARBAMOYL-TRANSFERASE ALB15 IS RESPONSIBLE FOR THE MODIFICATION OF ALBICIDIN.** *ACS Chemical Biology* **11**, 1198-1204.
IF 2017 : 4.592 - Web of Science citations : 7 - Doi: 10.1021/acschembio.5b01001

4. Pieretti I, Cociancich S, Bolot S, Carrère S, Morisset A, Rott P & Royer M (2015) **FULL GENOME SEQUENCE ANALYSIS OF TWO ISOLATES REVEALS A NOVEL *XANTHOMONAS* SPECIES CLOSE TO THE SUGARCANE PATHOGEN *XANTHOMONAS ALBILINEANS*.** *Genes* **6**, 714-733.
IF 2017 : 3.191 - Web of Science citations : 3 - Doi: 10.3390/genes6030714

5. Vieweg L, Kretz J, Pesic A, Kerwat D, Grätz S, Royer M, Cociancich S, Mainz A & Süssmuth RD (2015) **THE ALBICIDIN RESISTANCE FACTOR ALBD IS A SERINE ENDOPEPTIDASE THAT HYDROLYZES UNUSUAL OLIGOAROMATIC-TYPE PEPTIDES.** *Journal of the American Chemical Society* **137**, 7608-7611.
IF 2017 : 14.357 - Web of Science citations : 5 - Doi: 10.1021/jacs.5b04099

6. Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD & Cociancich S (2015) **WHAT MAKES *XANTHOMONAS ALBILINEANS* UNIQUE AMONGST XANTHOMONADS?** *Frontiers in Plant Science* **6**, 289. - Review
IF 2017 : 3.678 - Web of Science citations : 5 - Doi: 10.3389/fpls.2015.00289

7. Pieretti I, Bolot S, Carrère S, Barbe V, Cociancich S, Rott P & Royer M (2015) **DRAFT GENOME SEQUENCE OF *XANTHOMONAS SACCHARI* STRAIN LMG 476.** *Genome Announcements* **3**, e00146-15.
Web of Science citations : 3 - Doi: 10.1128/genomeA.00146-15

8. Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll J, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M & Süssmuth RD (2015) **THE GYRASE INHIBITOR ALBICIDIN CONSISTS OF *P*-AMINOBENZOIC ACIDS AND CYANOALANINE.** *Nature Chemical Biology* **11**, 195-197.
IF 2017 : 13.843 - Web of Science citations : 40 - Doi: 10.1038/nchembio.1734

9. Kretz J, Kerwat D, Schubert V, Grätz S, Pesic A, Semsary S, Cociancich S, Royer M & Süssmuth RD (2015) **TOTAL SYNTHESIS OF ALBICIDIN: A LEAD STRUCTURE FROM *XANTHOMONAS ALBILINEANS* FOR POTENT ANTIBACTERIAL GYRASE INHIBITORS.** *Angewandte Chemie International Edition* **54**, 1969-1973. **TOTALSYNTHESE VON ALBICIDIN – EINE LEITSTRUKTUR AUS *XANTHOMONAS ALBILINEANS* FÜR POTENTE ANTIBAKTERIELLE GYRASE-INHIBITOREN.** *Angewandte Chemie* **127**, 1992-1996. - Very important paper
IF 2017 : 12.102 - Web of Science citations : 18 - Doi: 10.1002/anie.201409584

10. Darrasse A, Carrère S, Barbe V, Boureau T, Arrieta-Ortiz ML, Bonneau S, Briand M, Brin C, Cociancich S, Durand K, Fouteau S, Gagnevin L, Guérin F, Guy E, Indiana A, Koebnik R, Lauber E, Munoz A, Noël LD, Pieretti I, Poussier S, Pruvost O, Robène-Soustrade I, Rott P, Royer M, Serres-Giardi L, Szurek B, Van Sluys M-A, Verdier V, Vernière C, Arlat M, Manceau C & Jacques M-A (2013) **GENOME SEQUENCE OF *XANTHOMONAS FUSCANS* SUBSP. *FUSCANS* STRAIN 4834-R REVEALS THAT FLAGELLAR MOTILITY IS NOT A GENERAL FEATURE OF XANTHOMONADS.** *BMC Genomics* **14**, 761. - Highly accessed paper
IF 2017 : 3.730 - Web of Science citations : 29 - Doi: 10.1186/1471-2164-14-761

11. Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin GP, Brin C, Carrere S, Gomez C, Hügelland M, Völler GH, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth RD & Cociancich S (2013) **GENOME MINING REVEALS THE GENUS *XANTHOMONAS* TO BE A PROMISING RESERVOIR FOR NEW BIOACTIVE NON-RIBOSOMALLY SYNTHESIZED PEPTIDES.** *BMC Genomics* **14**, 658. - Highly accessed paper
IF 2017 : 3.730 - Web of Science citations : 9 - Doi: 10.1186/1471-2164-14-658

12. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Poussier S, Ségurens B, Szurek B, Verdier V, Arlat M, Gabriel DW, Rott P & Cociancich S (2012) **GENOMIC INSIGHTS INTO STRATEGIES USED BY *XANTHOMONAS ALBILINEANS* WITH ITS REDUCED ARTILLERY TO SPREAD WITHIN SUGARCANE XYLEM VESSELS.** *BMC Genomics* **13**, 658.
IF 2017 : 3.730 - Web of Science citations : 22 - Doi: 10.1186/1471-2164-13-658

13. Marguerettaz M, Pieretti I, Gayral P, Puig J, Brin C, Cociancich S, Poussier S, Rott P & Royer M (2011) **GENOMIC AND EVOLUTIONARY FEATURES OF THE SPI-1 TYPE III SECRETION SYSTEM THAT IS PRESENT IN *XANTHOMONAS ALBILINEANS* BUT IS NOT ESSENTIAL FOR XYLEM COLONIZATION AND SYMPTOM DEVELOPMENT OF SUGARCANE LEAF SCALD.** *Molecular Plant-Microbe Interactions* **24**, 246-259.
IF 2017 : 3.588 - Web of Science citations : 17 - Doi: 10.1094/mpmi-08-10-0188

14. Rott P, Marguerettaz M, Fleites L, Cociancich S, Girard J-C, Pieretti I, Gabriel DW & Royer M (2011) **UNRAVELLING PATHOGENICITY OF *XANTHOMONAS ALBILINEANS*, THE CAUSAL AGENT OF SUGARCANE LEAF SCALD.** *International Sugar Journal* **113**, 490-496.
IF 2017 : 0.067 - Web of Science citations : 1

15. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M & Rott P (2009) **THE COMPLETE GENOME SEQUENCE OF *XANTHOMONAS ALBILINEANS* PROVIDES NEW INSIGHTS INTO THE REDUCTIVE GENOME EVOLUTION OF THE XYLEM-LIMITED *XANTHOMONADACEAE*.** *BMC Genomics* **10**, 616.
IF 2017 : 3.730 - Web of Science citations : 87 - Doi: 10.1186/1471-2164-10-616

16. Renier A, Vivien E, Cociancich S, Letourmy P, Perrier X, Rott PC & Royer M (2007) **SUBSTRATE SPECIFICITY-CONFERRING REGIONS OF THE NONRIBOSOMAL PEPTIDE SYNTHETASE ADENYLATION DOMAINS INVOLVED IN ALBICIDIN PATHOTOXIN BIOSYNTHESIS ARE HIGHLY CONSERVED WITHIN THE SPECIES *XANTHOMONAS ALBILINEANS*.** *Applied and Environmental Microbiology* **73**, 5523-5530.
IF 2017 : 3.633 - Web of Science citations : 7 - Doi: 10.1128/aem.00577-07

17. Vivien E, Pitorre D, Cociancich S, Pieretti I, Gabriel DW, Rott PC & Royer M (2007) **HETEROLOGOUS PRODUCTION OF ALBICIDIN: A PROMISING APPROACH TO OVERPRODUCING AND CHARACTERIZING THIS POTENT INHIBITOR OF DNA GYRASE.** *Antimicrobial Agents and Chemotherapy* **51**, 1549-1552.
IF 2017 : 4.255 - Web of Science citations : 18 - Doi: 10.1128/aac.01450-06

18. Champoiseau P, Daugrois J-H, Pieretti I, Cociancich S, Royer M & Rott P (2006) **HIGH VARIATION IN PATHOGENICITY OF GENETICALLY CLOSELY RELATED STRAINS OF *XANTHOMONAS ALBILINEANS*, THE SUGARCANE LEAF SCALD PATHOGEN, IN GUADELOUPE.** *Phytopathology* **96**, 1081-1091.
IF 2017 : 3.036 - Web of Science citations : 16 - Doi: 10.1094/phyto-96-1081

19. Vivien E, Megessier S, Pieretti I, Cociancich S, Frutos R, Gabriel DW, Rott PC & Royer M (2005) ***XANTHOMONAS ALBILINEANS* HTPG IS REQUIRED FOR BIOSYNTHESIS OF THE ANTIBIOTIC AND PHYTOTOXIN ALBICIDIN.** *FEMS Microbiology Letters* **251**, 81-89.
IF 2017 : 1.735 - Web of Science citations : 11 - Doi: 10.1016/j.femsle.2005.07.026

20. Cociancich SO, Park SS, Fiddock DA and Shahabuddin M (1999) **VESICULAR ATPASE-OVEREXPRESSING CELLS DETERMINE THE DISTRIBUTION OF MALARIA PARASITE OOCYSTS ON THE MIDGUTS OF MOSQUITOES.** *Journal of Biological Chemistry* **274**, 12650-12655.
IF 2017 : 4.010 - Web of Science citations : 30 - Doi: 10.1074/jbc.274.18.12650

21. Shahabuddin M, Cociancich S & Zieler H (1998) **THE SEARCH FOR NOVEL MALARIA TRANSMISSION-BLOCKING TARGETS IN THE MOSQUITO MIDGUT.** *Parasitology Today.* **14**, 493-497. - Review - Cover
IF 2017 : 7.929 - Web of Science citations : 27 - Doi: 10.1016/s0169-4758(98)01348-9

22. Chernysh S, Cociancich S, Briand J-P, Hetru C & Bulet P (1996) **THE INDUCIBLE ANTIBACTERIAL PEPTIDES OF THE HEMIPTERAN INSECT *PALOMENA PRASINA*: IDENTIFICATION OF A UNIQUE FAMILY OF PROLINE-RICH PEPTIDES AND OF A NOVEL INSECT DEFENSIN.** *Journal of Insect Physiology* **42**, 81-89.
IF 2017 : 2.733 - Web of Science citations : 67 - Doi: 10.1016/0022-1910(95)00085-2

23. Cociancich S, Dupont A, Hégy G, Lanot R, Holder F, Hetru C, Hoffmann JA & Bulet P (1994) **NOVEL INDUCIBLE ANTIBACTERIAL PEPTIDES FROM A HEMIPTERAN INSECT, THE SAP-SUCKING BUG *PYRRHOCORIS APTERUS*.** *Biochemical Journal* **300**, 567-575.
IF 2017 : 3.857 - Web of Science citations : 161 - Doi: 10.1042/bj3000567

24. Cociancich S, Bulet P, Hetru C and Hoffmann JA (1994) **THE INDUCIBLE ANTIBACTERIAL PEPTIDES OF INSECTS.** *Parasitology Today* **10**, 132-139. - Review - Cover
IF 2017 : 7.929 - Web of Science citations : 179 - Doi: 10.1016/0169-4758(94)90260-7

25. Cociancich S, Ghazi A, Hetru C, Hoffmann JA & Letellier L (1993) **INSECT DEFENSIN, AN INDUCIBLE ANTIBACTERIAL PEPTIDE, FORMS VOLTAGE-DEPENDENT CHANNELS IN *MICROCOCCUS LUTEUS*.** *Journal of Biological Chemistry* **268**, 19239-19245.
IF 2017 : 4.010 - Web of Science citations : 260

26. Cociancich S, Goyffon M, Bontems F, Bulet B, Bouet F, Menez A & Hoffmann J (1993) **PURIFICATION AND CHARACTERIZATION OF A SCORPION DEFENSIN, A 4 KDA ANTIBACTERIAL PEPTIDE PRESENTING STRUCTURAL SIMILARITIES WITH INSECT DEFENSINS AND SCORPION TOXINS.** *Biochemical and Biophysical Research Communications* **194**, 17-22.
IF 2017 : 2.559 - Web of Science citations : 126 - Doi: 10.1006/bbrc.1993.1778

27. Bulet P, Cociancich S, Reuland M, Sauber F, Bischoff R, Hégy G, Van Dorsselaer A, Hetru C & Hoffmann JA (1992) **A NOVEL INSECT DEFENSIN MEDIATES THE INDUCIBLE ANTIBACTERIAL ACTIVITY IN LARVAE OF THE DRAGONFLY *AESCHNA CYANEA* (PALEOPTERA, ODONATA).** *European Journal of Biochemistry* **209**, 977-984.
IF 2017 : 4.530 - Web of Science citations : 107 - Doi: 10.1111/j.1432-1033.1992.tb17371.x

28. Bulet P, Cociancich S, Dimarcq J-L, Lambert J, Reichhart J-M, Hoffmann D, Hetru C & Hoffmann JA (1991) **INSECT IMMUNITY. ISOLATION FROM A COLEOPTERAN INSECT OF A NOVEL INDUCIBLE ANTIBACTERIAL PEPTIDE AND OF NEW MEMBERS OF THE INSECT DEFENSIN FAMILY.** *Journal of Biological Chemistry* **266**, 24520-24525.
IF 2017 : 4.010 - Web of Science citations : 153

Articles dans des revues sans comité de lecture

1. Thibault F, Cociancich S, Paucod J-C, Gignoux F, Thibault I & Vidal D (1996) **PURIFICATION D'UNE EXOPROTEASE DE 42 KDA PRODUITE PAR *BURKHOLDERIA* (*PSEUDOMONAS*) *PSEUDOMALLEI*.** *Travaux Scientifiques des Chercheurs du Service de Santé des Armées* **17**, 39-40.
Web of Science citations : 0

Communications orales invitées dans un congrès national ou international

1. Cociancich S, Pesic A, Petras D, Duplan S, Hempel B, Hügelland M, Kemper S, Kerwat D, Kretz J, Mainz A, Marguerettaz M, Noëll J, Pieretti I, Schubert V, Semsary S, Uhlmann S, Vieweg L, von Eckardstein L, Rott P, Royer M & Süssmuth RD (2016) **MULTIDISCIPLINARY APPROACHES SHED LIGHT ON BIOSYNTHESIS AND UNUSUAL STRUCTURAL FEATURES OF THE ANTIMICROBIAL HYBRID PKS/NRPS-SYNTHESIZED ALBICIDIN.** AMP2016. 5th International Symposium on Antimicrobial Peptides. MONTPELLIER (France), 06-08/06/2016.

2. Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin GP, Brin C, Carrere S, Gomez C, Hügelland M, Völler GH, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth RD & Cociancich S (2013) **GENOME MINING REVEALS A NOVEL AND PROMISING NRPS GENE CLUSTER IN *XANTHOMONAS ALBILINEANS*, *XANTHOMONAS ORYZAE* AND *XANTHOMONAS TRANSLUCENS*.** 59° Congresso Brasileiro de Genética. AGUAS DE LINDOIA, SAO PAULO (Brésil), 16-19/09/2013.

3. Cociancich S & Shahabuddin M (2000) **DECIPHERING *PLASMODIUM*'S DEVELOPMENT IN ITS MOSQUITO VECTOR: NEW TOOLS FOR FIGHTING MALARIA.** XVth International Congress for Tropical Medicine and Malaria. CARTAGENA DE INDIAS (Colombie), 20-25/08/2000.

Communications orales dans un congrès international ou national

1. Sabri S, Royer M, Pesic A, Petras D, Mainz A, Süssmuth RD & Cociancich S (2018) **GENOME-BASED SURVEY OF NON-RIBOSOMAL PEPTIDE SYNTHETASE GENES SHARING FASCINATING FEATURES WITH THE ALBICIDIN BIOSYNTHESIS GENES.** 13èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 29/01-02/02/2018.

2. Cociancich S (2017) **MULTIDISCIPLINARY APPROACHES SHED LIGHT ON BIOSYNTHESIS AND UNUSUAL STRUCTURAL FEATURES OF THE ANTIMICROBIAL HYBRID PKS/NRPS-SYNTHESIZED ALBICIDIN.** MicrobiOccitanie 2017. 1ère Rencontre des Microbiologistes de la Région Occitanie. TOULOUSE (France), 24-26/04/2017.

3. Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll N, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M & Süssmuth RD (2016) **FULL ELUCIDATION OF THE HITHERTO UNKNOWN STRUCTURE OF ALBICIDIN, A POTENT ANTIBIOTIC PRODUCED BY *XANTHOMONAS ALBILINEANS*.** 12èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 11-15/01/2016.

4. Petras D, Hempel B, Kerwat D, Mainz A, Cociancich S, Royer M & Süssmuth RD (2015) **THE BIOSYNTHESIS OF ALBICIDINS.** 2nd European Conference on Natural Products. FRANKFURT-AM-MAIN (Allemagne), 06-09/09/2015.

5. Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll N, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M & Süssmuth (2015) **FULL ELUCIDATION OF THE HITHERTO UNKNOWN STRUCTURE OF ALBICIDIN, A POTENT ANTIBIOTIC PRODUCED BY *XANTHOMONAS ALBILINEANS*.** 5th *Xanthomonas* Genomics Conference. BOGOTA (Colombie), 08-11/07/2015.

6. Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll N, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M & Süssmuth (2015) **FULL ELUCIDATION OF THE HITHERTO UNKNOWN STRUCTURE OF ALBICIDIN, A POTENT ANTIBIOTIC PRODUCED BY *XANTHOMONAS ALBILINEANS*.** 9ème Colloque de la Société Française de Phytopathologie. COLMAR (France), 02-05/06/2015.

7. Sabri S, Royer M, Völler GH, Gomez C, Hügelland M, Noëll J, Süssmuth RD & Cociancich S (2014) **VERS LA CARACTERISATION DE LA FONCTION D'UNE NOUVELLE PETITE MOLECULE SYNTHETISEE PAR DES NRPS CHEZ *XANTHOMONAS ALBILINEANS*, LA BACTERIE RESPONSABLE DE L'ECHAUDURE DES FEUILLES DE LA CANNE A SUCRE.** 11èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 03-07/02/2014.

8. Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin GP, Brin C, Carrere S, Gomez C, Hügelland M, Völler G, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth RD & Cociancich S (2012) **GENOME MINING INDICATES THAT THE GENUS *XANTHOMONAS* IS A PROMISING RESERVOIR FOR NEW BIOACTIVE NON-RIBOSOMALLY SYNTHESIZED PEPTIDES.** 4th *Xanthomonas* Genomics Conference. ANGERS (France), 09-12/07/2012.

9. Darrasse A, Carrère S, Barbe V, Boureau T, Bernal A, Bonneau S, Brin C, Cociancich S, Durand K, Fouteau S, Gagnevin L, Gouzy J, Guérin F, Guy E, Indiana I, Koebnik R, Lauber E, Munoz A, Noël LD, Pieretti I, Poussier S, Pruvost O, Robène-Soustrade I, Rott P, Royer M, Szurek B, Van Sluys M-A, Verdier V, Vernière C, Arlat M, Manceau C & Jacques M-A (2012) **GENOME SEQUENCING OF *XANTHOMONAS AXONOPODIS* PV. *PHASEOLI* CFBP4834-R REVEALS THAT FLAGELLAR MOTILITY IS NOT A GENERAL FEATURE OF XANTHOMONADS.** 4th *Xanthomonas* Genomics Conference. ANGERS (France), 09-12/07/2012.

10. Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin G, Brin C, Carrère S, Gomez C, Hügelland M, Völler G, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth R & Cociancich S (2012) **GENOME MINING INDICATES THAT THE GENUS *XANTHOMONAS* IS A PROMISING RESERVOIR FOR NEW BIOACTIVE NON-RIBOSOMALLY SYNTHESIZED PEPTIDES.** 8ème Colloque de la Société Française de Phytopathologie. PARIS (France), 05-08/06/2012.

11. Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin GP, Brin C, CarrEre S, Gomez C, Hügelland M, Völler G, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth RD & Cociancich S (2012) **GENOME MINING INDICATES THAT THE GENUS *XANTHOMONAS* IS A PROMISING RESERVOIR FOR NEW BIOACTIVE NON-RIBOSOMALLY SYNTHESIZED PEPTIDES.** COST Action CM0804 meeting « Chemistry and target identification of natural products ». BUCAREST (Roumanie), 22-24/05/2012.

12. Cociancich S, Nachtigall J, Marguerettaz M, Duplan S, Pieretti I, Rott P, Süssmuth RD & Royer M (2010) **DECIPHERING SMALL MOLECULES AS NEW VIRULENCE FACTORS IN THE BACTERIAL SUGARCANE PATHOGEN *XANTHOMONAS ALBILINEANS*.** 12th International Conference on Plant Pathogenic Bacteria. SAINT-DENIS DE LA REUNION (France), 07-11/06/2010.

13. Rott P, Marguerettaz M, Fleites L, Cociancich S, Girard J-C, Pieretti I, Gabriel DW & Royer M (2010) **UNRAVELLING PATHOGENICITY OF *XANTHOMONAS ALBILINEANS*, THE CAUSAL AGENT OF SUGARCANE LEAF SCALD.** XXVII International Society of Sugar Cane Technologists Congress. VERACRUZ (Mexique), 07-11/03/2010.

14. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Poussier S, Ségurens B, Szurek B, Verdier V, Arlat M & Rott P (2009) **LA SEQUENCE COMPLETE DU GENOME DE *XANTHOMONAS ALBILINEANS* APPORTE UN ECLAIRAGE NOUVEAU SUR L'EVOLUTION DES XANTHOMONADACEAE DONT L'HABITAT EST LIMITE AU XYLEME.** Phytopathologie 2009 - 7ème Colloque National de la Société Française de Phytopathologie. LYON (France), 08-11/06/2009.

15. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Poussier S, Ségurens B, Szurek B, Verdier V, Arlat M & Rott P (2008) **LA SEQUENCE DU GENOME DE *XANTHOMONAS ALBILINEANS* DEVOILE DES PARTICULARITES SURPRENANTES CHEZ CETTE BACTERIE PATHOGENE DE LA CANNE A SUCRE.** 8èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 14-18/01/2008.

16. Vivien E, Cociancich S, Duplan S, Pieretti I, Gabriel DW, Rott PC & Royer M (2006) **EXPRESSION DE L'ENSEMBLE DES GENES DE BIOSYNTHESE DE L'ALBICIDINE CHEZ UN HOTE HETEROLOGUE.** 7èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 20-24/03/2006.

17. Shahabuddin M, Pimenta P, Zieler H & Cociancich S (1998) ***PLASMODIUM* DEVELOPMENT IN THE MOSQUITO MIDGUT: IMPLICATIONS IN DEVELOPING NEW MALARIA-TRANSMISSION BLOCKING STRATEGIES.** XIV Meeting of Brazilian Society of Protozoology. CAXAMBU, MG (Brésil), 11-13/11/1998.

18. Cociancich SO & Shahabuddin M (1998) **A NOVEL FACTOR FROM *AEDES* AND *ANOPHELES* MOSQUITOES BLOCKS THE ZYGOTE TO OOKINETE TRANSFORMATION OF *PLASMODIUM GALLINACEUM*.** 47th Annual Meeting of the American Society of Tropical Medicine and Hygiene. SAN JUAN, PUERTO RICO (USA), 18-22/10/1998. Congrès annulé.

19. Cociancich SO, Park SS & Shahabuddin M (1998) **VESICULAR ATPASE OVER-EXPRESSING CELLS IN THE MIDGUT OF MOSQUITOES: RELATIONSHIP TO MALARIA TRANSMISSION.** 47th Annual Meeting of the American Society of Tropical Medicine and Hygiene. SAN JUAN, PUERTO RICO (USA), 18-22/10/1998. Congrès annulé.

20. Thibault F, Cociancich S, Paucod J-C, Gignoux F, Thibault I & Vidal D (1995) **PURIFICATION OF A 42,000 MOL. WT PROTEASE FROM *BURKHOLDERIA PSEUDOMALLEI*.** 3èmes Journées des Rencontres en Toxinologie (Société Française pour l'Etude des Toxines) « Molecular Analysis of the Relations between Structure and Functions in Toxins ». PARIS (France), 07-08/12/1995.

21. Chernysh S, Cociancich S, Briand J-P, Hetru C and Bulet P (1994) **THE INDUCIBLE ANTIBACTERIAL PEPTIDES OF THE HEMIPTERAN INSECT *PALOMENA PRASINA*: IDENTIFICATION OF A UNIQUE FAMILY OF PROLINE-RICH PEPTIDES AND OF A NOVEL INSECT DEFENSIN.** 42nd National Meeting of the Entomological Society of America "Mediators of insect immunity". DALLAS, TX (USA), 13-17/12/1994.

22. Cociancich S (1993) **ETUDE DU MODE D'ACTION DES DEFENSINES D'INSECTES.** Réunion du Mont Sainte-Odile « Aspects Structuraux, Cellulaires et Moléculaires de la Biologie des Invertébrés ». MONT SAINTE-ODILE (France), 14-16/03/1993.

23. Bulet P & Cociancich S (1993) **REPONSE IMMUNITAIRE DE LA DROSOPHILE: CARACTERISATION COMPLETE D'UN NOUVEAU PEPTIDE ANTIBACTERIEN O-GLYCOLSYLE, LA DROSOCINE.** Réunion du Mont Sainte-Odile « Aspects Structuraux, Cellulaires et Moléculaires de la Biologie des Invertébrés ». MONT SAINTE-ODILE (France), 14-16/03/1993.

24. Dimarcq J-L, Bulet P, Cociancich S, Hetru C, Hoffmann D, Reichhart J-M & Hoffmann J (1992) **THE INSECT DEFENSINS: INDUCIBLE ANTIBACTERIAL PEPTIDES OF THE INSECT HOST DEFENSE.** 25th Annual Meeting of the Society for Invertebrate Pathology. HEIDELBERG (Allemagne), 16-21/08/1992.

Communications par affiches dans un congrès international ou national

1. Hugouvieux-Cotte-Pattat N, Royer M & Cociancich S (2018) **COMMUNICATION INTERCELLULAIRE DANS LE GENRE *DICKEYA* : INTERCHAGEABILITE DU SIGNAL SPECIFIQUE ?.** 13èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 29/01-02/02/2018.

2. Dang T, Petras D, Pesic A, Cociancich S, Royer M, Dorrenstein P & Süssmuth R (2016) **BIOACTIVITY-GUIDED MASS SPECTRAL NETWORKING REVEALS A NEW SET OF ALBICIDIN DERIVATIVES FROM *XANTHOMONAS ALBILINEANS*.** 34th European Peptide Symposium & 8th International Peptide Symposium. LEIPZIG (Allemagne), 04-09/09/2016.

3. Hempel B-F, Petras D, Kerwat D, Pesic A, Semsary S, von Eckardstein L, Arasté J, Marguerettaz M, Royer M, Cociancich S & Süssmuth R (2016) **THE O-CARBAMOYL-TRANSFERASE ALB15 IS RESPONSIBLE FOR THE MODIFICATION OF ALBICIDIN.** 34th European Peptide Symposium & 8th International Peptide Symposium. LEIPZIG (Allemagne), 04-09/09/2016.

4. Sabri S, Royer M, Pesic A, Mainz A, Süssmuth RD & Cociancich S (2016) **GENOME-BASED SURVEY OF NON-RIBOSOMAL PEPTIDE SYNTHETASE GENES SHARING FASCINATING FEATURES WITH THE ALBICIDIN BIOSYNTHESIS GENES.** AMP2016. 5th International Symposium on Antimicrobial Peptides. MONTPELLIER (France), 06-08/06/2016.

5. Pieretti I, Cociancich S, Carrère S, Morisset A, Rott P & Royer M (2015) **WHOLE-GENOME SEQUENCING OF 18 STRAINS REVEALS HIGH POLYMORPHISM IN THE LIPOPOLYSACCHARIDES LOCUS OF *XANTHOMONAS ALBILINEANS*.** 5th *Xanthomonas* Genomics Conference. BOGOTA (Colombie), 08-11/07/2015.

6. Hempel B-F, Petras D, Mainz A, Cociancich S, Royer M & Süssmuth RD (2015) **THE BIOSYNTHESIS OF ALBICIDIN.** 12th German Peptide Symposium. DARMSTADT (Allemagne), 18-21/03/2015.

7. Petras D, Mainz A, Hempel B, Cociancich S, Royer M & Süssmuth RD (2015) **THE BIOSYNTHESIS OF ALBICIDIN.** 10th Status Seminar on Chemical Biology. FRANKFURT-AM-MAIN (Allemagne), 20-21/01/2015.

8. Vieweg L, Kretz J, Pesic A, Cociancich S, Royer M & Süssmuth RD (2014) **HYDROLYTIC DETOXIFICATION OF ALBICIDIN.** 4. Gemeinsamer Kongress von DGHM (Deutsche Gesellschaft für Hygiene und Mikrobiologie) und VAAM (Vereinigung für Allgemeine und Angewandte Mikrobiologie). DRESDEN (Allemagne), 05-08/10/2014.

9. Cociancich S, Pesic A, Uhlmann S, Petras D, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll J, Pieretti I, Hügelland M, Kemper S, Rott P, Royer M & Süssmuth RD (2014) **TOWARDS STRUCTURE ELUCIDATION OF ALBICIDIN, A POTENT DNA GYRASE INHIBITOR PRODUCED BY THE PLANT PATHOGEN *XANTHOMONAS ALBILINEANS*.** AMP2014. 4th International Symposium on Antimicrobial Peptides. LORIENT (France), 04-06/06/2014.

10. Pieretti I, Marguerettaz M, Bolot S, Carrere S, Cociancich S, Rott P & Royer M (2014) **L'ANALYSE DU POLYMORPHISME DE LOCI CRISPR PERMET DE MIEUX COMPRENDRE L'EVOLUTION DE *XANTHOMONAS ALBILINEANS*.** 11èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 03-07/02/2014.

11. Pieretti I, Bolot S, Carrere S, Lefeuvre P, Cociancich S, Rott P & Royer M (2014) **TOWARDS IDENTIFICATION OF MUTATIONS ASSOCIATED WITH VARIATION IN PATHOGENICITY AMONG NINE GENETICALLY CLOSELY RELATED STRAINS OF *XANTHOMONAS ALBILINEANS* FROM GUADELOUPE.** 11èmes Rencontres Plantes-Bactéries. AUSSOIS (France), 03-07/02/2014.

12. Völler GH, Cociancich S, Gomez C, Hügelland M, Magno C, Noëll J, Süssmuth RD & Royer M (2013) **OVERPRODUCTION AND ISOLATION FROM *XANTHOMONAS ALBILINEANS* OF A NEW NON-RIBOSOMALLY SYNTHESIZED PEPTIDE PUTATIVELY INVOLVED IN PLANT-BACTERIA INTERACTIONS.** FEMS 2013. 5th Congress of European Microbiologists. LEIPZIG (Allemagne), 21-25/07/2013.

13. Pieretti I, Marguerettaz M, Bolot S, Carrere S, Cociancich S, Rott P & Royer M (2013) **THE CRISPR-ASSOCIATED SEQUENCE DIVERSITY PROVIDES NEW INSIGHTS INTO EVOLUTION OF *XANTHOMONAS ALBILINEANS*, THE SUGARCANE LEAF SCALD PATHOGEN.** Biochemical Society focused meeting: CRISPR: evolution, mechanisms and infection. SAINT-ANDREWS (Royaume Uni), 17-19/06/2013.

14. Pieretti I, Marguerettaz M, Bolot S, Carrère S, Cociancich S, Gouzy J, Rott P & Royer M (2012) **CRISPR-ASSOCIATED SEQUENCE DIVERSITY WITHIN *XANTHOMONAS ALBILINEANS*, THE CAUSAL AGENT OF LEAF SCALD DISEASE OF SUGARCANE.** 4th *Xanthomonas* Genomics Conference. ANGERS (France), 09-12/07/2012.

15. Pieretti I, Marguerettaz M, Bolot S, Carrere S, Cociancich S, Gouzy J, Rott P & Royer M (2012) **CRISPR-ASSOCIATED SEQUENCE DIVERSITY WITHIN *XANTHOMONAS ALBILINEANS*, THE CAUSAL AGENT OF LEAF SCALD DISEASE OF SUGARCANE.** 8ème Colloque de la Société Française de Phytopathologie. PARIS (France), 05-08/06/2012.

16. Pieretti I, Cociancich S, Barbe V, Carrere S, Koebnik R, Champoiseau P, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Marguerettaz M, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M, Gabriel DW, Rott P & Royer M (2011) **THE COMPLETE GENOME SEQUENCE OF *XANTHOMONAS ALBILINEANS* PROVIDES INSIGHTS INTO PATHOGENICITY OF THIS SUGARCANE PATHOGEN AND ALLOWS FURTHER ASSESSMENTS OF THE LARGE DIVERSITY WITHIN THIS SPECIES.** ProkaGENOMICS 2011. 5th European Conference on Prokaryotic and Fungal Genomics. GÖTTINGEN (Allemagne), 18-21/09/2011.

17. Pieretti I, Royer M, Cociancich S, Duplan S, Rott P & Gabriel DW (2010) **TOWARDS IDENTIFICATION OF PATHOGENICITY GENES SPECIFIC TO *XANTHOMONAS ALBILINEANS* STRAINS CLOSELY ASSOCIATED WITH SUGARCANE LEAF SCALD DISEASE OUTBREAKS.** 12th International Conference on Plant Pathogenic Bacteria. SAINT-DENIS DE LA RÉUNION (France), 07-11/06/2010.

18. Marguerettaz M, Pieretti I, Gayral P, Puig J, Brin C, Cociancich S, Poussier S, Rott P & Royer M (2010) **EVOLUTIONARY HISTORY AND PUTATIVE ROLES OF THE SPI-1 TYPE III SECRETION SYSTEM OF *XANTHOMONAS ALBILINEANS*.** 12th International Conference on Plant Pathogenic Bacteria. SAINT-DENIS DE LA REUNION (France), 07-11/06/2010.

19. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M & Rott P (2009) **THE COMPLETE GENOME SEQUENCE OF *XANTHOMONAS ALBILINEANS* PROVIDES NEW INSIGHTS INTO THE REDUCTIVE GENOME EVOLUTION OF THE XYLEM-LIMITED *XANTHOMONADACEAE*.** XIV International Congress on Molecular Plant-Microbe Interactions. QUÉBEC CITY (Canada), 19-23/07/2009.

20. Cociancich S, Nachtigall J, Schneider K, Duplan S, Gabriel DW, Süssmuth RD, Rott P & Royer M (2009) **PROGRES RECENTS DANS LA CARACTERISATION STRUCTURALE DE L'ALBICIDINE, UN PUISSANT INHIBITEUR DE L'ADN GYRASE A.** Phytopathologie 2009 - 7ème Colloque National de la Société Française de Phytopathologie. LYON (France), 08-11/06/2009.

21. Marguerettaz M, Darroussat M-J, Muller M, Duplan S, Pieretti I, Girard J-C, Cociancich S, Rott P & Royer M (2009) **PLUSIEURS METABOLITES SECONDAIRES SYNTHETISES PAR DES NRPSS SONT IMPLIQUES DANS LA PATHOGENIE DE *XANTHOMONAS ALBILINEANS*.** Phytopathologie 2009 - 7ème Colloque National de la Société Française de Phytopathologie. LYON (France), 08-11/06/2009.

22. Cociancich S, Schneider K, Duplan S, Pieretti I, Gabriel DW, Süssmuth RD, Rott PC & Royer M (2007) **HETEROLOGOUS PRODUCTION AND CHARACTERIZATION OF ALBICIDIN, A POTENT DNA GYRASE INHIBITOR.** ProkaGENOMICS 2007. 3rd European Conference on Prokaryotic Genomics. GÖTTINGEN (Allemagne), 07-10/10/2007.

23. Champoiseau P, Daugrois J-H, Cociancich S, Pieretti I, Royer M & Rott P (2006) **TOWARDS IDENTIFICATION OF GENES INVOLVED IN PATHOGENICITY OF *XANTHOMONAS ALBILINEANS*, THE PATHOGEN CAUSING SUGARCANE LEAF SCALD.** 11th International Conference on Plant Pathogenic Bacteria. EDINBURGH (Ecosse), 10-14/07/2006.

24. Cociancich S, Vivien E, Pitorre D, Duplan S, Pieretti I, Gabriel DW, Rott P & Royer M (2006) **ENGINEERED TRANSFER OF THE PKS/NRPS BIOSYNTHESIS PATHWAY OF ALBICIDIN: A PROMISING APPROACH TO OVERPRODUCE THIS POTENT ANTIBIOTIC.** 2nd FEMS Congress of European Microbiologists. MADRID (Espagne), 04-08/07/2006.

25. Koenig S, Cociancich S, Shahabuddin M & Fales HM (1999) **THE IDENTIFICATION OF UNKNOWN PEPTIDELIKE STRUCTURES – HOW EASY IS IT?** 47th American Society of Mass Spectrometry Conference on Mass Spectrometry and Allied Topics. DALLAS, TX (USA), 13-17/06/1999.

26. Koenig S, Cociancich S, Shahabuddin M & Fales H (1998) **ISOLATION AND STRUCTURAL CHARACTERIZATION OF A MOSQUITO FACTOR THAT BLOCKS THE DEVELOPMENT OF THE MALARIA PARASITE.** NIH Research Festival '98. BETHESDA, MD (USA), 06-09/10/1998.

27. Cociancich S & Shahabuddin M (1998) **SPATIAL DISTRIBUTION OF VESICULAR ATPASE OVER-EXPRESSING CELLS IN THE MIDGUT OF MALARIA VECTORS.** Third International Symposium on Molecular Insect Science. SNOWBIRD, UT (USA), 05-10/06/1998.

28. Cociancich S & Shahabuddin M (1998) **A MOSQUITO FACTOR BLOCKS THE ZYGOTE TO OOKINETE TRANSFORMATION OF *PLASMODIUM GALLINACEUM*.** Third International Symposium on Molecular Insect Science. SNOWBIRD, UT (USA), 05-10/06/1998.

29. Troyer JM, Cociancich S, Noden B, Radulovic S & Azad AF (1998) **IMMUNE PEPTIDE PRODUCTION IN FLEAS IN RESPONSE TO BACTERIAL CHALLENGE.** Keystone Symposia on Molecular and Cellular Biology « Toward the Genetic Manipulation of Insects ». TAOS, NM (USA), 09-15/01/1998.

30. Cociancich S & Shahabuddin M (1997) **SOLUBLE FACTORS IN MOSQUITOES BLOCK ZYGOTE TO OOKINETE TRANSFORMATION OF *PLASMODIUM GALLINACEUM*.** Molecular Parasitology Meeting VIII. WOODS HOLE, MA (USA), 24-28/09/1997.

31. Troyer JM, Noden B, Radulovic S, Azad AF & Cociancich S (1997) **CHARACTERIZATION OF AN ANTI-BACTERIAL PEPTIDE PRODUCED IN FLEAS IN RESPONSE TO BACTERIAL CHALLENGE.** 13th Sesquiannual Meeting of the American Society for Rickettsiology. CHAMPION, PA (USA), 21-24/09/1997.

32. Letellier L, Cociancich S, Hetru C and Hoffmann JA (1993) **INSECT DEFENSIN, AN INDUCIBLE ANTIBACTERIAL PEPTIDE, FORMS VOLTAGE-DEPENDENT CHANNELS IN *MICROCOCCUS LUTEUS*.** 2nd International Workshop on Pore-forming Toxins. MAINZ (Allemagne), 29/09-02/10/1993.

33. Bulet P, Ohresser S, Cociancich S, Hetru C, Dimarcq J-L, Levashina E, Reichhart J-M & Hoffmann J (1993) **ISOLATION AND CHARACTERIZATION OF INDUCIBLE ANTIBACTERIAL PEPTIDES OF *DROSOPHILA*.** 2nd International Symposium on Molecular Insect Science. FLAGSTAFF, AZ (USA), 17-22/07/1993.

Chapitres d'ouvrages scientifiques

1. Royer M, Pieretti I, Cociancich S & Rott P (2018) **RECENT PROGRESS IN UNDERSTANDING THREE MAJOR BACTERIAL DISEASES OF SUGARCANE: GUMMING, LEAF SCALD AND RATOON STUNTING.** Achieving sustainable cultivation of sugarcane. Volume 2: breeding, pests and diseases. (ed. Philippe Rott) pp311-336, Burleigh Dodds Science Publishing, Cambridge, UK.
   Web of Science citations: 0

2. Hetru C, Bulet P, Cociancich S, Dimarcq J-L, Hoffmann D & Hoffmann JA (1994) **ANTIBACTERIAL PEPTIDES/POLYPEPTIDES IN THE INSECT HOST DEFENSE. A COMPARISON WITH VERTEBRATE ANTIBACTERIAL PEPTIDES/POLYPEPTIDES.** *Phylogenetic Perspectives in Immunity* (Eds Hoffmann JA, Janeway CA and Natori S) pp 34-65, RG Landes Company, Austin, TX, USA.
   Web of Science citations: 36

Publications de vulgarisation

1. Cociancich S, Royer M & Rott P (2017) **UN ANTIBIOTIQUE DECOUVERT DANS LA CANNE A SUCRE.** *La Recherche*, n°524 (juin 2017), pp 62-65.

Brevets

1. Süssmuth R, Kretz J, Schubert V, Pesic A, Hügelland M, Royer M, Cociancich S, Rott P, Kerwat D & Grätz S (2014) **ALBICIDIN DERIVATIVES, THEIR USE AND SYNTHESIS.** PCT WO 2014/125075 A1.

Rapports

1. Miller L, Shahabuddin M, Wang P, Docherty P, Pimenta P, Zieler H & Cociancich S (1999) **MALARIA: IDENTIFICATION OF MOSQUITO MIDGUT RECEPTORS FOR *PLASMODIUM* OOKINETES.** UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, Project N° 950476. TDR Final Report Series n° 15.

Diplômes

1. Cociancich S (1994) **ISOLEMENT ET CARACTERISATION DE PEPTIDES ANTIBACTERIENS INDUCTIBLES CHEZ LES INSECTES.** Thèse de l'Université Louis Pasteur, Strasbourg I (145 pages).

2. Cociancich S (1991) **PURIFICATION CHEZ UN INSECTE COLEOPTERE D'UN NOUVEAU PEPTIDE ANTIBACTERIEN ET DE DEUX NOUVEAUX MEMBRES DE LA FAMILLE DES DEFENSINES D'INSECTES.** Diplôme d'Etudes Approfondies (Biologie Moléculaire et Cellulaire, option Biologie du Développement), Université Louis Pasteur, Strasbourg I (18 pages).

Publications internes

1. Vivien E, Cociancich S, Duplan S, Pieretti I, Rott P & Royer M (2004) **VOIES DE BIOSYNTHESE DE L'ALBICIDINE.** *Cirad, Rapport annuel Programme canne à sucre.* pp 13-14.

Seminaires & Réseaux – Communications orales

1. Cociancich S (2016) **WHAT CAN WE LEARN FROM THE COMPARISON OF THE BIOSYNTHETIC GENES OF ALBICIDIN IN TWO PHYLOGENETICALLY DISTANT BACTERIA : *XANTHOMONAS ALBILINEANS* AND *CYSTOBACTER* SP. ?** Réunion du réseau FNX (French Network on Xanthomonads). ANGERS (France), 13-15/09/2016.

2. Cociancich S (2016) **QUE NOUS APPREND LA COMPARAISON DES GENES DE BIOSYNTHESE DE L'ALBICIDINE DE DEUX BACTERIES PHYLOGENETIQUEMENT TRES ELOIGNEES : *XANTHOMONAS ALBILINEANS* ET *CYSTOBACTER* SP. ?** Journée scientifique de l'UMR BGPI. MONTPELLIER (France), 24/06/2016.

3. Cociancich S (2016) **UNE APPROCHE MULTIDISCIPLINAIRE A PERMIS LA CARACTERISATION DE LA STRUCTURE ET DES VOIES DE BIOSYNTHESE DE L'ALBICIDINE, UN PUISSANT INHIBITEUR DE L'ADN GYRASE SYNTHETISE PAR LA VOIE NON-RIBOSOMIQUE.** Séminaire interne de l'UMR Microbiologie, Adaptation et Pathogénie (MAP). VILLEURBANNE (France), 11/03/2016.

4. Cociancich S (2014) **WHAT MAKES THE ALBICIDIN BIOSYNTHESIS PATHWAY UNIQUE AMONGST ALL KNOWN NON-RIBOSOMALLY SYNTHESIZED PEPTIDE ASSEMBLY LINES?** Réunion du réseau FNX (French Network on Xanthomonads). BRIANDES (France), 06-08/10/2014.

5. Cociancich S (2013) **LES NRPS CHEZ LES *XANTHOMONAS*: ALBICIDINE ET AUTRES PETITES MOLECULES PROMETTEUSES.** Séminaire interne de l'UMR Diversité, Génomes & Interactions Microorganismes-Insectes (DGIMI). MONTPELLIER (France), 14/10/2013.

6. Cociancich S (2013) **NRPS IN *XANTHOMONAS*: THE SUCCESSFUL STORY OF ALBICIDIN DESERVES TO BE PURSUED FOR OTHER PROMISING SMALL MOLECULES.** Réunion du réseau FNX (French Network on Xanthomonads). MOULIS (France), 02-04/10/2013.

7. Cociancich S (2011) **GENOME MINING INDICATES THAT THE GENUS *XANTHOMONAS* IS A PROMISING RESERVOIR FOR NEW BIOACTIVE NON-RIBOSOMALLY SYNTHETIZED PEPTIDES.** Réunion du réseau FNX (French Network on Xanthomonads). MONTPELLIER (France), 14-15/11/2011.

8. Cociancich S (2011) **DECIPHERING SMALL MOLECULES AS NEW VIRULENCE FACTORS IN THE BACTERIAL SUGARCANE PATHOGEN *XANTHOMONAS ALBILINEANS*.** AK Süssmuth internal seminars, Technische Universität Berlin. BERLIN (Allemagne), 08/09/2011.

9. Cociancich S (2008) **PURIFICATION AND CHARACTERIZATION OF ALBICIDIN, A BACTERIAL TOXIN PRODUCED BY *XANTHOMONAS ALBILINEANS*, THE CAUSAL AGENT OF SUGARCANE LEAF SCALD DISEASE.** AK Süssmuth internal seminars, Technische Universität Berlin. BERLIN (Allemagne), 08/04/2008.

10. Cociancich S (1999) **RECHERCHE ET ETUDE CHEZ LE CHARANÇON DE LA (DES) CIBLE(S) MOLECULAIRE(S) DE LA PROTEINE ENTOMOTOXIQUE DU POIS.** Séminaire interne du Laboratoire de Biologie Appliquée INRA/INSA. VILLEURBANNE (France), 11/03/1999.

11. Cociancich S (1998) **AN UPDATE ON THE CHARACTERIZATION OF AE-TBF1, A FACTOR IN *AEDES AEGYPTI* THAT BLOCKS THE TRANSFORMATION FROM ZYGOTE TO OOKINETE OF *PLASMODIUM GALLINACEUM* OR I KNOW WHAT IT ISN'T, BUT I DON'T KNOW WHAT IT IS (!).** The Laboratory of Malaria Research internal seminar, NIAID, NIH. BETHESDA, MD (USA), 25/11/1998.

12. Cociancich S (1998) **A MOSQUITO FACTOR BLOCKS *PLASMODIUM*'S ZYGOTE TO OOKINETE TRANSFORMATION.** Vector Biology Forum, Johns Hopkins University. BALTIMORE, MD (USA), 30/10/1998.

13. Cociancich S (1998) **V-ATPASE OVEREXPRESSING CELLS MAY DETERMIN THE PATTERN OF *PLASMODIUM GALLINACEUM* OOCYSTS DISTRIBUTION ON THE MIDGUTS OF MALARIA VECTORS.** The Laboratory of Parasitic Diseases internal seminar, NIAID, NIH. BETHESDA, MD (USA), 06/10/1998.

14. Cociancich S (1998) **A MOSQUITO FACTOR BLOCKS *PLASMODIUM*'S ZYGOTE TO OOKINETE TRANSFORMATION.** The Laboratory of Parasitic Diseases internal seminar, NIAID, NIH. BETHESDA, MD (USA), 17/03/1998.

15. Cociancich S (1998) **CHARACTERIZATION IN MOSQUITOES OF THE V-ATPASE B-SUBUNIT.** The Laboratory of Parasitic Diseases internal seminar, NIAID, NIH. BETHESDA, MD (USA), 30/01/1998.

16. Cociancich S (1997) **CHARACTERIZATION OF MOSQUITO FACTORS THAT INFLUENCE *PLASMODIUM* DEVELOPMENT.** The Laboratory of Malaria Research internal seminar, NIAID, NIH. BETHESDA, MD (USA), 05/11/1997.

17. Cociancich S (1997) **MOLECULAR CLONING OF THE B-SUBUNIT OF *AEDES AEGYPTI* V-ATPASE.** The Laboratory of Parasitic Diseases internal seminar, NIAID, NIH. BETHESDA, MD (USA), 25/07/1997.

18. Cociancich S (1997) **CHARACTERIZATION OF MOSQUITO FACTORS THAT INFLUENCE *PLASMODIUM* DEVELOPMENT.** The Laboratory of Malaria Research internal seminar, NIAID, NIH. BETHESDA, MD (USA), 17/01/1997.

19. Cociancich S (1996) **COMPARATIVE STUDY OF SUSCEPTIBLE AND REFRACTORY STRAINS OF MOSQUITOES USING BIOCHEMICAL TECHNIQUES. CHARACTERIZATION OF ONE/SEVERAL MOLECULES INVOLVED IN THE REFRACTORY MECHANISMS OBSERVED IN SOME MOSQUITO STRAINS.** The Laboratory of Parasitic Diseases internal seminar, NIAID, NIH. BETHESDA, MD (USA), 06/09/1996.

20. Cociancich S (1994) **ISOLEMENT ET CARACTERISATION DE PEPTIDES ANTIBACTERIENS INDUCTIBLES CHEZ LES INSECTES.** Séminaires de l'Unité de Microbiologie du CRSSA "Emile Pardé". LA TRONCHE (France), 25/11/1994.

21. Hetru C & Cociancich S (1993) **LES PEPTIDES ANTIBACTERIENS INDUCTIBLES CHEZ LES INSECTES.** Journées Scientifiques CRSSA-DRET « Neurotoxines animales et canaux ioniques ». LA TRONCHE (France), 07/12/1993.

22. Cociancich S (1993) **ETUDE DU MODE D'ACTION D'UNE DEFENSINE D'INSECTE.** Séminaire interne du groupe de Franc Pattus, EMBL. HEIDELBERG (Allemagne), 21/06/1993.

23. Cociancich S (1990) **ISOLEMENT ET CARACTERISATION DE PEPTIDES NEUROHORMONAUX ET IMMUNITAIRES.** Séminaire interne du Laboratoire de Biologie Générale « Endocrinologie et Immunologie des insectes ». STRASBOURG (France), 12/12/1990.

Séminaires & Réseaux - Affiches

1. Marguerettaz M, Darroussat M-J, Müller M, Duplan S, Pieretti I, Girard J-C, Cociancich S, Rott P & Royer M (2010) **PLUSIEURS METABOLITES SECONDAIRES SYNTHETISES PAR DES NRPSS SONT IMPLIQUES DANS LA PATHOGENIE DE *XANTHOMONAS ALBILINEANS*.** 3ème Printemps de Baillarguet. MONTFERRIER-SUR-LEZ (France), 29/04/2013.

Récapitulatif des publications

Journal name	5-year mean IF (data from 2017)	Number of publications	First/last authorship
Genome Announcements	-	1	-
International Sugar Journal	0,113	1	-
FEMS Microbiology Letters	2,096	1	-
Biochemical and Biophysical Research Communications	2,455	1	1
Journal of Insect Physiology	2,863	1	-
Phytopathology	3,200	1	-
Genes	3,286	1	-
Biochemical Journal	3,850	1	1
Molecular Plant-Microbe Interactions	4,078	1	-
European Journal of Biochemistry (FEBS Journal)	4,237	1	-
Journal of Biological Chemistry	4,253	3	2
BMC Genomics	4,257	4	2
Applied and Environmental Microbiology	4,272	1	-
Frontiers in Plant Science	4,354	1	1
Antimicrobial Agents and Chemotherapy	4,504	1	-
ACS Chemical Biology	4,794	1	-
Chemistry - a European Journal	4,950	1	-
Parasitology Today (Trends in Parasitology	7,178	2	1
Annual Review of Phytopathology	11,484	1	-
Angewandte Chemie International Edition	11,954	1	-
Journal of the American Chemical Society	13,613	1	-
Nature Chemical Biology	13,990	1	1
Total		28	9

INTRODUCTION

For any researcher, the « Habilitation à Diriger des Recherches (HDR) » is the sesame allowing to officially supervise PhD students. Through the redaction of a memoir and a defense, the candidate has to demonstrate scientific qualities, students' supervision skills, as well as the ability to implement research projects. It is said that the redaction of the HDR memoir is a unique occasion to withdraw for a short time from the never-ending whirl of the laboratory life, when experiments, publications and projects writings, participations to conferences, lab-meetings and student supervision race in a hectic pace. Indeed, the time has come for me to take stock of my career and imagine its future with respect to the past and the present.

During the 1990-1999 timeframe, which mostly corresponds to my master 2, PhD and post-docs, I participated to several projects aimed at studying host/pathogen interactions (e.g. the insect immunity in response to bacterial infections, the development of the malaria parasite in mosquitoes, or the insecticidal toxicity of a pea protein). For most of these projects, at a time when molecular biology was the main technique implemented in almost every research laboratory, I developed biochemical methods which efficiently complemented the molecular and genetic studies, and allowed me to isolate the molecules of interest and to characterize fully or partially their structures. An additional training in molecular biology and in electrophysiology completed my technical background.

In 2003, I joined at UMR BGPI the team led by Philippe Rott, whose interest was to study the pathogenicity of the bacterium *Xanthomonas albilineans*, the causal agent of the leaf scald disease of sugarcane. A striking feature of *X. albilineans* is the production of albicidin, a phytotoxin which exhibits a potent antibiotic activity and whose structure was still elusive almost 20 years after its discovery. Interestingly, this secondary metabolite is synthesized by the non-ribosomal peptide synthesis pathway. A feature of bacteria and fungi, the non-ribosomal peptide synthesis and the polyketide synthesis correspond to the (ribosome-independent) enzymatic biosynthesis of secondary metabolites, which are natural products encompassing a broad range of biologically-relevant activities, and often displaying complex structures. Non-ribosomal pathways involve large genomic sequences which contain genes encoding multi-modular mega-enzymes responsible for the biosynthesis of the peptidic and/or polyketidic backbone of the metabolites. These biosynthetic genes are often clustered with other genes involved in either modification, transport, regulation or resistance to the final product, for instance, thus forming biosynthetic genes clusters (BGCs). As such, the *in silico* analysis of the genomic organization and sequence of these BGCs, and in particular of the biosynthetic enzymes encoding genes, provides many data which are useful to predict biosynthetic pathways and structural features of the corresponding secondary metabolites. For a full structural characterization of complex molecules such as secondary metabolites, competences in *in silico* analysis are therefore a required complement to the knowledge in molecular biology and biochemistry necessary to drive their production and isolation.

Over a decade, I extensively studied albicidin. This work, combining predictive data from *in silico* analyses of the BGC with experimental data from molecular biology and biochemistry experiments leading to decipher the complex structure of albicidin, has been a proof of concept for the importance of predictive data obtained through *in silico* analysis of this BGC. With this work, I gained expertise on non-ribosomal biosynthetic pathways and mechanisms, and I currently develop another project based on a similar approach for the study of an unprecedented non-ribosomally-synthesized quorum sensing signal involved in the phytopathogenicity of *Dickeya* bacteria, which are involved in many crops diseases.

As it has been the case for albicidin, the classical approach used almost exclusively up to the end of the 20th century to discover new secondary metabolites was based on screenings for specific biological activities.

However, we can now use the knowledge generated by the *in silico* analyses of many BGCs to screen the actual exponentially-raising number of available microbial genome sequences for the presence of BGCs (or for specific features of BGCs). This "genome mining" approach allows notably to reveal the real secondary metabolic potential of microorganisms which appeared to be completely under-estimated by the biological activity-based classical approach since these BGCs may be either silent or non-expressed in usual laboratory conditions. The subsequent *in silico* analysis of these newly-discovered BGCs supposedly gives hints towards the biosynthetic pathway and the structure of the encoded compounds prior their eventual expression, isolation and characterization. Genome mining should be considered as a required predictive approach which complements molecular studies to fully decipher the structure of new natural products and to unravel their complex biosynthesis pathways.

I currently take part to and develop several projects based on the "genome mining" approach aiming at screening genomes of bacteria of the genus *Xanthomonas* (projects "Meta-B" and "XanthoNRPS") and *Streptomyces* (project "CONTACT") for discovering new secondary metabolites exhibiting specific features from the non-ribosomal peptide synthesis pathway. These projects are based on multi-disciplinary approaches, *i.e.* involving genome mining and subsequent *in silico* analysis prior to characterization of the natural products by experimental techniques. They are expected to generate fundamental knowledge as well as knowledge potentially providing applications in development, for instance, of new antibiotics or other pharmaceutically- or agrochemically-relevant compounds.

To date, I co-supervised several master and PhD students, and recently co-directed a PhD on the non-ribosomal peptide synthesis in the genus *Xanthomonas*.

This document contains a synthetic report of my past research achievements as well as insights into my current and future scientific projects. First, I shortly report my research activities and main results obtained during my master2/PhD in Strasbourg and my post-doctoral fellowship at the National Institutes of Health in Bethesda (USA). The second part describes my past work and present projects performed at UMR BGPI, prefiguring future projects that I plan to implement in the team in the coming years. When appropriate, I will outline the valorization of my research and my involvement in the supervision of students.

PART 1
MASTER/PHD AND POST-DOCTORAL FELLOWSHIP (1990/1999)

STUDY OF ANTIBACTERIAL PEPTIDES IN INSECTS

STUDY OF *PLASMODIUM*'S DEVELOPMENT IN MOSQUITOES

"Time passes in moments. Moments which, rushing past, define the path of a life [...]. How rarely do we stop to examine that path, to see the reasons why all things happen, to consider whether the path we take in life is our own making or simply one into which we drift with eyes closed? But what if we could stop, pause to take stock of each precious moment before it passes? Might we then see the endless forks in the road that have shaped a life, and seeing those choices, choose another path?" **Dana Scully, The X-Files, 7X17**

I chose to limit this chapter to the works done during my Master 2 and PhD in Strasbourg (1990/1994) on insect immunity and during my post-doc at NIH (1995/1998) on *Plasmodium* / mosquito interactions, as they had a "founder effect" on the rest of my career. This chapter is meant to give the reader an overview of my training.

I. Isolation and characterization of inducible antibacterial peptides in insects

Master2 and PhD at UPR RIDI "Réponse Immunitaire et Développement chez les Insectes"
Strasbourg (1990-1994) - Dir: Jules Hoffmann

After I graduated in biology and biochemistry at University of Strasbourg, my scientific career started as a master student at the "Laboratoire de Biologie Générale" directed by Jules Hoffmann in Strasbourg. I was initially expected to study the insect molt hormone ecdysone in locusts as most of the laboratory efforts were devoted to investigate insect endocrinology. However, preliminary investigations into insect immunity were already on-going since insects have been observed to be highly resistant to infections, as shown by their strong capacity to survive to "chirurgical" operations (such as brain transfers for the ecdysone studies) which were done in non-sterile environments. Promising data therefore prompted my re-assignation to the new challenging "innate immunity" topic.

I.1. Insect innate immunity: state of the art in 1990

Insect immunity has been studied since the end of the 19th century, when insects were shown to rely on phagocytosis to fight bacterial infections (Cuénot, 1896). Besides these cellular reactions, another set of data later brought to light inducible humoral mechanisms based on the appearance of an antibacterial activity in the hemolymph following the injection of heat-killed bacteria (Glaser, 1918; Paillot, 1920, Metalnikow, 1920). The following decades offered no real progress in the identification of these factors. A first antibacterial compound was characterized only 60 years later in the hemolymph of bacteria-challenged pupae from the Lepidoptera *Hyalophora cecropia*, and described as a 4-kDa anti-Gram-negative cationic peptide named cecropin (Steiner *et al.,* 1981). Further studies allowed the discovery of additional inducible peptides isolated from the immune hemolymph of insects belonging to the orders Lepidoptera, Diptera and Hymenoptera (Neoptera Oligoneoptera, most recent insects in evolution). These antibacterial peptides could be sorted in four families:

Cecropins were isolated in several species from Lepidoptera and Diptera (reviewed in Boman *et al.,* 1991). These 4-kDa peptides display an anti-Gram-negative activity. A nuclear magnetic resonance (NMR) spectroscopy study revealed the 3D-structure of cecropin A from *H. cecropia* as a N-terminal amphipatic α-helix linked by a three-aminoacid (aa) bend to a hydrophobic α-helix (Holak *et al.,* 1988).

Insect defensins were isolated in Diptera. They are 4-kDa anti-Gram-positive peptides containing six cysteines engaged in three disulfide bridges (Matsuyama & Natori, 1988; Lambert *et al.,* 1989). The 3D-structure of insect defensins consists of a N-terminal flexible loop followed by an α-helix and an antiparallel β-sheet (Bonmatin *et al.,* 1992). The array of cysteines and the connectivity of the disulfide bridges result in a 3D-conformation called CS$\alpha\beta$ motif (Cysteine-stabilized α-helix β-sheet motif) in which the α-helix is linked to the β-sheet by two disulfide bridges (figure 1). Interestingly, this structural motif is also found in some scorpion toxins.

Large glycine-rich peptides are anti-Gram-negative peptides found in Lepidoptera (attacines, 22 kDa) (Hultmark *et al.,* 1983) and in Diptera (diptericines, 9 kDa). Diptericines exhibit a characteristic short N-terminal proline-rich domain separated from a large glycine-rich domain by a stretch of five consecutive glycine residues (Dimarcq *et al.,* 1988).

Small proline-rich peptides are 2- to 4-kDa anti-Gram-negative peptides isolated from the honeybee *Apis mellifera* (Hymenoptera). They contain up to 1/3 of proline residues (Casteels *et al.,* 1989).

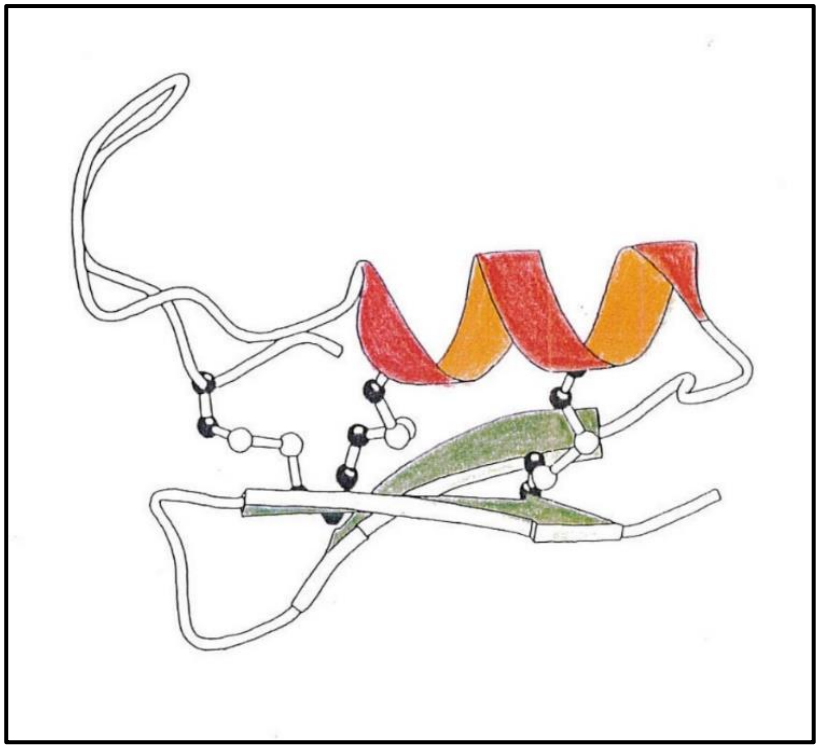


Figure 1. Schematic representation of the 3D structure of insect defensin A from *Phormia terranovae* (Diptera). The α-helix is represented in red and orange whereas the β-sheet is represented in green (adapted from Bonmatin *et al.,* 1992; taken from Cociancich, 1994a).

I.2. Description of new inducible antibacterial peptides from insects and other Arthropods

As I started my master2, we investigated whether coleopteran insects (chosen because they represent over one million species) were able to produce antibacterial peptides in response to a bacterial challenge, and if yes, which ones. The injection of heat-killed bacteria to larvae of the tropical species *Zophobas atratus* led to the appearance of an antibacterial activity in the hemolymph within hours post-immunization. We developed a protocol of purification based on biochemical techniques to isolate and characterize the molecule(s) responsible for this antibacterial activity. Briefly, the challenged cell-free hemolymph was fractionated by solid-phase extraction. The fractions were monitored for their antibacterial activity against either Gram-negative or Gram-positive bacteria by a biological assay. Positive fractions were submitted to chromatography by HPLC using various chromatographical techniques (reversed-phase, size exclusion) to isolate the antibacterial compounds. Once purified, each antibacterial peptide was sequenced by Edman degradation and its molecular mass was determined by mass spectrometry (MS) measurement (collaboration: Alain Van Dorsselaer, Strasbourg). Finally, the activity spectrum of the antibacterial peptides was determined by assessing their capacity to inhibit the growth of a panel of Gram-positive and Gram-negative bacteria. I purified three novel inducible antibacterial peptides, two of them being new members of the anti-Gram-positive insect defensins family (67% sequence similarity with insect defensin A from *P. terranovae*, including the six cysteines involved in three disulfide bridges). The third antibacterial peptide, coleoptericin, is a large (8 kDa) glycine-rich (18%) anti-Gram-negative peptide showing no homology with other known antibacterial peptides (Bulet *et al.,* 1991).

These results were the proof of concept that the workflow we established was efficient enough to investigate further insect species. At this stage, only holometabolous insects, which are the most recent in evolution, had been studied for their capacity to produce antibacterial peptides in response to a bacterial infection. These studies were therefore widened during my PhD with a focus on species representative of more ancient insect orders. Indeed, I could isolate and characterize almost two dozen new inducible antibacterial peptides from about ten insect species representative of the insect class, including a species belonging to the ancient order Odonata. All the newly-discovered antibacterial peptides in frame of my PhD belong to one of the four above-mentioned families. Below is a summary of my main results and key findings.

In addition to the two coleopteran defensins already described, I isolated seven other insect defensins in the immune hemolymph from species belonging to Diptera, Trichoptera, Hemiptera and Odonata. They share common features such as the size (approx. 4 kDa), the biological activity (anti-Gram-positive) and the six cysteines engaged in three disulfide bridges compatible with the CS$\alpha\beta$ motif (Bulet *et al.,* 1992; Cociancich, 1994a; Cociancich *et al.,* 1994b). They exhibit between 58 and 95% similarity when compared between orders, with the remarkable exception of the *Aeschna cyanea* defensin (Odonata, very ancient insect order which appeared 490 million years ago) which is only 37% similar to the dipteran defensins. Interestingly, I also isolated in the hemolymph of another arthropod, the scorpion *Leiurus quinquestriatus*, a defensin exhibiting over 71% similarity with the *A. cyanea* defensin (Cociancich *et al.,* 1993a). Amazingly, these two species, which are among the most primitive arthropods leaving on Earth, were capable to fight bacterial infection through the release of very similar antibacterial peptides.

I also isolated and characterized short proline-rich anti-Gram-negative peptides in several hemipteran species, namely pyrrhocoricin in *Pyrrhocoris apterus*, and metalnikowins in *Podisus maculiventris* and *Palomena prasina*. Pyrrhocoricin is a 20-aa peptide containing 25% of proline residues mostly arranged with arginine residues in proline-arginine-proline triplets. Pyrrhocoricin shows a high degree of similarity with drosocin, a proline-rich antibacterial peptide isolated previously in *Drosophila melanogaster* at the laboratory (Bulet *et al.,* 1993). Similarly to drosocin, pyrrhocoricin is decorated with an N-acetyl-galactosamine residue linked through an O-glycosylation to the threonine residue in position 11 (Cociancich *et al.,* 1994b). Indeed, this threonine is present within a consensus sequence for O-glycosylation of hydroxylated aas (*i.e.* a proline at positions -1 and/or +3 with respect to a threonine or a serine increases the probability of these latter residues to be O-glycosylated). I showed that the glycosylation of pyrrhocoricin is required for its biological activity. Because the acid conditions used during the purification of the molecule are prone to weaken glycosidic bonds, it is possible that pyrrhocoricin is substituted by a more complex oligosaccharide, as is the case for drosocin in which a di-saccharidic substitution has been evidenced. Metalnikowins represent seven isoforms of 15- to 17-aa molecules exhibiting between 53 and 67% of sequence similarity with pyrrhocoricin. They contain up to 33% of proline residues, and the typical proline-arginine-proline arrangement is observed. However, since no threonine or serine have been found in their sequence, none of the metalnikowins is O-glycosylated (Cociancich, 1994a).

In addition to coleoptericin, I purified three other large (9 to 17 kDa) glycine-rich anti-Gram-negative peptides, namely hemiptericin in *P. apterus* (Hemiptera) (Cociancich *et al.,* 1994b), trichoptericin in *Limnephilus stigma* (Trichoptera), and a new member of the diptericin family in *Calliphora vicina* (Diptera) (Cociancich, 1994a). Sequence-wise, despite their high content in glycine, all these molecules have in common to have nothing in common (!).

In conclusion, I showed during my master and PhD that many insect species belonging to several orders (including very ancient ones) rely on the production in their hemolymph of a battery of broad-spectrum antibacterial peptides which are induced in response to a bacterial infection (Cociancich, 1994a; reviewed in Cociancich *et al.,* 1994c). Among the antibacterial peptides that I isolated and characterized, insect defensins are the more widespread (found in six orders of insects, including an ancient one, and one order of arachnids). Hemipteran and dipteran insects also produce short proline-rich antibacterial peptides, eventually O-glycosylated. Additionally, I described large glycine-rich peptides and cecropins in some recent orders of insects. Eventually, I also investigated some insect species in which I could not find any antibacterial peptide but it was impossible to conclude whether this was the result of technical limitation due to detection thresholds being too high or because there were no antibacterial peptides at all. This was the case, for instance, of the locust *Locusta migratoria* (Orthoptera). We observed however that this species developed a stronger than usual cellular immune response which we hypothesized as being sufficient to fight the bacterial infections.

I.3. Deciphering the mode of action of an insect defensin

I previously observed that insect defensins kill Gram-positive bacteria within few minutes and hence suspected an action at the level of the plasma membrane. In collaboration with Lucienne Letellier at Orsay, I studied the effect of a recombinant insect defensin on the Gram-positive bacterium *Micrococcus luteus*. We showed that the insect defensin triggers within seconds a rapid potassium efflux from the bacteria (10^5 to 10^6 molecules of insect defensin per bacterium). Bacteria are completely depleted in potassium in less than two minutes (figure 2A). This membranolytic activity is abolished if the 3D-structure of the insect defensin is impaired or if the fluidity of the membrane is decreased by the action of divalent cations (reversible phenomenon with the addition of chelating agents) (figure 2B). Additional experiments showed that the activity of the insect defensin lowers the bacterial transmembrane potential to half of its normal value (200 mV), indicating however that the plasma membrane keeps some level of integrity (figures 2C and 2D) (Cociancich *et al.,* 1993b).

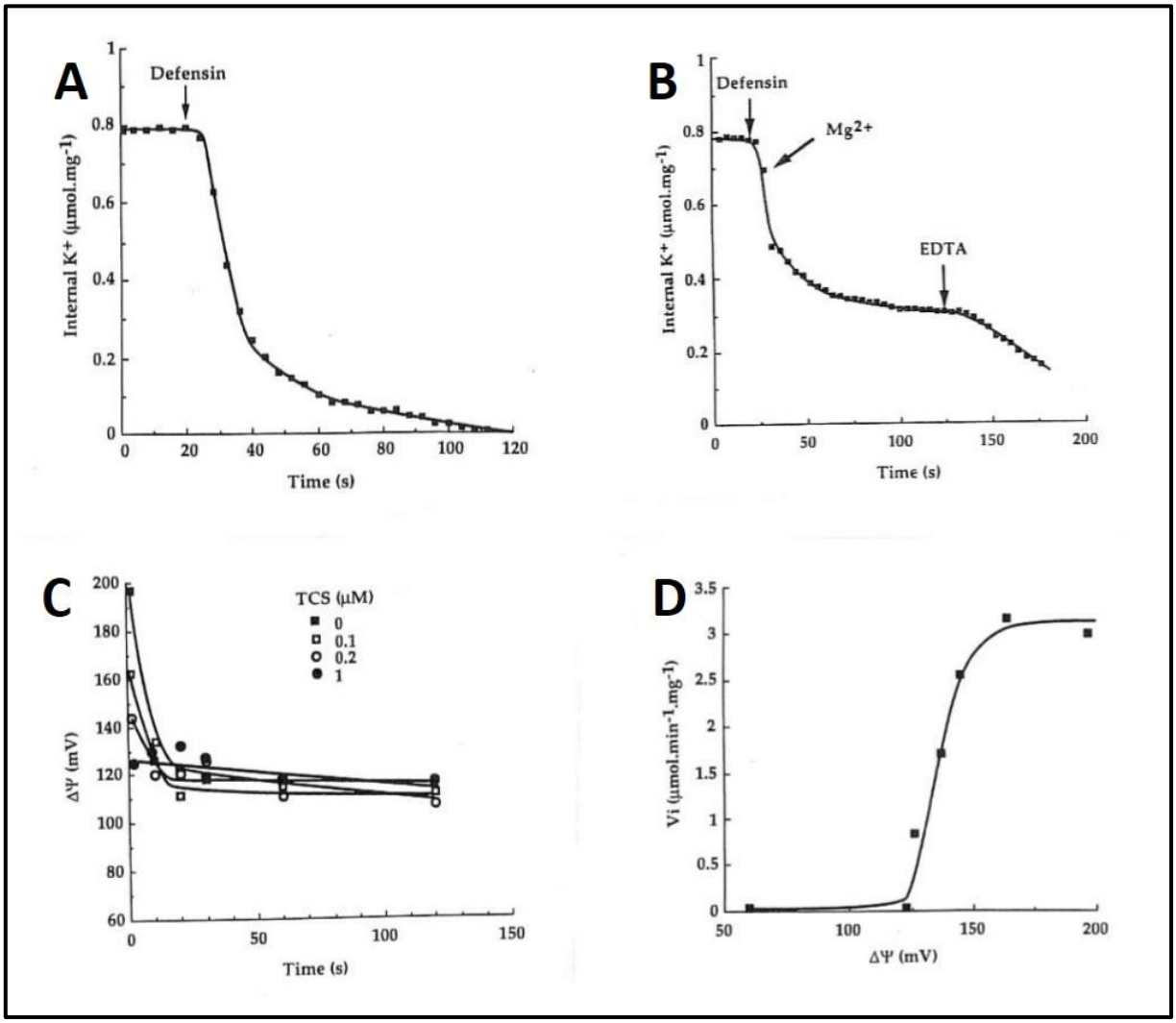


Figure 2. Physiological effects of recombinant insect defensin A from *P. terranovae* on the Gram-positive bacterium *M. luteus*. A: Potassium efflux induced by insect defensin. Bacteria are depleted within two minutes. B: Inhibitory effect of divalent cations on the insect defensin-induced potassium efflux, and reversibility of the phenomenon by addition of a chelating agent. C: Changes in the bacterial transmembrane potential provoked by the insect defensin. The transmembrane potential decreases to approx. 100 mV, indicating that the membrane keeps some integrity. D: Effect of the transmembrane potential value on the insect defensin-induced potassium efflux. A minimal transmembrane potential of approx. 100 mV is required for insect defensin to trigger the potassium efflux. For detailed experimental conditions, see Cociancich *et al.,* 1993b.

All together, these and other results are in favor of the formation by insect defensins of voltage-gated channels in the cytoplasmic membrane of *M. luteus*. This hypothesis has been ascertained by a complementary patch-clamp study in which insect defensin induced, as long as a transmembrane potential value of at least 100 mV is applied, opening/closing events of channels (with variations in duration of opening and in size) in giant liposomes (figure 3) (Cociancich *et al.,* 1993b).

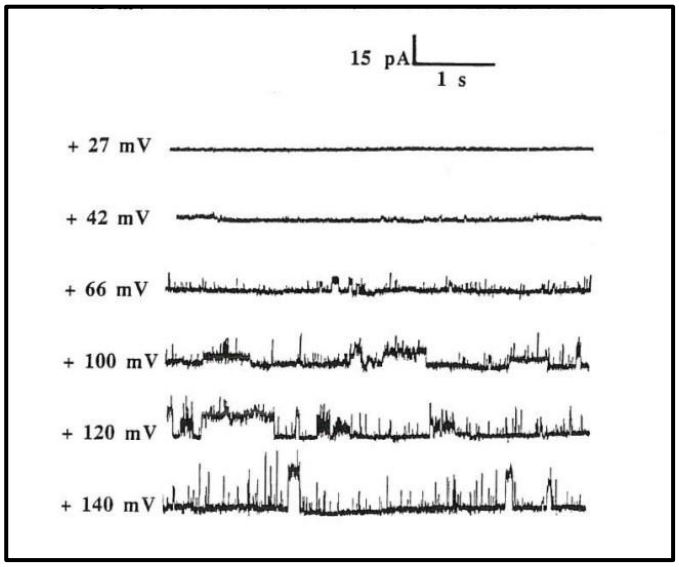


Figure 3. Single voltage-gated channels currents induced by insect defensin A from *P. terranovae*. In this patch-clamp experiment with giant liposomes in which insect defensin has been incorporated, one can easily observe heterogeneity in conductance sizes and in openings and closings durations. For detailed experimental conditions, see Cociancich *et al.,* 1993b.

And since then, what's new?

Almost 25 years later, literally thousands of papers have been published on antibacterial peptides (or rather "antimicrobial" peptides [AMPs] as they are now referred to as since antiviral, antifungal and antiparasitic activities have also been discovered) from various animal (including mammals) and plant sources. Regarding insects, a few hundreds of AMPs have now been discovered. Insect defensins (almost 50 described to-date) and proline-rich AMPs (over two dozen described, half of them being O-glycosylated) are still the larger families. Recently, the mode of action of a proline-rich AMP has been deciphered as being very different from the mode of action of insect defensins: indeed, pyrrhocoricin has been shown to pass through the bacterial membrane and to inhibit the bacterial growth by binding to the ribosome and blocking the protein synthesis. More precisely, pyrrhocoricin binds to the ribosome exit tunnel with a reverse orientation to a nascent polypeptide chain (reviewed in Graf *et al.,* 2017). And finally, AMPs have recently been identified in Orthoptera by a transcriptomic approach showing that transcripts for insect defensins are present in various tissues in locusts upon infection (Lv *et al.,* 2016).

Interstingly, during my PhD, all the teams of the Hoffmann lab switched to the "insect innate immunity" theme. Our team laid grounds by showing with biochemical methods that almost all insects use AMPs as a first line of defense against infections. The next question to be answered was "how" they do it. Accordingly, teams of geneticists and molecular biologists started to work on the *Drosophila* model, and therefore got access to genetic tools to unravel the cascade of events implemented by the insect from the recognition of conserved bacterial motifs (pathogens-associated molecular patterns [PAMPs] resulting in the activation of two major signaling cascades (the Toll and immune deficiency [IMD] pathways) to the regulation of the transcription of AMPs-encoding genes by members of the nuclear factor-kappa B (NF-κB) family of inducible transactivators (including the dorsal-related immunity factor [DIF], Relish and Dorsal (reviewed in Imler, 2014)). These discoveries on the activation of innate immunity led the Nobel Committee to award the 2011 Nobel Prize in physiology or medicine to Jules Hoffmann (prize shared with Bruce Beutler and Ralph Steinman).

Supervision

Within our team, I worked daily with a younger PhD student, Pascale Fehlbaum, and although I was not officially in charge of supervising her, I spent time providing her advices and help.

Collaborations

* Alain Van Dorsselaer (Laboratoire de spectrométrie de masse bio-organique, Université de Strasbourg)

* Lucienne Letellier (Laboratoire des biomembranes, Université de Paris-Sud, Orsay)

* Max Goyffon (Laboratoire d'études et de recherches sur les arthropodes irradiés, Museum National d'Histoire Naturelle, Paris)

Funding

* Allocation from the Ministère de la Recherche et de la Technologie

Valorization

The work performed on insect innate immunity has been valorized in 6 publications (all as first/second author), 1 review (first author), 1 book chapter and 8 communications in meetings (oral or poster) (not listed here, see list of publications):

Bulet P, **Cociancich S**, Dimarcq J-L, Lambert J, Reichhart J-M, Hoffmann D, Hetru C & Hoffmann JA **(1991)** Insect immunity: isolation from a coleopteran insect of a novel inducible antibacterial peptide and of new members of the insect defensin family. *J. Biol. Chem.* 266, 24520-24525.

Bulet P, **Cociancich S**, Reuland M, Sauber F, Bischoff R, Hégy G, Van Dorsselaer A, Hetru C & Hoffmann JA **(1992)** A novel insect defensin mediates the inducible antibacterial activity in larvae of the dragonfly *Aeschna cyanea* (Paleoptera, Odonata). *Eur. J. Biochem.* 209, 977-984.

Cociancich S, Goyffon M, Bontems F, Bulet B, Bouet F, Menez A & Hoffmann J **(1993)** Purification and characterization of a scorpion defensin, a 4 kDa antibacterial peptide presenting structural similarities with insect defensins and scorpion toxins. *Biochem. Biophys. Res. Commun.* 194, 17-22.

Cociancich S, Ghazi A, Hetru C, Hoffmann JA & Letellier L **(1993)** Insect defensin, an inducible antibacterial peptide, forms voltage-dependent channels in *Micrococcus luteus*. *J. Biol. Chem.* 268, 19239-19245.

Cociancich S, Dupont A, Hégy G, Lanot R, Holder F, Hetru C, Hoffmann JA & Bulet P **(1994)** Novel inducible antibacterial peptides from a hemipteran insect, the sap-sucking bug *Pyrrhocoris apterus*. *Biochem. J.* 300, 567-575.

Chernysh S, **Cociancich S**, Briand J-P, Hetru C & Bulet P **(1996)** The inducible antibacterial peptides of the hemipteran insect *Palomena prasina*: identification of a unique family of proline-rich peptides and of a novel insect defensin. *J. Ins. Physiol.* 42, 81-89.

Cociancich S, Bulet P, Hetru C and Hoffmann JA **(1994)** The inducible antibacterial peptides of insects. *Parasitol. Today* 10, 132-139.

Hetru C, Bulet P, **Cociancich S**, Dimarcq J-L, Hoffmann D & Hoffmann JA **(1994)** Antibacterial peptides/polypeptides in the insect host defense. A comparison with Vertebrate antibacterial peptides/polypeptides. *Phylogenetic Perspectives in Immunity* (Eds Hoffmann JA, Janeway CA and Natori S) pp 34-65, RG Landes Company, Austin, TX, USA.

II. Study of *Plasmodium* / mosquito interactions

Post-doctoral fellowship at NIH - NIAID - Medical Entomology Section - Laboratory of Parasitic Diseases
Bethesda, MD (USA) (1995-1998) - Dir: Louis H Miller

II.1. *Plasmodium* / mosquito interactions: state of the art in 1995

At the twilight of the 20th century, despite decades of international efforts, malaria still was the most devastating insect-borne parasitic disease for humans, with 300 to 500 million new cases arising and over a million death recorded yearly in tropical countries. Drug-resistant parasites spread, mosquitoes resistant to insecticides emerged, and an efficient vaccine was still lacking. The development of the causal agent of malaria—the mosquito-borne *Plasmodium* parasite—consists in a complex multi-steps cycle with alternate human and mosquito phases. Controlling malaria through manipulations of the vector competence was an option seriously considered, but it required additional knowledge on molecular and cellular mosquito / parasite interactions.

II.2. Partial unraveling of an anti-*Plasmodium* factor from mosquitoes

At the Laboratory of Parasitic Diseases from NIH, studies focused on the development of *Plasmodium* in its mosquito vector. When a mosquito takes a blood meal from an infected vertebrate host, a complete transformation of the parasite occurs within the insect. Briefly, right after the blood intake, *Plasmodium* male and female gametes emerge from the red blood cells and fertilize to form zygotes. Within hours, round zygotes transform into banana-shaped motile invasive ookinetes which evade from the blood bolus. They penetrate the peritrophic matrix and cross the midgut epithelium until they reach the basal lamina where they round up and form oocysts. Days later, each oocyst releases into the hemolymph thousands of sporozoites which invade the salivary glands, allowing the mosquito to infect new vertebrate hosts (figure 4). Each stage of the malaria parasite development is therefore a potential target for disrupting its transmission cycle.

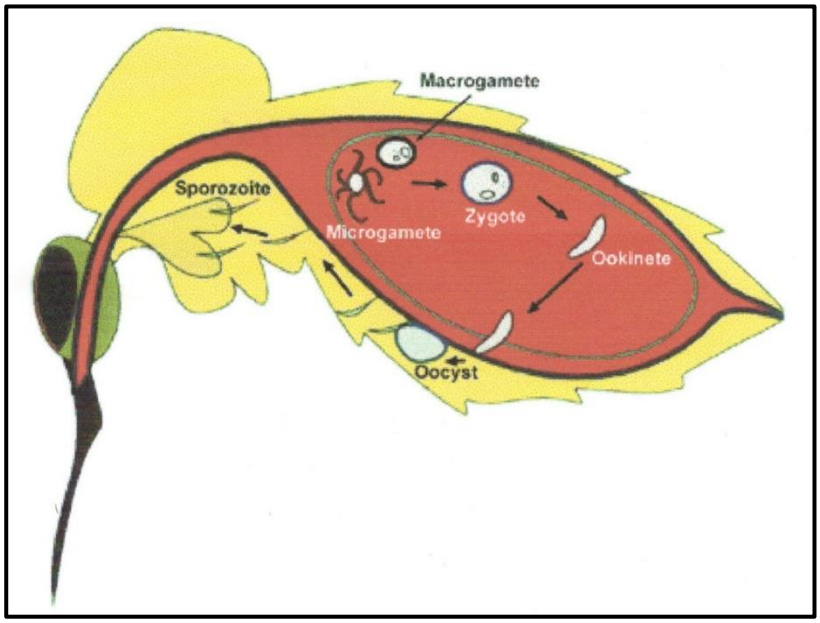


Figure 4. Schematic development cycle of the malaria parasite *Plasmodium* in its mosquito vector.

Only a few species of mosquitoes permit the successful development of the malaria parasite. For instance, anopheline mosquitoes are good vectors for the human malaria parasite, as opposed to *Aedes* or *Culex* species. On the contrary, *Aedes aegypti* is a susceptible mosquito for the avian malaria parasite *Plasmodium gallinaceum*, whereas *Anopheles gambiae* is refractory. These observations indicate that mosquito factors might play an important role in the development of the parasite. These factors can be multiple, especially

because of the many different biological forms of the parasite (gamete, zygote, ookinete, oocyst, sporozoite) and its presence in many compartments within the insect (blood bolus, midgut cells, hemolymph, salivary glands). They can be physical barriers, digestive enzymes, hormones, molecules from the immune system, etc. Using the *P. gallinaceum* model, I observed a sequential decrease in the number of parasites after each stage of the development cycle in refractory mosquitoes but also, to a lower extent, in susceptible mosquitoes. In all cases, the higher rate of parasite attrition is observed during the zygote-to-ookinete transformation. Therefore, mainly because it was available in much larger quantity in the insectarium, I investigated *Ae. aegypti* (susceptible species but in which a parasite attrition happens) for a zygote-to-ookinete transformation-blocking activity.

Crude extracts of *P. gallinaceum*-infected bloodfed *Ae. aegypti* mosquito batches were fractionated by solid-phase extraction. The fractions were screened using an *in vitro* biological assay. Briefly, I prepared twice-weekly *P. gallinaceum* zygotes batches from freshly collected infected chicken blood. When put in an adequate medium, zygotes transform *in vitro* into ookinetes. Using this biological assay, I identified in *Ae. aegypti* extracts a zygote-to-ookinete transformation-blocking activity. These extracts were submitted to several steps of reversed-phase HPLC. After each round of separation, all the resulting fractions were tested for their ability to block the zygote-to-ookinete transformation.

I finally detected a single zygote-to-ookinete transformation-blocking fraction in the HPLC chromatogram and purified the corresponding molecule. I named it *Ae*TBF-1 for "*Aedes aegypti* transformation-blocking factor 1" (figure 5). In collaboration with Henry Fales at Bethesda, we determined by MS the molecular weight of the molecule ($[M+H]^+$ = 503.2 Da). Additional techniques aiming at describing *Ae*TBF-1, such as Edman degradation and MS fragmentation pattern in ionic bombardment, indicated that it is not a conventional peptide. However, we depicted several characteristics of the molecule: acetylation experiments indicated the presence of three primary or secondary amines, whereas comparison of MS^n spectra with available standards in databases led us to propose that a glutamic acid stands at the N-terminus whereas the C-terminal part corresponds to a glycine. Because of the limited amount of pure material available, we failed in obtaining exploitable NMR spectra.

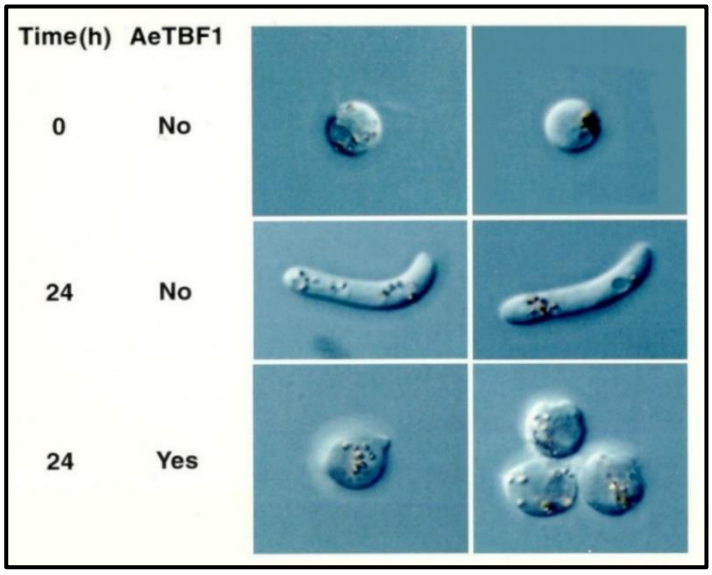


Figure 5. Zygote-to-ookinete transformation-blocking activity of *Ae*TBF-1. Top panels: *P. gallinaceum* control zygotes. Middle panels: *P. gallinaceum* control ookinetes resulting from the 24h-transformation of untreated zygotes. Bottom panels: *Ae*TBF-1-treated *P. gallinaceum* zygotes (24h-treatment). The transformation into ookinetes has been aborted. Note the unusual irregular shape of the untransformed zygotes. All observations were performed with optical microscopy (X400) (Cociancich, unpublished results).

Although this is the first report of a purified mosquito molecule that may contribute to impair the development of the malaria parasite, this project unfortunately stopped before I could fully characterize *Ae*TBF1 as my fellowship ended.

II.3. Study of specific mosquito midgut cells preferentially invaded by the *Plasmodium* parasite

The malaria parasite faces several physical barriers during its development in the mosquito. One of them is the midgut epithelium, which has to be crossed by ookinetes from the luminal side towards the basal side where they stand and transform into oocysts. Interestingly, oocysts are primarily distributed at the posterior part of the posterior midgut. This biased distribution is commonly explained by gravity since mosquitoes rest for several hours in a vertical position after a blood meal, allowing ookinetes to settle down at the bottom portion of the posterior midgut. But as we submitted for several hours blood-fed mosquitoes to a device which nullifies gravity, the biased oocysts distribution towards the posterior part of the posterior midgut was not impaired.

A previous study performed at the laboratory outlined that a scarce specific cell type (namely Ross cells) within the midgut epithelium is preferentially invaded by *P. gallinaceum* ookinetes. Compared to regular columnar midgut cells, Ross cells are less basophilic, less osmiophilic in electron microscopy, devoid of microvilli, and they over-express a vesicular ATPase (v-ATPase) (Shahabuddin & Pimenta, 1998). Their presence in uninfected midguts proves that these characteristics are not the consequence of the *Plasmodium* invasion.

Since *Plasmodium* seems to invade preferentially Ross cells, and oocysts develop mainly at the very end of the posterior midgut, we decided to determine whether the distribution pattern of Ross cells within the midgut epithelium shows any bias towards the posterior part of the posterior midgut. As Ross cells over-express a v-ATPase, we decided to use the *Ae. aegypti* v-ATPase as a marker to examine their distribution in the midgut epithelium. I cloned the *Ae. aegypti* v-ATPase B-subunit and expressed it as a recombinant protein to produce an antibody. I ascertained by western-blot the specificity of the v-ATPase antiserum which I subsequently used in immunofluorescence experiments with dissected mosquito midguts. As a result, I observed an over-expression of the v-ATPase in a limited number of midgut cells. Interestingly, their pattern of distribution matches the biased distribution of the oocysts in the midguts. Finally, the observation by immunofluorescence of a co-localization between Ross cells and midgut-invading ookinetes added up to this “Ross cells model” for the *Plasmodium* ookinetes invasion of the mosquito midgut epithelium (figure 6) (Cociancich *et al.,* 1999).

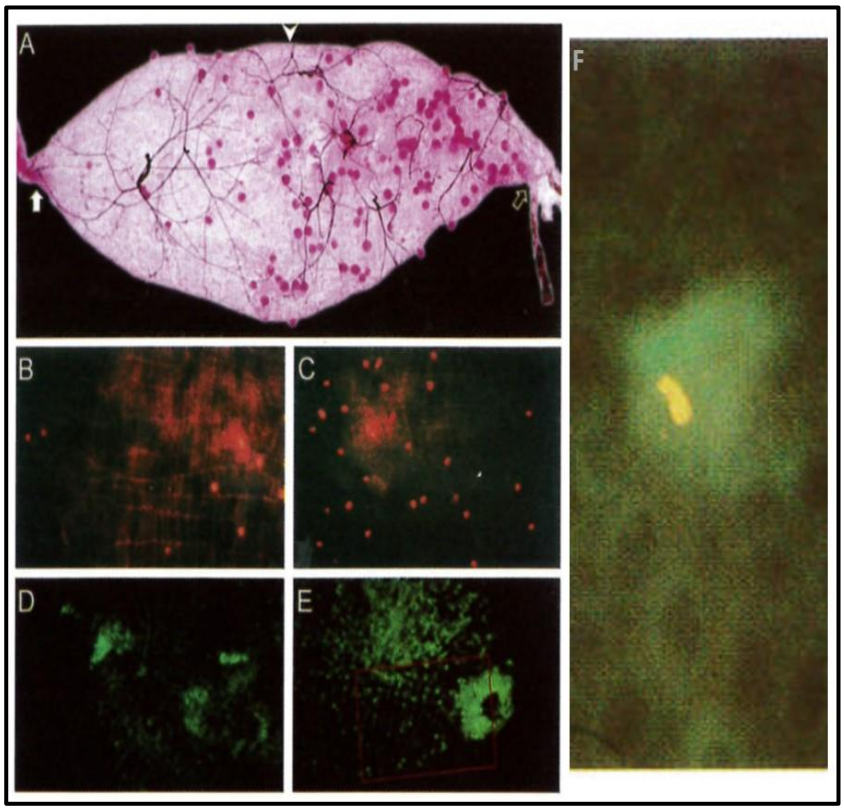


Figure 6. *P. gallinaceum* ookinetes invasion and oocysts development in the posterior midguts of gravity-deprived mosquitoes. A: Oocysts distribution eight days post-blood meal. Note the biased distribution of oocysts towards the posterior half of the posterior midgut. B: Ookinete invasion 42 hours post-blood meal in the anterior half and C: in the posterior part of the posterior midgut. Note the higher frequency of ookinetes in the posterior half. D: v-ATPase over-expressing cells in the anterior half and E: in the posterior part of the posterior midgut. Note the higher frequency of v-ATPase over-expressing cells in the posterior half. F: Co-localisation of a v-ATPase over-expressing cell (Ross cell) and of an ookinete. For detailed experimental conditions, see Cociancich *et al.,* 1999.

And since then, what's new?

Two decades later, malaria still kills near half a million people (mostly children) each year. *Plasmodium* / mosquito interactions studies are still a hot topic, and several programs involving the release of sterile or genetically-modified mosquitoes have been launched, only with poor results. Regarding my results on the preferential invasion by *Plasmodium* of specific midgut Ross cells, a competing team working on the *Anopheles*-transmissible rodent malaria parasite *P. berghei* proposed later another model, referred to as the "time-bomb model" to report the crossing of the mosquito midgut by *Plasmodium* ookinetes (Han *et al.,* 2000). They state that *P. berghei* ookinetes do not invade specific midgut cells. Instead, they provoke detrimental effects to the invaded cells, the latter being extruded from the midgut wall into the lumen side, forcing ookinetes to quickly evade these cells before they are removed. A controversial debate took place over which model was the right one (Shahabuddin, 2002; Baton & Ranford-Cartwright, 2005), and even if as of today, the "time-bomb model" seems to have the preference of many scientists over the "Ross cells model", no definitive picture is yet available.

Supervision

During my fellowship, I supervised a BS student, Rizwan Ahmed, during his summer internship. I trained him for HPLC purification of mosquito crude extracts, and he took an active part to the purification of *Ae*TBF-1. Also, I dedicated time to help three post-doctoral fellows, Chuong Huynh, Soon Park and Paul Docherty, as they joined the team.

Collaborations

* Henry Fales (Laboratory of applied mass spectrometry, National Institutes of Health, Bethesda, MD, USA)

* Jill Troyer, Bruce Noden & Abdu Azad (School of medicine, university of Maryland, Baltimore, MD, USA)

Funding

* Fogarty International Center / NIH / John T & Catherine D McArthur foundation fellowship

* Project n° 950476: Identification of mosquito midgut receptors for *Plasmodium* ookinetes. 1996-1998. UNDP / World Bank / WHO Special programme for research and training in tropical diseases. 88750 $

Valorization

The work performed on *Plasmodium* / mosquitoes interactions has been valorized in 1 publication (first author), 1 review (second author) and 9 communications (1 invited) in meetings (oral or poster) (not listed here, see list of publications):

Cociancich SO, Park SS, Fiddock DA & Shahabuddin M **(1999)** Vesicular ATPase-overexpressing cells determine the distribution of malaria parasite oocysts on the midguts of mosquitoes. *J. Biol. Chem.* 274, 12650-12655.

Shahabuddin M, Cociancich S & Zieler H **(1998)** The search for novel malaria transmission-blocking targets in the mosquito midgut. *Parasitol. Today.* 14, 493-497.

PART 2
SCIENTIFIC STUDIES AT UMR BGPI (2003/ON GOING)

PAST WORK + PRESENT & FUTURE PROJECTS

CONTRIBUTION OF GENETICS FOR THE CHARACTERIZATION AND DISCOVERY OF NON-RIBOSOMALLY-SYNTHESIZED NATURAL PRODUCTS IN PLANT-ASSOCIATED MICROORGANISMS

"Au fond de nous, l'étincelle..." **Etienne Daho, Blitz, Virgin**

General outline

Microbial secondary metabolites are natural products displaying, for most of them, therapeutically- or agrochemically-relevant biological activities (e.g. antibiotics, antifungal or anti-proliferative agents, siderophores etc). Interestingly, many of these compounds are derived from non-ribosomal biosynthetic enzymatic pathways which are able to generate near-infinite complex structures. For centuries, new bioactive secondary metabolites have been identified through biological activity screenings, before being isolated and structurally-characterized. However, at the twilight of the 20th century, the number of new biologically-relevant microbial secondary metabolites decreased drastically, notably because of a higher rate of rediscovery of known compounds or an increased interest in creating huge sets of new synthetic molecules using combinatorial chemistry compatible with high-throughput screenings. This prompted the research community to turn down its interest towards finding new natural compounds because it considered that the reservoir of new microbial natural products was about to dry up. But the actual exponential increase of available genomic sequences (including sequences from thousands complete microbial genomes) suddenly gives access to a huge amount of information which reveals, on the very contrary, that the potential of microorganisms to produce secondary metabolites exhibiting new structural scaffolds, and therefore new biological activities, is largely underestimated. However, the actual challenge remains to exploit this potential.

Since I joined, in 2003, the team "Génomique et analyse moléculaire de la pathogénie des bactéries phytopathogènes" led by Philippe Rott at UMR BGPI ("Biologie et Génétique des Interactions Plante/Parasite) from Cirad, I have been interested in studying microbial secondary metabolites synthesized by the non-ribosomal pathway in plant-associated bacteria. Having been trained as a biochemist, I initially favored biochemical analytical approaches to isolate and characterize molecules of interest. To some extent, however, this has been proven to be not sufficient when it comes to the identification and complete characterization of structurally-complex natural products such as those synthesized *via* the non-ribosomal pathway. Indeed, in this chapter, going through past works as well as current and future projects, I want to show (i) how the *in silico* analysis of non-ribosomal biosynthesis genes efficiently contributes, in complement to biochemical and molecular biological techniques, to the over-production and structural characterization of already-discovered natural products, and (ii) how the genome mining approach allows to discover new secondary metabolites through the screening of bacterial genomes for the presence of biosynthetic genes clusters (BGCs).

The present part of the document is divided in two chapters. The first chapter is devoted to works and projects on structural characterization of already-discovered bacterial natural products. It starts by an introductive section summarizing the basics of the non-ribosomal peptide synthesis mechanisms and how the *in silico* analysis of biosynthesis genes provides insights into the biosynthetic pathways and the molecular structures of the natural products. This section is by no way meant to be exhaustive but detailed enough to appreciate the whereabouts of my works on microbial non-ribosomal secondary metabolites. I will then detail my work on albicidin, a long-term effort which introduced me in the first place to the fascinating world of the non-ribosomal peptide synthesis. As a matter of fact, this work could only be completed by combining biological data obtained through analytical techniques performed on the purified molecule with data provided by *in silico* analyses of the natural product biosynthesis machinery. Our team was finally rewarded after a decade-long struggle resulting in the almost complete description of the biosynthetic pathway of albicidin and in the determination of its structure, which displays many unique features making this antibiotic a lead structure for drug development. As such, this study is an archetypal example of the multidisciplinary efforts required for the full characterization of microbial non-ribosomal natural products. During this work on albicidin, I acquired significant knowledge and expertise on the non-ribosomal biosynthetic processes, and I have been recently solicited to develop a new promising project requiring *in silico* analysis of the biosynthetic gene cluster of a new type of quorum sensing signal in the plant pathogen *Dickeya* [project "VFM"].

The second chapter starts with an introductive section describing the genome mining approach for revealing new BGCs in microbial genomes. This section is followed by a general description of a common strategy that I implemented (or plan to implement) in the course of three current and future projects which will be subsequently described. A first project, still on-going, is devoted to the study of a promising non-ribosomally synthesized metabolite [project "Meta-B"]. We discovered its BGC as we mined the sequenced and annotated genome of *X. albilineans*. No information was available regarding the Meta-B production by the bacterium, nor about its biological function. Hence, *in silico* analyses of the corresponding biosynthetic genes were used to predict partially the structure of the Meta-B molecule, and to drive molecular biology and biochemical methods in order to fully describe the compound. Remarkably, our interest in Meta-B was reinforced as we identified similar BGCs in the genomes of some other *Xanthomonas* species after we launched a global genome mining of all available genome sequences from *Xanthomonas* species. This genome mining effort rendered data indicating that the *Xanthomonas* genus is an under-explored reservoir for new secondary metabolites. I therefore proposed a project aiming at sequencing and mining for BGCs the genomes of a large number of *Xanthomonas* strains representative of the diversity of the whole genus [project "XanthoNRPS"].

Capitalizing on the knowledge acquired over the years on non-ribosomal secondary metabolites, our team (recently renamed METABOL, "Antibiotiques et autres métabolites secondaires chez les micro-organismes phytopathogènes"), now led by Monique Royer, has recently been solicited to build a collaborative project aiming at discovering new non-ribosomal secondary metabolites in *Streptomyces* bacteria. I therefore co-developed a project concerning the mining and biological screening of a collection of *Streptomyces* strains for the presence of promising new secondary metabolites [project "CONTACT"]. This project has recently been funded by the Agropolis Foundation and has just started.

All these past works and present and future projects represent collaborative efforts with several partner teams and involve(d) many students whose respective contributions will be acknowledged in the document.

Chapter 1
Contribution of *in silico* analyses of biosynthesis genes to the characterization of already-discovered natural products

I. The non-ribosomal peptide synthesis

I.1. Generalities

All living organisms fundamentally rely on ribosomes to synthesize peptides and proteins. The elucidation of the universal genetic code in the early 1960s allowed to translate nucleotidic sequences of DNA and RNA into aminoacid (aa) sequences, establishing a correspondence between each of the 64 possible 3-nucleotide codons and one of the 20 proteinogenic α-L-aas incorporated into a growing polypeptidic chain (with the exception of the three "STOP" codons). However, many bacteria and fungi possess an additional mechanism for peptidic synthesis, independent from the ribosome activity, and thus called non-ribosomal peptide synthesis. Interestingly, most of secondary metabolites from microorganisms are synthesized *via* the non-ribosomal peptide synthesis machinery, including antibiotics (e.g. vancomycin, precursor of penicillin, bacitracin, chloramphenicol, streptomycin, kanamycin), immunosuppressing molecules (e.g. cyclosporine), antitumoral agents (e.g. bleomycin, actinomycin), siderophores (e.g. enterobactin, yersiniabactin), toxins etc. Non-ribosomal peptide synthesis is dependent upon the activity of non-ribosomal peptide synthetases (NRPSs), which are mega-enzymes bearing a modular architecture. They function as thio-template assembly lines, in which each module governs the recognition and activation of a specific aa substrate and its sequential incorporation into a nascent polypeptidic chain. The colinearity rule mostly applies, meaning that five modules are required for the biosynthesis of a pentapeptide, for instance. The substrate for a NRPS module is either one of the 20 proteinogenic α-aas or any small unit bearing both an amino and a carboxyl group, *i.e.* β-aas (Miyanaga *et al.,* 2014), δ-aminobenzoic acids (Cociancich *et al.,* 2015) or even α-hydroxy-acids (Feifel *et al.,* 2007). To date, over 500 of these "building blocks" or "monomers" have been identified (for an updated list, see the NORINE website at http://bioinfo.lifl.fr/norine/) and participate to the structural diversity of the non-ribosomal peptides (NRPs).

I.2. Mechanism and genetic organization

Mechanisms underlying non-ribosomal peptide synthesis are utterly complex (for a comprehensive review, see Süssmuth and Mainz, 2017) but can be summarized as follows: as previously mentioned, NRPS exhibit a multi-modular architecture. Each NRPS module is typically composed of three obligatory domains referred to as adenylation (A), peptidyl carrier protein (PCP, also called T for thiolation) and condensation (C) domains.

An A domain is approx. 500-aa long, and specifically selects and activates by adenylation an aa substrate according to an internal signature sequence involving ten scattered aas as deduced from the analysis of the first-ever structure of an A domain, *i.e.* the phenylalanine-activating A domain from GrsA–a NRPS involved in gramicidin S biosynthesis (Conti *et al.,* 1997). The lateral chains of these ten aas point towards the substrate-binding pocket of the A domain and, depending on their nature, stabilize a specific aa-building block. The correlation between specific 10-aa signature sequences within A domains and their cognate specific substrates allowed to determine a specificity-conferring code (or non-ribosomal code) (Stachelhaus *et al.,* 1999; Challis *et al.,* 2000) that has subsequently been used to develop algorithms and softwares to help for predictions in genome mining approaches (see below for more details). Nevertheless, the non-ribosomal code is not strict since smaller substrates such as glycine require only part of the ten aas constituting the substrate-specificity conferring signature for their binding into the substrate pocket (Drake *et al.,* 2016). However, the Asp in first position and the Lys in 10th position in the substrate-specificity conferring signature are of particular importance since they stabilize the amino- and carboxy-groups of the substrate, respectively. The Lys residue is strictly conserved, whereas the Asp residue can sometimes be replaced or found at another position within the signature sequence to accommodate the interaction with substrates such as β-aas or

aminobenzoic acids, for instance (figure 7). The full code governing the NRPS substrate recognition is not yet fully deciphered, and new "unknown" signature sequences within A domains are found as new bacterial genome sequences are available.

A domain	specificity signature										substrate
	1	2	3	4	5	6	7	8	9	10	
GrsA	D	A	W	T	I	A	A	I	C	K	Phe
ApnA-A1	D	V	E	S	I	G	V	I	A	K	Arg/Tyr
CytC1	D	F	w	n	i	G	g	V	F	K	D-Val
DltA	D	L	M	t	l	C	t	V	a	K	D-Ala
AB3403	D	I	l	q	l	g	v	V	W	K	Gly
BlmIV-A1	V	D	W	V	I	S	L	A	D	K	β-Ala
ESyn-A1	G	A	L	M	V	V	G	S	I	K	α-HIV
Vlm1-A1	A	A	L	W	I	A	V	S	G	K	α-KIV

Figure 7. Some examples of substrate specificity-conferring signature sequences from NRPSs A domains and their corresponding substrates (L/D-α/β-aas with variable sizes of lateral chains and α-hydroxy/keto acids). In the signatures, only the residues shown in capital are involved in substrate binding. The conserved Asp and Lys residues are highlighted in red. The boxed Ser residues replace the Asp residue for the stabilization of the α-hydroxy/keto-acid substrates. Adapted from Süssmuth and Mainz, 2017.

A PCP domain is approx. 100-aa long, and does not have any catalytic activity. An A domain covalently loads its specifically-activated adenylated aa substrate onto a 4'-phosphopantheteine (pPant) prosthetic group present on the immediate downstream PCP domain, creating a thioester bond (figure 8). The flexible pPant arm is beforehand brought by a 4'-phosphopantetheinyl-transferase (PPTase), which is an enzyme mandatorily required for the post-translational activation of NRPS systems, and allows the PCP-tethered thioesters to be transferred to the next domain along the biosynthetic assembly line.

The adjacent C-domain (approx. 450-aa long) then catalyzes the coupling (or condensation) reaction between the aa or nascent polypeptidic chain present on the PCP domain of module n-1 and the aa carried by the PCP domain of module n. The condensation is catalyzed through the nucleophilic attack of the amine group of the acceptor aa (module n) towards the thioester group of the donor aa (module n-1), and results in the formation of a peptidic (amide) bond (figure 8). Usually, the first module of a NRPS is devoid of C domain since it incorporates the first building block. Sometimes, however, the initiation module includes a C domain (referred to as "starter C domain") which performs the linkage of the nascent polypeptidic chain onto an acyl moiety, resulting in the biosynthesis of a lipopeptide. Also, some C domains from elongation and termination modules do epimerize the aa incorporated by the module n-1 (dual C/E domains; condensation/epimerization) (for reviews, see Bloudoff and Schmeing, 2017; Süssmuth and Mainz, 2017). Indeed, the presence of D-aa participates to the structural diversity as well as it increases the overall resistance of the peptide to proteolysis. Finally, the last (termination) module of the NRPS usually contains an addidional terminal domain named TE for thio-esterase, which releases by hydrolysis the polypeptidic chain from the modular NRPS template. Other modifications can also be achieved by the terminal TE domain in the NRP assembly line, such as intramolecular cyclisation or intramolecular hydrolysis (thioesterase-type), or even epimerization or dimerization (Payne *et al.,* 2017).

The overall complex structural diversity of NRPs results not only from the incorporation within the elongating peptidic chain of building blocks of various chemical nature, but also eventually from the action of additional domains. Indeed, besides the basic [C-A-PCP] domain organization for a NRPS module, additional domains are sometimes present which catalyze the co-synthetic tailoring of the growing peptide, resulting in methylation, oxydation, cyclization, formylation or hydroxylation reactions, among others. Although most NRPS systems are dependent upon the colinearity rule *(i.e.* the final assembled peptide is entirely determined by the number and order of modules within the corresponding NRPS), some others alternatively use an iterative functioning in which part of or all modules are re-used during the biosynthesis of their final products (the siderophore enterobatin, for instance) (Drake *et al.,* 2006).

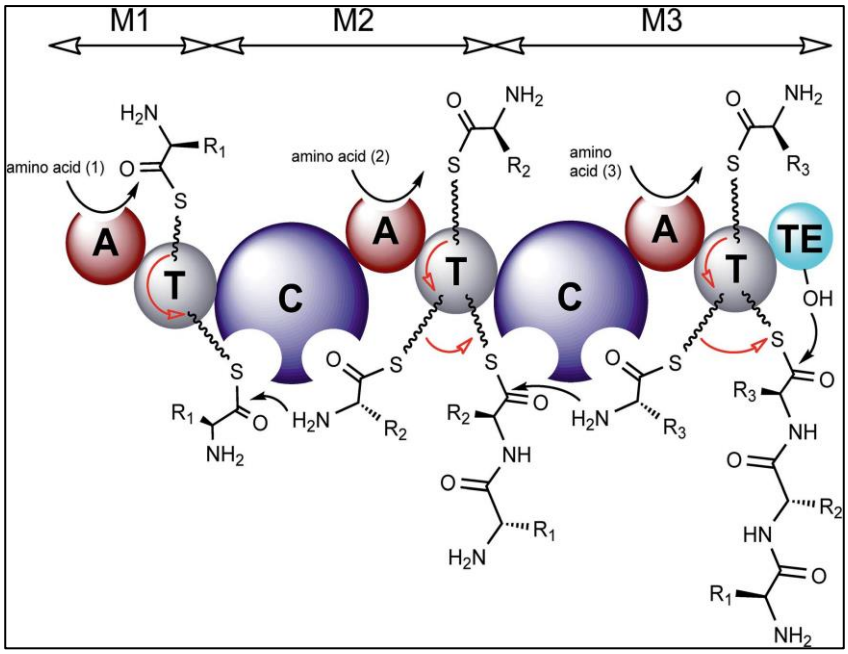


Figure 8. Canonical modular organization of a NRPS and associated model of biosynthesis. M1, M2 and M3 stand for the modules referred to as initiation, elongation and termination, respectively. aa-substrates are specifically activated by the adenylation domains (A) and transferred onto the thiol group of the pPant arm located on the adjacent thiolation (also called PCP) domains (T). Condensation domains (C) catalyze the peptide bond formation between thioesters tethered on adjacent T domains. At the end of the termination module, a thioesterase domain (TE) catalyzes the release of the final peptide by hydrolysis. The growing peptide is prone to co-synthetic tailoring by surnumerary optional domains, and the final peptide can be post-synthetically modified by the products of tailoring genes present within the same BGC. Taken from Winn *et al.,* 2016.

Because of their modular structure, NRPSs are variable in size, ranging from a stand-alone module up to 18 modules (encompassing molecular weights > 2 MDa). The corresponding encoding genes therefore also display huge sizes, for one NRPS module is encoded by a 3-kb portion of genome. Moreover, it happens frequently that a single NRP is biosynthesized by a NRPS assembly line encoded by several physically-clustered genes whose products arrange themselves to reconstitute the full logic synthetic unit. Indeed, in bacterial and fungal genomes, genes encoding NRPSs are usually part of larger regions called genomic islands or biosynthetic gene clusters (BGCs). Besides NRPS-encoding gene(s), BGCs do contain several other genes involved either in the biosynthesis of specific building blocks being assembled into the final natural product and/or in the modification, decoration, transport and regulation of the NRP, for instance.

NRPSs are not the only multi-modular machines able to synthesize natural products independently from the ribosome: polyketide synthases (PKSs) are modular mega-enzymes sharing characteristics with NRPSs, with the remarkable difference that they catalyze the incorporation and condensation of acyl-thioester extension units such as malonyl-CoA and methylmalonyl-CoA. Their domain organization within modules resembles the one from NRPSs: briefly, an acyl-transferase (AT) domain specifically recognizes and activates an extension unit which is transferred and covalently-linked to the pPant prosthetic group onto the downstream acyl carrier protein (ACP) domain. A ketosynthase (KS) domain subsequently links covalently the ACP-tethered thioesters from modules n-1 and n. As for NRPs, polyketides (PKs) can be co-synthetically and/or post-synthetically modified by additional intramolecular domains (*e.g.* ketoreductase [KR], dehydratase [DH], enoyl-reductase [ER]) and/or by tailoring genes present within the BGC. In fungi, PKSs tend to function preferentially according to an iterative pattern, at least much more than NRPSs. Strikingly, many secondary metabolites are synthesized as hybrid PKS/NRPS.

This rapid description is not meant to be exhaustive and only barely explores the detailed core-mechanisms underlying the non-ribosomal peptide synthesis pathways which involve far more intricate mechanisms accounting for the full diversity of molecular structures that have been implemented by Nature. Nevertheless, it provides to the reader enough data to follow my work on non-ribosomal microbial secondary metabolites.

II. Multi-disciplinary approaches for the structural characterization of albicidin, a potent antibiotic discovered in 1985

II.1. State of the art at the beginning of my work

Albicidin is produced by the pathogenic bacterium *Xanthomonas albilineans*, which is the causal agent of leaf scald disease. This yellow-pigmented mono-flagellated Gram-negative bacterium invades and colonizes the xylem of sugarcane, and provokes symptoms varying from a single, white, narrow, sharply-defined stripe to full wilting and necrosis of leaves, leading to plant death (figure 9) (For review, see Pieretti *et al.,* 2015).

Figure 9. Symptoms of sugarcane leaf scald disease. Symptoms are of incresing severity from left to right (taken from Champoiseau, 2006).

The origin of this work lies over 30 years ago, when Robert Birch and colleagues (University of Queensland, Australia), who discovered albicidin, showed that it is responsible for the characteristic white pencil-line symptoms of sugarcane leaf scald disease (see figure 9) (Birch & Patil, 1985a; Birch & Patil, 1985b). Albicidin, which also exhibits a very powerful antibiotic activity at nanomolar concentrations against *Escherichia coli*, impairs DNA synthesis, and the same group revealed that it is a potent inhibitor of the supercoiling activity of plant and bacterial DNA gyrases, acting in a different way than the fluoroquinolone antibiotics currently used on the market (Birch & Patil, 1985a; Hashimi *et al.,* 2007). Birch and colleagues developed a protocol for albicidin purification and, while detecting a mixture of antibiotic compounds, they finally succeeded in purifying the main form of albicidin (referred to as β-albicidin) which was analyzed by MS and proton- and ^{13}C-NMR. The resulting data indicated that β-albicidin has a molecular weight of 842 Da, possesses approx. 38 carbon atoms, several aromatic rings, and a carboxyl group. The NMR spectroscopy data nevertheless appeared to be insufficient to achieve the full structural characterization of β-albicidin. In the same study, the biological activity of albicidin was shown to be undiminished after heat- and pronase-treatments (Birch & Patil, 1985b).

Progress towards albicidin structural characterization have been later obtained through the description of its biosynthetic machinery. In the late 1990s, two independant teams, *i.e.* Birch's group and Philippe Rott's group at Cirad (in collaboration with Prof. Dean Gabriel at University of Florida), investigated the genes involved in the biosynthesis of albicidin. If both teams identified multiple genes with a modular architecture characteristic of PKSs and NRPSs (Rott *et al.,* 1996; Huang G *et al.,* 2001), the team from Montpellier described through a very elegant work the complete set of genes responsible for the biosynthesis of albicidin (Royer *et al.,* 2004). In addition, this work provided an extensive *in silico* analysis of the corresponding PKSs and NRPSs, yielding hypotheses that appeared later to be crucial for the structure determination of albicidin. The key results of this paper are summarized below (see figure 10, and for extensive information, Royer *et al.,* 2004).

- The study of 47 out of 50 albicidin-deficient mutants (picked from 7000+ *X. albilineans* mutants obtained through a Tn*5* mutagenesis experiment) allowed to unravel two genomic regions involved in albicidin production. The first region spans over 56 kb and contains a BGC of 20 open reading frames. Interestingly, three out these 20 genes encode PKSs and NRPSs. Other genes from this cluster are involved in secretion, regulation, decoration or self-resistance to albicidin. The second genomic region contains a single gene encoding a PPTase which, as a reminder, is required for the post-translational activation of PKSs and NRPSs.

- A single assembly line is formed by the products of genes *albI* and *albIX*. *albI* encodes a hybrid PKS/NRPS with three PKS modules followed by three NRPS modules encompassing the canonical [C-A-PCP] domain organization (referred to as NRPS-1, -2 and -3, respectively) and an additional single C domain (incomplete module NRPS-4). The product of *albIX* first displays an incomplete NRPS module bearing an A and a PCP domains (incomplete module NRPS-6, which connects with the incomplete module NRPS-4 to form a complete module), followed by the terminal NRPS module (NRPS-7) ending with a TE domain (figure 10).

- The PKS modules were proposed to incorporate successively an acetyl-CoA substrate and two malonyl-CoA substrates with the help of a trans-acting acyl-transferase encoded by *albXIII*. From the study of the substrate-specificity conferring signatures of the NRPS modules, the authors concluded that modules NRPS-1 and -3 possess a similar, although unknown, substrate-specificity signature sequence, and thus are predicted to incorporate the same substrate named X1. This is also the case for modules NRPS-6 and -7, both predicted to incorporate the same unknown substrate X2. In all these modules, the Lys is conserved in 10th position of the signature sequence, indicating that substrates X1 and X2 are carboxylated, but the Asp in first position of the signature sequence is absent, indicating that the substrates X1 and X2 are not α-aas. However, similarities with signature sequences from NRPS modules involved in biosynthesis of rifamycin plead for the incorporation of aromatic substrates (Admiraal *et al.,* 2001).

- The analysis of the sequence of module NRPS-2 indicates that its substrate-binding pocket is non-functional. The authors propose module NRPS-2 to be trans-complemented by the stand-alone module NRPS-5 encoded by the gene *albIV*, containing only an A and a PCP domains. The corresponding signature sequence is predicted to drive the incorporation of an asparagine residue. Interestingly, the authors noted the presence within the A domain of an additional domain of over 300 aas which includes an ATP-binding site and shows homologies to enzymes belonging to the adenosine nucleotide α-hydrolases superfamily (αANH).

- Based on this thorough *in silico* analysis, a theoretical structure for the backbone of albicidin has been proposed (figure 10). The hypothesis that substrates X1 and X2 are aromatic carboxylated building blocks is backed up by the presence, within the albicidin BGC, of several genes encoding enzymes involved in the metabolism of aromatic moieties such as a *para*-aminobenzoate synthase, a hydroxybenzoate-CoA ligase and a hydroxybenzoate synthase.

- Additional modifications are expected to be performed by the products of other genes present in the BGC such as, for instance, a carbamoyl-transferase or two methyltransferases.

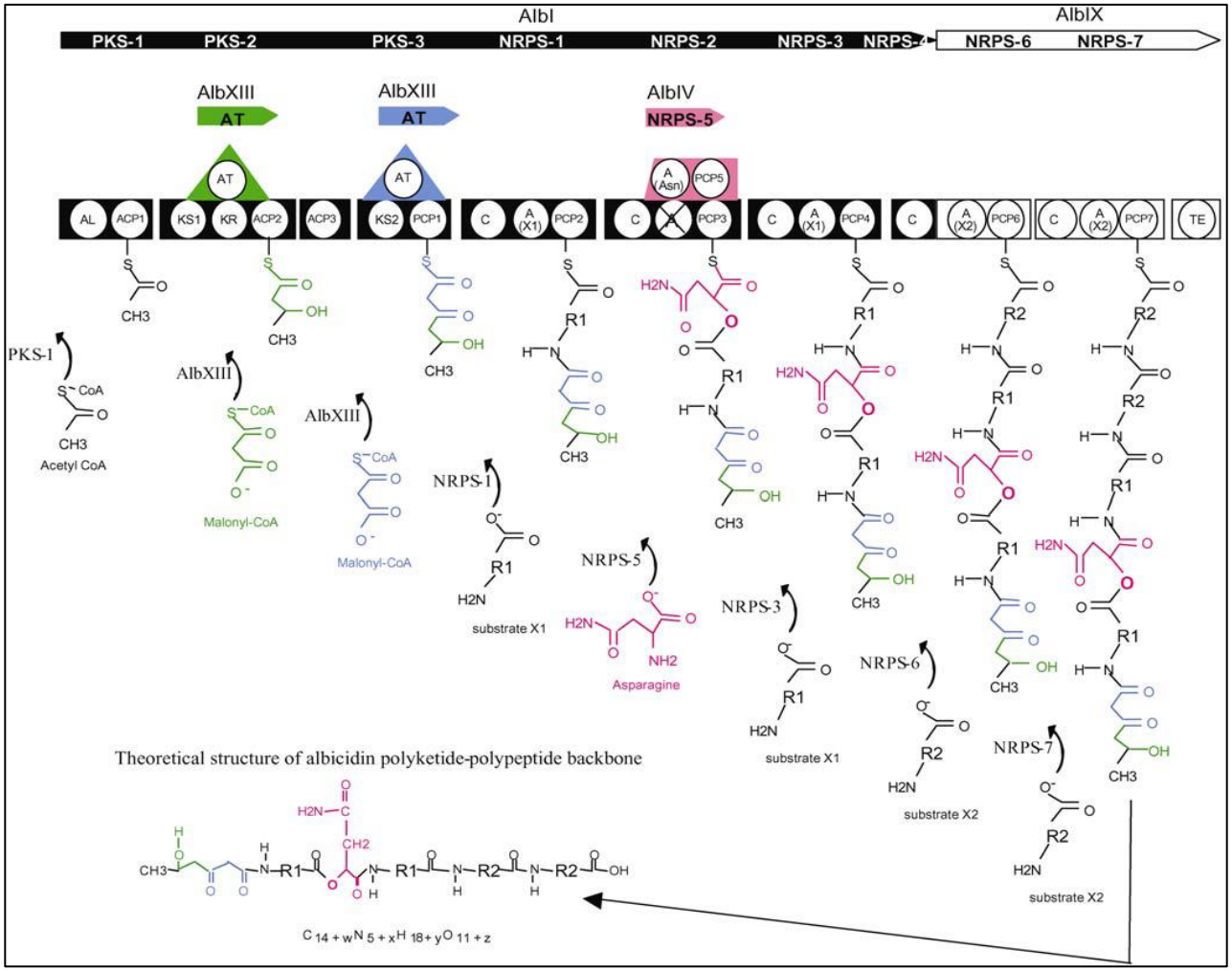


Figure 10. Hybrid PKS/NRPS assembly line for the biosynthesis of albicidin and theoretical structure of its backbone. Domain organization of the modules and correponding substrates are shown (taken from Royer *et al.,* 2004).

II.2. Discovery of a HtpG protein also required for albicidin biosynthesis

In frame of the PhD of Eric Vivien, the analysis of three yet-unstudied mutants from the transposon mutagenesis performed by Philippe Rott, all located in a region of the genome different than the albicidin BGC and the PPTase gene, allowed to identify a new gene which encodes a large protein sharing over 60% homology with the *E. coli* heat-shock protein HtpG, a prokaryotic homologue of the eukaryotic Hsp90 protein. This HtpG-encoding gene is required for biosynthesis of albicidin, as I showed the restoration of albicidin production in these mutants complemented by the wild-type (WT) gene. The role of the HtpG protein in albicidin synthesis is not yet defined but it has been hypothesized that it may act as a chaperone (Vivien *et al.,* 2005). Interestingly, the requirement of this protein has been more recently proved in the biosynthesis of two other NRPs, *i.e.* the cyclic lipopeptide arthrofactin in *Pseudomonas* sp. MIS38 (Washio *et al.,* 2010) and the contiguous genotoxin colibactin and siderophore yersiniabactin in *E. coli* (Garcie *et al.,* 2016).

II.3. Engineering of a heterologous system to over-produce albicidin

Since its discovery, and notably due to its potent antibiotic activity, albicidin has been considered as a strong lead candidate for drug development. The proposed theoretical structure reinforced its pharmaceutical interest because of its original features. However, it was of utmost importance to determine the full structure of the molecule. I developed biochemical approaches to purify the molecule, but as I started to work on the isolation

of albicidin from supernatants of *X. albilineans*, I faced the first of many yet-to-come bottlenecks, *i.e.* the low yield of production of albicidin by its natural producer.

To solve this problem, we decided to develop the heterologous expression of albicidin. This work has been performed by Eric Vivien in frame of his PhD. The targeted genes to be transferred to the heterologous host were obviously those present within the BGC identified during the transposon mutagenesis experiments, including the PKS/NRPS encoding genes, but also the PPTase gene and, finally, the HtpG-encoding gene. Complex molecular techniques were used by Eric Vivien to prepare a two-plasmid expression system carrying the albicidin biosynthetic machinery (Vivien *et al.,* 2007).

Several bacterial species have been tested as hosts for the transfer of the albicidin expression system. This operation failed with *E. coli*, maybe because the sigma transcription factors are very different between *E. coli* and *X. albilineans*. Thus, several *Xanthomonas* species have been tested, as we expected that the regulation of gene expression in a heterologous host from the same genus will be facilitated. Interestingly, a successful transformation of the fast-growing *Xanthomonas axonopodis* pv. *vesicatoria* (which has been re-classified later as *Xanthomonas perforans* [Schwartz *et al.,* 2015]) was achieved, as attested by the production of an antibiotic activity by the transformed strain hereafter referred to as *Xvesalb,* which does not naturally produce albicidin (the albicidin production is a specific feature of *X. albilineans* within the *Xanthomonas* genus) (Vivien *et al.,* 2007). We observed a production yield approx. 6- to 10-fold greater compared to *X. albilineans*, which represents a main breakthrough towards the isolation of albicidin (Vivien *et al.,* 2007). I additionally showed that the heterologously-produced antibiotic molecule has the same chromatographical behavior in thin layer chromatography than the native albicidin from *X. albilineans* (Vivien *et al.,* 2007). Moreover, as the natural albicidin, the heterologous albicidin is sensitive to the AlbD esterase, an albicidin-detoxifying enzyme previously isolated in *Pantoea dispersa* (Zhang & Birch, 1997). These results not only indicate that we had identified all the genes required for albicidin biosynthesis, but also that the heterologous molecule shares its chromatographical behaviour and biological activities with the native molecule. In addition, the *Xvesalb* heterologous producer grows in minimal medium, and this is of crucial importance for the purpose of chromatographic isolation of albicidin.

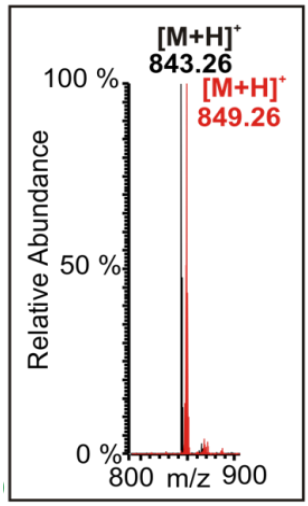
II.4. Unraveling the structure and biosynthesis pathway of albicidin

Nevertheless, the following tentative purifications of albicidin rapidly turned into a recurrent nightmare as the albicidin produced by *Xvesalb* cultures tend to stick permanently on whatever surface it encounters (the glass of the Erlenmeyer flasks used for the bacterial cultures, solid-phase extraction cartridges, filters, etc.). In response to this problem, we added XAD-7 Amberlite resin in *Xvesalb* cultures grown in minimal medium. Consequently, freshly produced albicidin is continuously adsorbed onto the resin during bacterial growth and can be subsequently eluted from the resin.

With the production of increased amounts of albicidin, and the possibility to concentrate the secreted albicidin with a resin, I resumed the purification process. Meanwhile, I initiated in 2005 a (still on-going) collaboration with Roderich Süssmuth at the Technical University of Berlin (TUB), a specialist in structural characterization of natural products, who was interested in supporting us with MS and NMR spectroscopy analyses.

Over the years, we produced several hundred liters of *Xvesalb* cultures that were processed to purify sufficient amounts of albicidin for MS and NMR spectroscopy studies. Concentrated methanolic extracts of the XAD-7 Amberlite resin taken out from 5-day old *Xvesalb* cultures were loaded onto a polymeric column (PRP-1, 300 x 7.5 mm, 5 µm, Hamilton, Bonaduz, Switzerland) and were submitted to separation by HPLC (Agilent 1200 series equipped with a diode-array detector). Chromatography was performed under isocratic conditions (44% tetrahydrofuran in water, 1% acetic acid; 1 ml/min, detection at λ 308 nm corresponding to the maximum of absorption of albicidin). All collected fractions were tested for their antibacterial activity. Several zones of the

chromatogram were positive (corresponding to similar zones previously observed by Birch and colleagues). Through a second round of similar chromatographical conditions, I purified the main active peak, corresponding to β-albicidin. Still, the purification yield was very low, and only after years, I managed to isolate 3 mg of β-albicidin that were dedicated to structure elucidation at the Süssmuth laboratory in Berlin. Encouraging results were obtained after MS and MS/MS experiments. The molecular mass of β-albicidin was confirmed to be $[M+H]^+$ = 843.2621 Da. Unfortunately, even when assuming that the molecule contains only C, H, O and N atoms, several molecular formulae still fit with this specific mass. We therefore started back from scratch and produced several dozens of liters of *Xvesalb* culture grown in ^{15}N-containing minimal medium. ^{15}N is a non-radioactive isotope of nitrogen and, when incorporated into a molecule, it adds to the molecular mass 1 Da per nitrogen atom compared to the unlabeled molecule. When analyzed in MS, the purified ^{15}N-labeled β-albicidin showed an addition of 6 Da compared to the unlabeled albicidin, indicating that it contains six nitrogen atoms (figure 11).



The MS/MS comparison of ^{15}N- *vs* unlabeled albicidin provided additional data about mass and nitrogen content of the monomer moieties of the molecule (figure 12). Albicidin was also analyzed with various techniques of NMR spectroscopy (COSY, HMBC, NOESY, etc.). All together, data obtained from the *in silico* analysis put together with data resulting from the various MS and NMR spectroscopy analyses of the pure molecule allowed us to decipher the structure of albicidin and to unravel its specific and never-seen-before features (figure 12) (Cociancich *et al.,* 2015).

Figure 11. Overlay of HR-ESI-MS full scans of ^{15}N-labeled (red signal) *vs* unlabeled (black signal) β-albicidin. Note the mass shift of +6 Da for the ^{15}N-labeled molecule (taken from Cociancich *et al.,* 2015).

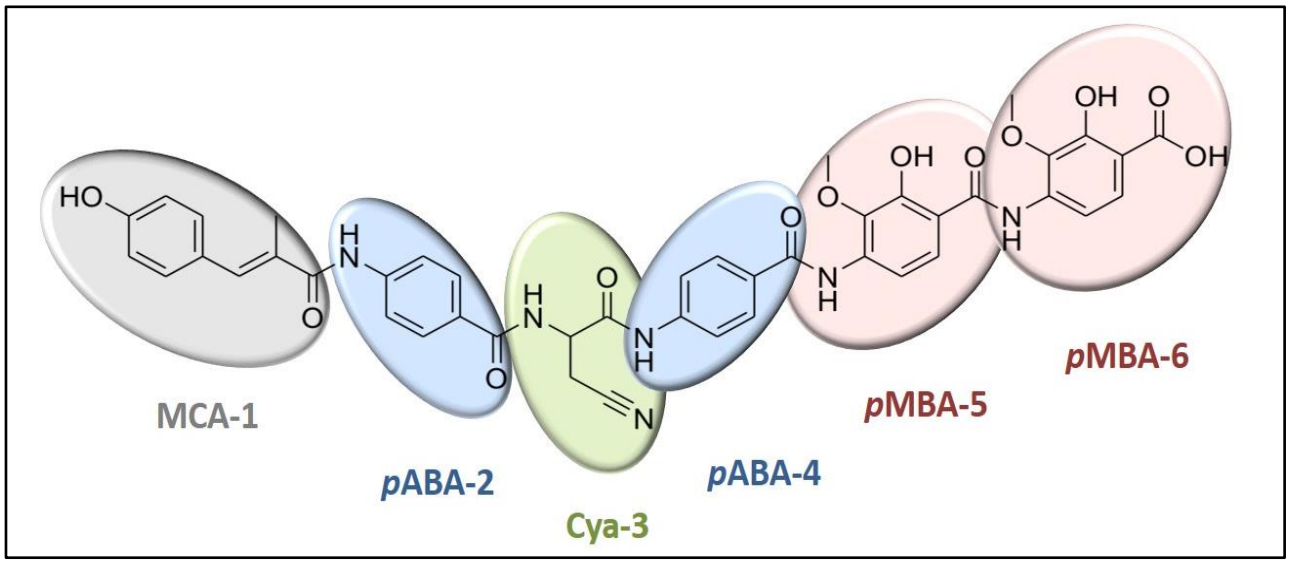


Figure 12. Structure of β-albicidin. The six building blocks linked together by amide bonds are color-coded as follows: grey: derivative of *p*-coumaric acid; blue: *p*-aminobenzoic acid; green: cyanoalanine; pink: hydroxylated and methoxylated derivative of *p*-aminobenzoic acid (adapted from Cociancich *et al.,* 2017).

The hybrid PKS/NRPS albicidin is composed of six building blocks linked together by five peptidic (amide) bonds. The N-terminal block is synthesized by the PKS part of biosynthetic machinery and is a derivative of

para-coumaric acid (MCA-1, see figure 12). Following is a first *para*-aminobenzoic acid (*p*ABA-2) moiety linked to a central L-cyanoalanine (Cya-3) as determined by chiral GC-MS, followed by a second *p*ABA residue (*p*ABA-4). Finally, the C-terminal part of the molecule contains two contiguous hydroxylated and methoxylated derivatives of *p*ABA (4-amino 2-hydroxy 3-methoxy benzoic acid [*p*MBA-5/*p*MBA-6]) (figure 12) (Cociancich *et al.,* 2015). *p*ABA-2 and *p*ABA-4 correspond therefore to the unknown substrate X1 proposed previously (Royer *et al.,* 2004), whereas *p*MBA-5 and *p*MBA-6 correspond to the unknown substrate X2. Notably, as it was speculated, both the X1 and X2 substrates are indeed aromatic moieties. It is interesting to note that albicidin is the first example of NRP containing *p*ABA or *p*ABA-derivatives.

Yet another unique feature of albicidin is the presence of the central cyanoalanine. According to the colinearity rule, the corresponding module within the assembly line is module NRPS-2 from gene *alb01* (figure 13A). *[Note: figure 13 (actual model for biosynthesis of albicidin) and figure 10 (theoretical model from 2004) show similarly organized assembly lines. However, the numbering of the modules is different, and genes carry now arabic numbers instead of roman. The text referring to each figure uses the respective nomenclature]. In silico* analyses predicted the A domain of module NRPS-2 to be inactive due to deletions in the corresponding gene (Royer *et al.,* 2004). We therefore proposed that, as is the case for the biosynthesis of the antibiotic bleomycin (Du *et al.,* 2000), NRPS-2 module is *trans*-complemented by the stand-alone module NRPS-2* encoded by gene *alb04*. The A domain of module NRPS-2* is predicted to incorporate asparagine. Though, the presence within this domain of a 342-aa insertion containing an ATP binding site and resembling a domain found in αANH enzymes led us to propose that this unknown domain, hereafter referred to as αANH-like, is responsible for the conversion of asparagine into cyanoalanine by a phosphorylation/dephosphorylation reaction and the subsequent elimination of a molecule of water (figure 13B).

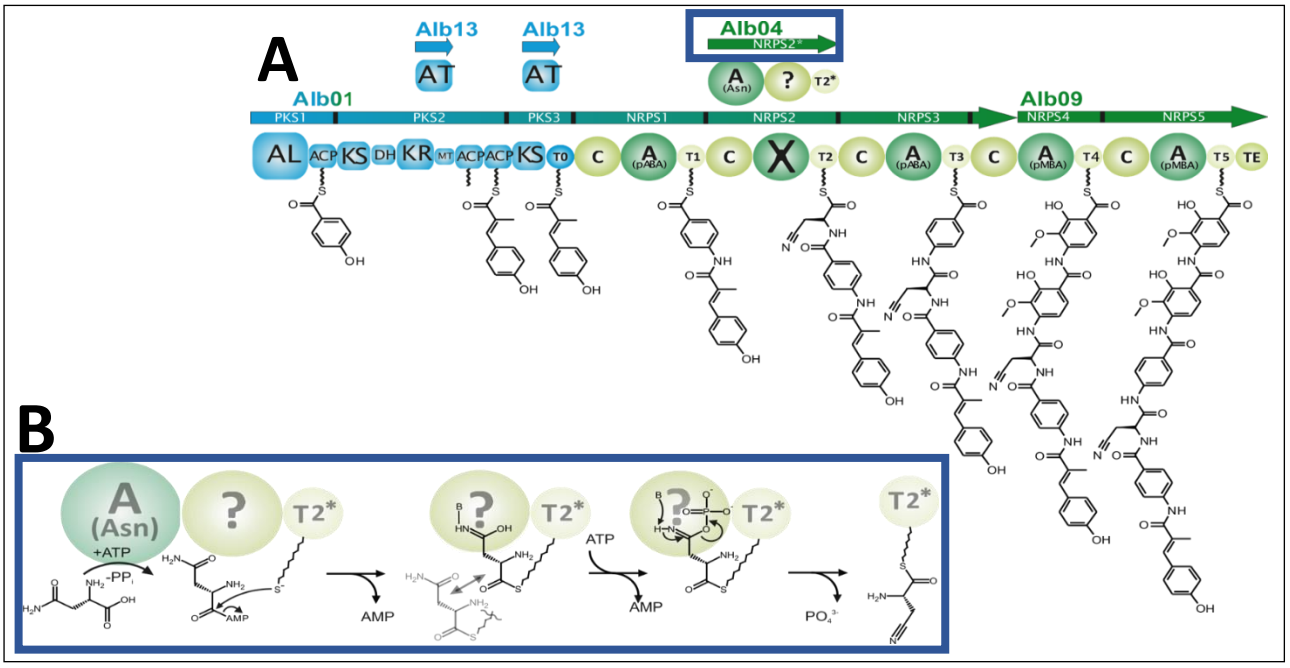


Figure 13. Proposed biosynthetic assembly line for β-albicidin. A: PKS and NRPS modules are color-coded in blue and green, respectively. Numbering of modules and of genes has been slightly modified compared to the initial numbering in Royer *et al.,* 2004. The PKS-NRPS hybrid consists of the following domains: acyl-CoA ligase (AL), acyl carrier protein (ACP), trans-acting acyltransferase (AT), ketoreductase (KR), ketosynthase (KS), dehydrogenase (DH), methyltransferase (MT), peptide carrier protein or thiolation (T), condensation (C), adenylation (A) as well as thioesterase (TE) domains. Substrates of the NRPS are indicated at the A-domains.? stands for the 342-aa insertion domain in the *trans*-acting module NRPS-2*. B: Hypothetical mechanism for the asparagine to cyanoalanine conversion involving the additional αANH-like domain within module NRPS-2* (taken from Cociancich *et al.,* 2015).

Radioactive ATP-PPi exchange assays allowing to figure out the substrate specificity of modules confirmed that module NRPS-2 is unable to activate any of the tested substrates (including L- and D-cyanoalanine, L- and D-asparagine, and L-aspartic acid). On the other hand, module NRPS-2* was shown to preferentially incorporate L-asparagine twice as efficiently as L-cyanoalanine. The efficiency of incorporation of the other substrates was much lower (Cociancich *et al.,* 2015; supplemental data).

II.5. Isolation and characterization of structural analogues of β-albicidin

Although it took us years to obtain the full structure of albicidin, this was however only the starting point for further charcterization of the molecule. In frame of two internships of a BTS student (Julie Arasté), we isolated by HPLC several other natural analogues of albicidin, since I previously detected additional smaller and less resoluted peaks showing antibacterial activity. We implemented additional steps of purification in reversed-phase HPLC to bring to purity small amounts of three analogues of albicidin. Their analysis in MS yielded masses as $[M+H]^+$ = 861.27 Da, 891.27 Da and 886.27 Da, respectively. Meanwhile, besides these three additional albicidin natural analogues, a protocole involving MS-networking allowed to visualize in extracts from both the *Xvesalb* heterologous producer and the *X. albilineans* natural producer seven additional albicidin derivatives. MS^n analyses revealed many structural modifications compared to the structure of β-albicidin (figure 14) (von Eckardstein *et al.,* 2017). One of these modifications is a N-terminal carbamoylation, which is compatible with the presence of a carbamoyl-transferase encoded within the albcidin BGC (Petras *et al.,* 2016). Compared to the reference structure for β-albicidin, other structural variations include one or several modifications such as N-terminal carbamoylation, N-terminal truncation, N-terminal methylation, dehydroxylation of either the *p*MBA5 or *p*MBA-6 building blocks, methoxylation of the cyanoalanine or replacement of the cyanoalanine by an asparagine (figure 14) (Petras *et al.,* 2016; von Eckardstein *et al.,* 2017). A previously identified natural albicidin derivative, which corresponds to a β-albicidin bearing an asparagine instead of a cyanoalanine ($[M+H]^+$ = 861.27 Da) which I previously isolated by HPLC has not been recovered in this MS-driven analysis, and is therefore not represented on figure 14.

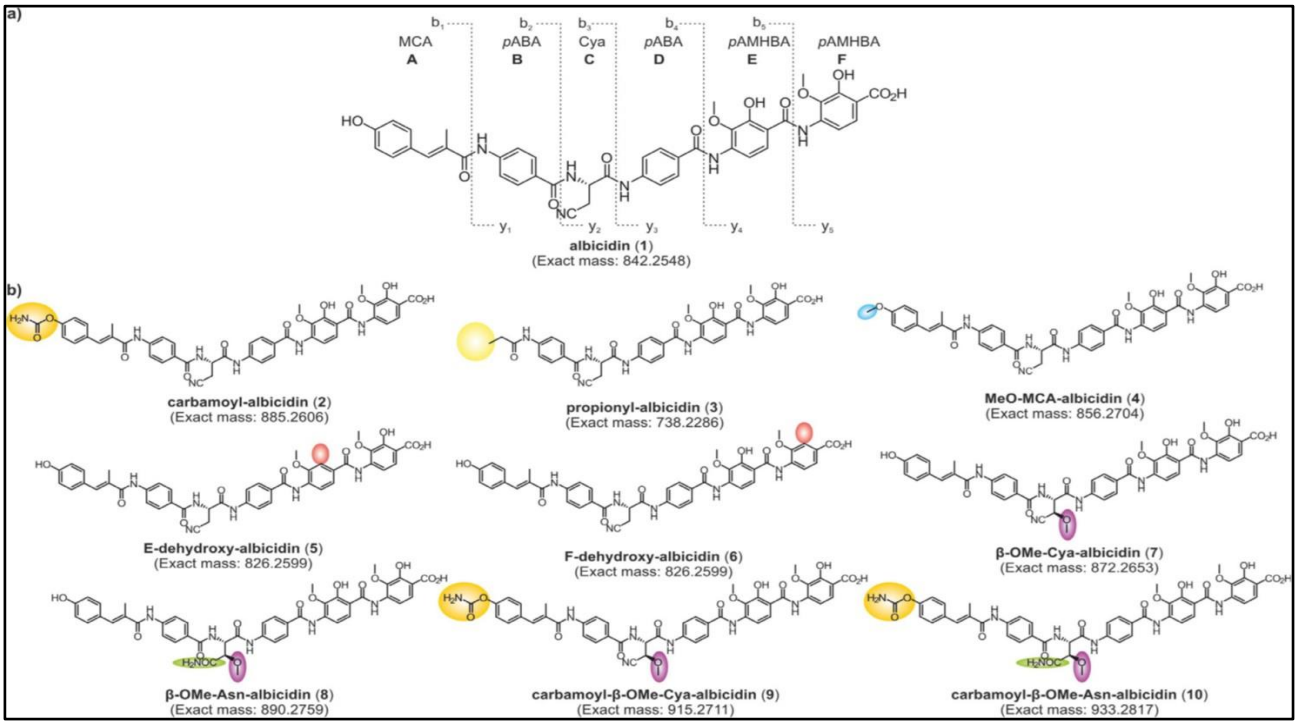


Figure 14. Structures of β-albicidin and derivatives from *X. albilineans* and *Xvesalb*. Modifications from the β-albicidin structure are color-coded as follows: carbamoylation (orange), truncated N-terminal building block (yellow), dehydroxylation (red), methylation (blue), methoxylation (purple), and nitrile to amide exchange conversion (green). Actual nomenclature and exact mass of each albicidin derivative are indicated (taken from von Eckardstein *et al.,* 2017).

Another series of experiments concerns the study of the albicidin detoxifying enzyme AlbD, isolated from *P. dispersa*, and initially described as an esterase (Zhang & Birch, 1997). This enzyme is able to abolish within minutes the antibacterial activity of albicidin. We showed that it cleaves albicidin between the *p*ABA-4 and *p*MBA-5 blocks as illustrated by the molecular masses of the respective yielded fragments ($[M-H]^-$ = 511.17 Da and 347.10 Da). Hence, the AlbD "esterase" is indeed an endopeptidase. Since it is the first peptidase described as being able to cleave an amide bond between two δ-aas, we proposed AlbD to be reclassified as the type example of a novel sub-family of the S9 family of peptidases (Vieweg *et al.,* 2015).

II.6. Conclusion

The complex and unique structure of albicidin has finally been deciphered, together with ten natural derivatives, more than three decades after its discovery. The success of this work holds within the multiple approaches required to reconstruct the puzzle. Many useful information came from the *in silico* analysis of the BGC, revealing the non-ribosomal nature of the molecule, and yielding predictions regarding the respective elementary building blocks.

The construction of the over-producing heterologous host was also a giant leap towards success. It allowed the isolation of the compound (and its isotopic-labeled counterpart) whose analysis in MS and NMR spectroscopy rendered enough data to unravel the structure of the molecule.

In conclusion, within this project, we deciphered the structures of β-albicidin and of about ten derivatives. The PKS/NRPS biosynthetic assembly line and the associated clustered genes account for most of the original structural features of albicidins, although some biosynthetic mechanisms remain unknown, such as the hydroxylation mechanisms. The albicidin molecule stands now as a lead candidate for drug development and, currently, hundreads of synthetic derivatives are being evaluated by our german colleagues.

The originality of its structure, its unique mode of action and its potent antibacterial activity against bacteria responsible for nosocomial infections make albicidin a potential lead structure for drug development. The Süssmuth team developed a chemical total synthesis protocole for albicidin, allowing to get larger quantities of the molecule for biological activity testing, as well as to generate analogues to be tested for improved properties (solubility, biological activity, etc.) (Kretz *et al.,* 2015). The german colleagues could, for instance, already show that the replacement of the cyanoalanine by a threonine lowers by half the IC_{50} for DNA gyrase inhibition (Grätz *et al.,* 2016). The protocole of chemical synthesis of albicidin has been patented (Süssmuth *et al.,* 2014).

III. How genetics can help to characterize the structure of the quorum sensing signal VFM of the plant pathogenic bacteria of the genus *Dickeya*?

Nicole Cotte-Pattat, a colleague from UMR MAP (Microbiologie, Adaptation and Pathogénie) in Lyon solicited us to build a collaborative project on the study of a new quorum-sensing (QS) system, named Vfm, that her team discovered a few years ago in their model bacterium, the plant pathogenic bacterium *Dickeya dadantii*. Indeed, the QS signal has been shown to be synthesized *via* the non-ribosomal pathway and has, so far, resisted to isolation and structural characterization attempts for several years. The whole *vfm* BGC has been identified and our main objective is to perform an *in silico* analysis of the non-ribosomal biosynthetic genes in order to get predictive information towards the structure of the molecule. With respect to our expertise on non-ribosomally-synthesized compounds, Nicole Cotte-Pattat proposed us to imagine a project aiming at describing this new QS signaling molecule in order to develop new control strategies targeting the QS system in these virulent bacteria. With the help of our colleague Roderich Süssmuth in Berlin, and together with the MAP team in Lyon, we co-constructed the following project submitted to the 2018 ANR/DFG franch-german call (including a grant for a future PhD student supervised by me), and for which I am the French coordinator.

III.1. State of the art

The bacteria belonging to the genus *Dickeya* (formerly *Erwinia chrysanthemi*) are listed among the top ten most important plant bacterial pathogens (Mansfield *et al.,* 2012). These pectinolytic bacteria are the causative agents of soft rot diseases worldwide, affecting numerous plant species of agronomical importance including crops and vegetables (e.g. potato, maize, rice, chicory, tomato, carrot, banana tree, pineapple, sugarcane) and several ornamental plants. The occurrence of *Dickeya* species (*D. dadantii, D. paradisiaca* and others) was previously associated with tropical regions (or greenhouses in temperate regions) but other species, e.g. *D. dianthicola, D. chrysanthemi,* appear to be better adapted to temperate climates (Adeolu *et al.,* 2016). Recent alarming damages of potato fields in Europe have been attributed to what is considered a highly virulent emerging pathogen, *D. solani* (Toth *et al.,* 2011).

Eventhough other determinants are involved, the virulence of *Dickeya* species mainly correlates with their high capacity of production and secretion of plant cell-wall degrading enzymes (PCWDEs), with a major role of pectate lyases (Hugouvieux-Cotte-Pattat *et al.*, 1996; reviewed in Hugouvieux-Cotte-Pattat, 2016). These enzymes are responsible for the maceration of plant tissues, a symptom observed for *Dickeya* infections.

As for all bacteria, the *Dickeya* virulence is based on quorum sensing (QS), a finely-tuned communication system which allows bacteria from the same species to regulate gene expression as a result of the perception of changes in bacterial cell density. QS-based communication depends on small diffusible molecules which act as signals and which can be perceived through sensors by other bacteria. QS is widely distributed in the bacterial world (figure 15). Interestingly, disrupting QS was proposed as a mean to fight animal but also plant diseases (reviewed in Helman & Chernin, 2015, and in Grandclément *et al.,* 2016). For most of bacteria, the QS systems rely on signaling molecules which are, depending on the bacterial species, derivatives of homoserine lactone (HSL), with the notable exception of *Xanthomonas* bacteria which rely on DSF, a fatty acid signaling molecule, for their QS-based virulence (for review, see Grandclément *et al.,* 2016).

Since decades, *D. dadantii 3937* has been the most studied strain of the genus *Dickeya*. Its arsenal of PCWDEs, as well as the regulators controlling their production, has been fully characterized (Reverchon & Nasser, 2013). *D. dadantii* utilizes a complex regulatory network that is able to rapidly respond to environmental stimuli in order to start infection at an appropriate point of time (Charkowski *et al.,* 2012, Hugouvieux-Cotte-Pattat *et al.,* 1996; Reverchon *et al.,* 2016). So far, two QS mechanisms have been described in *D. dadantii* 3937. The first system, termed Exp, is a classic system dependent on homoserine lactone synthesis (Reverchon *et al.,* 1998). This HSL-dependent QS system is widespread among Gram-negative bacteria (see figure 15). The second is the QS system Vfm (Virulence factor modulating),

discovered by the UMR MAP (Nasser *et al.,* 2013). Both systems Exp and Vfm are present in all sequenced species of *Dickeya*, except for *D. paradisiaca* which possesses only the vfm locus. The interplay between the two QS systems was recently studied in *D. solani* (Potrykus *et al.,* 2017). Most of the analysed *D. solani* vfm or exp mutants showed a significant attenuation of plant maceration and decreased PCWDE production as well as a restricted motility. This indicates that both VFM and HSL signals are involved in the regulation of *D. solani* virulence. However, the inactivation of *vfm* genes has a preponderant effect on the expression of the virulence genes, confirming the main role of the Vfm QS system in the pathogenicity of *D. solani* (Potrykus *et al.,* 2017).

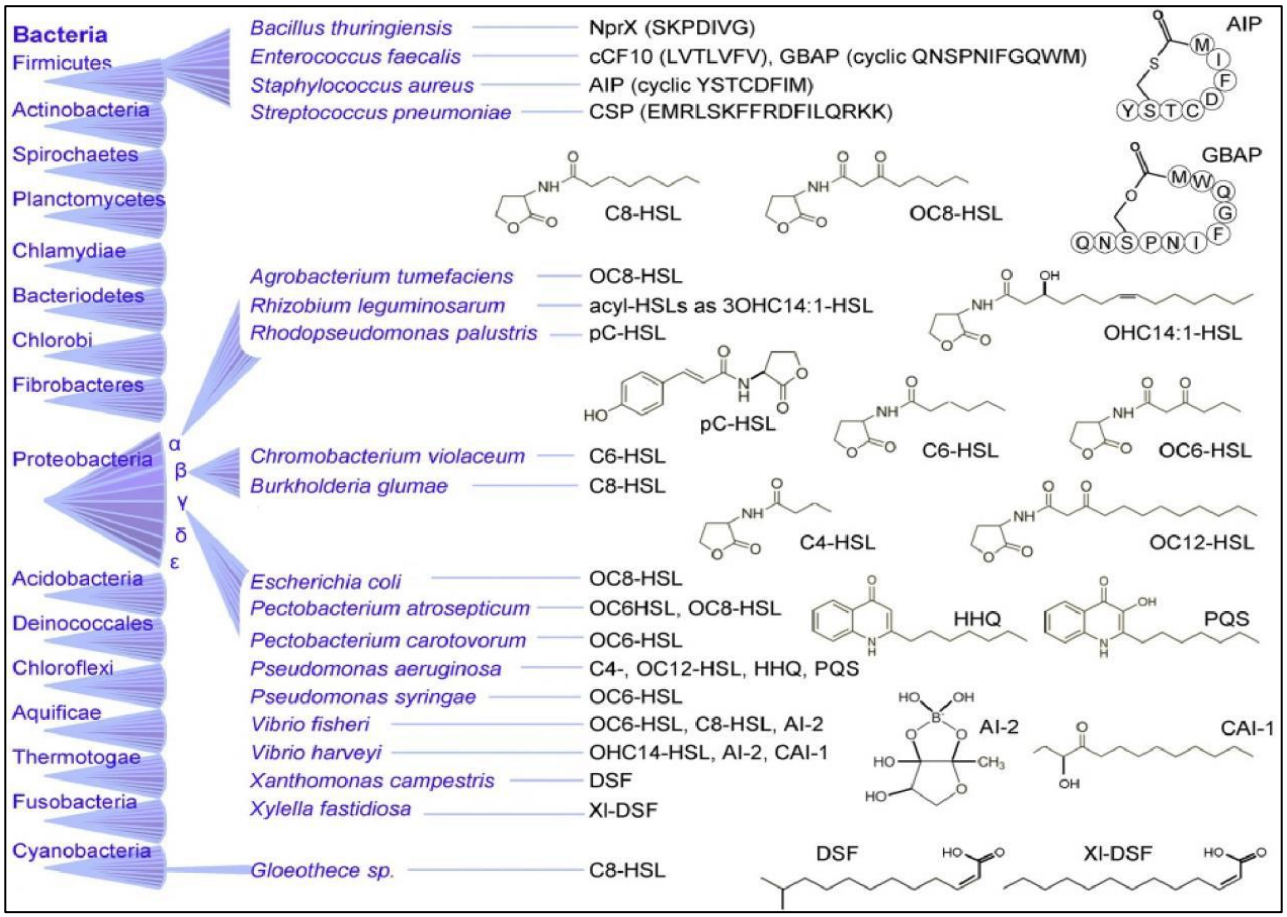


Figure 15. Diversity of QS signals in various clades of bacteria (taken from Grandclément *et al.,* 2016).

So far, the extracellular signal of the Vfm QS system has escaped detection by analytical chemistry methods. However, by using a biological assay, the locus *vfm*, encoding the QS system, has been experimentally characterized in the strain *D. dadantii* 3937 (Nasser *et al.,* 2013). This locus does not share any feature with already known QS loci, indicating that the signal VFM significantly differs from all already known QS signals. The locus *vfm* contains 26 genes, annotated as *vfmA* to *vfmZ*. It includes genes involved in the perception of and response to the extracellular signal, as well as several biosynthesis genes coding for the synthesis of a putative small molecule acting as the signaling factor (figure 16). The transfer of this locus to *E. coli* led to the heterologous production of the signal VFM, indicating that it contains all the *Dickeya* genes required for the biosynthesis of the signal (Nasser *et al.,* 2013). In this context, it is important to note that the heterologous production of the signal VFM has been detected solely on the basis of its biological effect. The perception of the signal is achieved by a two-component system encoded by genes *vfmH* and *vfmI*, which activates the expression of the AraC regulator VfmE which will, in turn, activate the transcription of PCWDE genes, as well as its own transcription (auto-induction mechanism) (Nasser *et al.,* 2013).

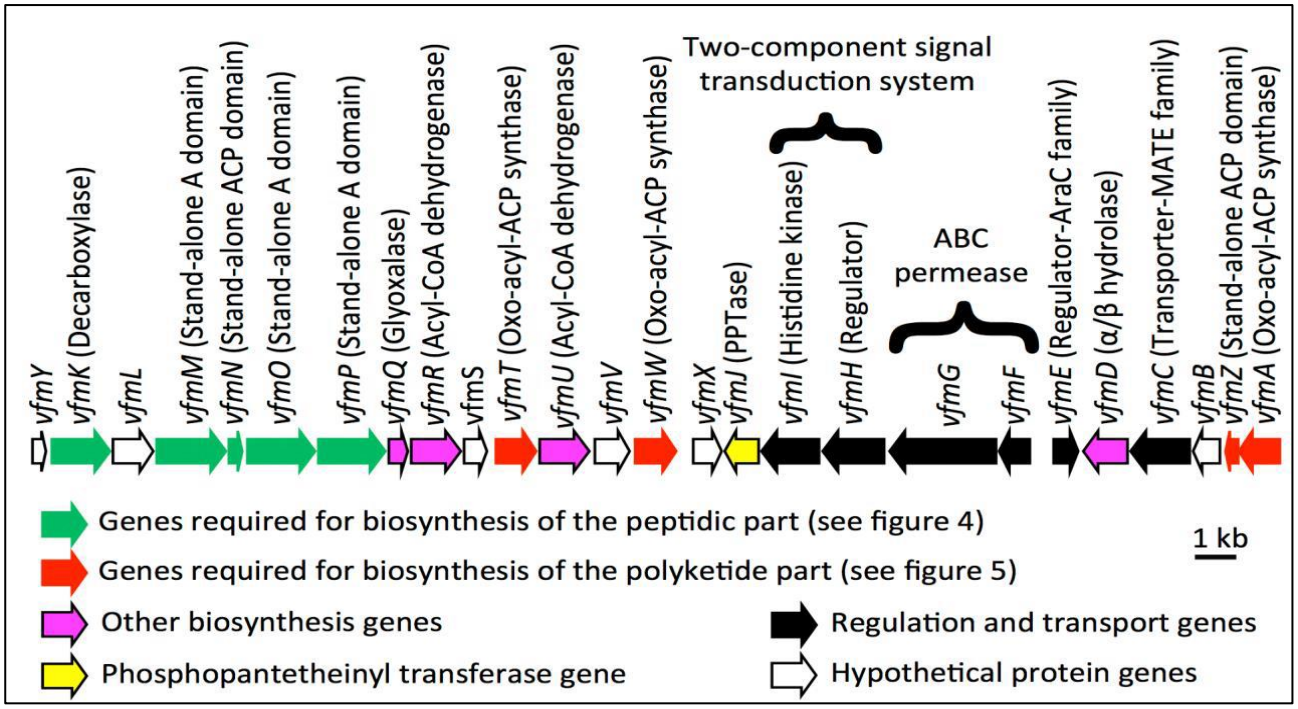


Figure 16. Description of the vfm locus of *D. dadantii* 3937. Genes are represented by arrows which are colored according to the legend. Respective names of the genes and their putative functions are also provided. A: adenylation; ACP: acyl carrier protein; ABC: ATP binding cassette. For detailed information, see Nasser *et al.,* 2013.

III.2. Preliminary results

With regard to the biosynthesis of the putative signaling molecule VFM, the locus *vfm* encodes proteins sharing features of NRPSs and PKSs. In a preliminary work, we performed an extensive *in silico* analysis of the biosynthesis genes present within the *vfm* BCG. This analysis allowed us to predict that the final VFM molecule should be synthesized as an hybrid PK/NRP. We observed very unusual domain arrangements in these NRPSs and PKSs, with a remarkable absence of C and KR domains. However, we could predict some structural data for the VFM signal compound which confirm its novelty. However, due to the competitive environment on this topic, I consider these predictive data and the corresponding biosynthetic model as confidential, and therefore I will not describe them in this document.

Interestingly, an *in silico* analysis of the biosynthesis genes from the *vfm* BCG in 49 genomes of *Dickeya* bacteria available in GenBank revealed a strain-specific polymorphism in the NRPS substrate specificity-conferring signature sequences, suggesting that several analogues of VFM can be produced by various strains.

Using a protocole similar to the one successfully employed for albicidin, we succeeded in extracting the VFM signal from supernatants of the heterologous host *E. coli* containing the complete *vfm* locus of *D. dadantii*, which has never been achieved before by our colleagues of UMR MAP despite a two-year collaboration with chemists specialized in isolation of QS signals. Nonetheless, our first tentative isolation in HPLC did not yield a specific peak compared to the similarly treated control *E. coli* extracted supernatant, indicating a probable low yied of production. Still, some VFM biological ativity could be detected in several contiguous HPLC fractions covering a large zone of the chromatogram, suggesting that either the chemistry of the column and/or the separation conditions are not optimal, or that we observe, as is the case for albicidin, the presence of multiple analogues. In a different approach, we sent to our chemist colleagues in Berlin two batches of XAD-resin which have been used to extract approx 200 liters of both the VFM-producer *E. coli* strain and the control non-producing strain. The analyses of the corresponding extracts are currently taking place.

III.3. Proposed strategies to over-produce, isolate and characterize the VFM signal molecule

Considering the low yield of production, we anticipate the need to over-produce the QS signal in this strain. In a first approach, we propose to over-express the transcriptional activator VfmE of the AraC family to up-regulate the expression of all *vfm* biosynthesis genes. We already used successfully this approach in another of our projects (see project "Meta-B"). This work is schedulded to be performed in frame of the master2 of Agathe Ropars. In parallel, we will check whether cell-lysates contain the QS signal and thus constitute a viable source for isolation. The yield of production of the QS signal will be estimated on the basis of its biological activity.

Regarding further VFM isolation and structural characterization, we plan to basically stick to a similar approach than the one we developed for albicidin. Briefly, supernatants from large culture volumes of the yet-to-be constructed VFM-overproducing heterologous host will be treated using Amberlite XAD-7 resin prior to chromatographic separation using various HPLC techniques (predominantly size-exclusion and reversed-phase). In parallel, the control *E. coli* strain which does not carry the *D. dadantii vfm* cluster will be submitted to a similar procedure and thus facilitate a differential comparison. Bioactive fractions will be identified using the *D. dadantii* reporter strain (Nasser *et al.,* 2013). Depending on the purity of positively tested HPLC fractions, additional chromatographic steps may be necessary. Technically-wise, all purification procedures will be greatly facilitated by employing minimal medium for bacterial cultivation. Such minimal medium containing either casamino acids or a defined set of amino acids has already been optimized for the production of the VFM signal of *D. dadantii* 3937 from the heterologous host. Subsequently, optimized minimal media will be complemented with ^{15}N-labeled and ^{13}C-labeled media components for isotopic labeling of VFM. This strategy has been successfull for the establishment of the structure of the antibiotic albicidin (Cociancich *et al.,* 2015; for ^{13}C, not published). Such labeling procedure is expected to be a bottleneck-breaker to facilitate the structural characterization of the VFM signal.

Once purified, the VFM compound will be anlysed through a battery of MS- and NMR-derived techniques which will be implemented by our german chemist partners in Berlin.

III.4. Conclusion

The *in silico* analyses and preliminary experimental data indicate several unique and specific features of the Vfm QS system, *i.e.* the biosynthetic NRPS-PKS pathway is highly unusual, and as a result, the structure of the signal differs greatly from all already known QS signals.

The objective of the project is to elucidate the structure of different analogues of the Vfm QS system from *Dickeya* strains, including the model strain *D. dadantii* 3937 and two *D. solani* strains. The heterologous host *E. coli* containing the locus *vfm* of *D. dadantii* 3937 will facilitate the overexpression of the biosynthesis genes in order to improve the yield of production of the signal VFM by this strain and to facilitate the isolation task.

A basic understanding of the structure and biosynthesis of VFM, a novel QS signal in bacterial world, is expected from this project. This involves a biosynthetic assignment of enzyme functions. Since the targeted microorganisms are agronomically important pathogens, this project will *de facto* contribute to a deeper understanding of *Dickeya*'s pathogenicity. In a later and more exploratory stage, structural data on the VFM signal(s) will be used to conceptualize, design and chemically synthesize putative VFM binding competitors or to identify a bacterium secreting a VFM modifying enzyme (quorum quenching, this bacterium could be a candidate for biocontrol). As a long term output, the fight against *Dickeya* pathogens should be facilitated courtesy of the data arising from the current project.

Chapter 2
Contribution of genome mining to the discovery of new natural products

I. The genome mining, a different approach to discover new microbial secondary metabolites

I.1. Generalities

For ages before the advent of the genomic era, biologically- or therapeutically-relevant microbial natural products have been identified through a "classical" approach, *i.e.* by monitoring their biological activity, eventually followed by their extraction, fractionation, isolation as pure compounds and finally structural characterization (figure 17). Biological screenings resulted in the discovery of thousands of antibiotics, antifungal agents, antiproliferative molecules etc. produced by microorganisms. But towards the end of the 20th century, the number of new interesting natural products discovered on the basis of their biological activity waned because of multiple re-discoveries of compounds and the development of chemical (semi-)synthetic libraries of molecules. Hopes for detecting new secondary metabolites from their biological activity were turned down and the feeling that the source of new natural products was about to dry up was predominant.

However, the genomic sequencing era brought back enthusiasm within the research community as the newly-sequenced genomes of microorganisms started to reveal their secrets (figure 17). As an example, when the first-ever *Streptomyces* genome was sequenced and annotated more than a decade ago (Bentley *et al.,* 2002), over 20 BGCs potentially encoding secondary metabolites have been found, corresponding for most of them to previously uncharacterized BGCs. As of today, given the access to millions of genomic sequences and thousands of complete microbial genomes, it is a clear picture that the potential of microorganisms for synthesis of natural products has been under-estimated by several orders of magnitude.

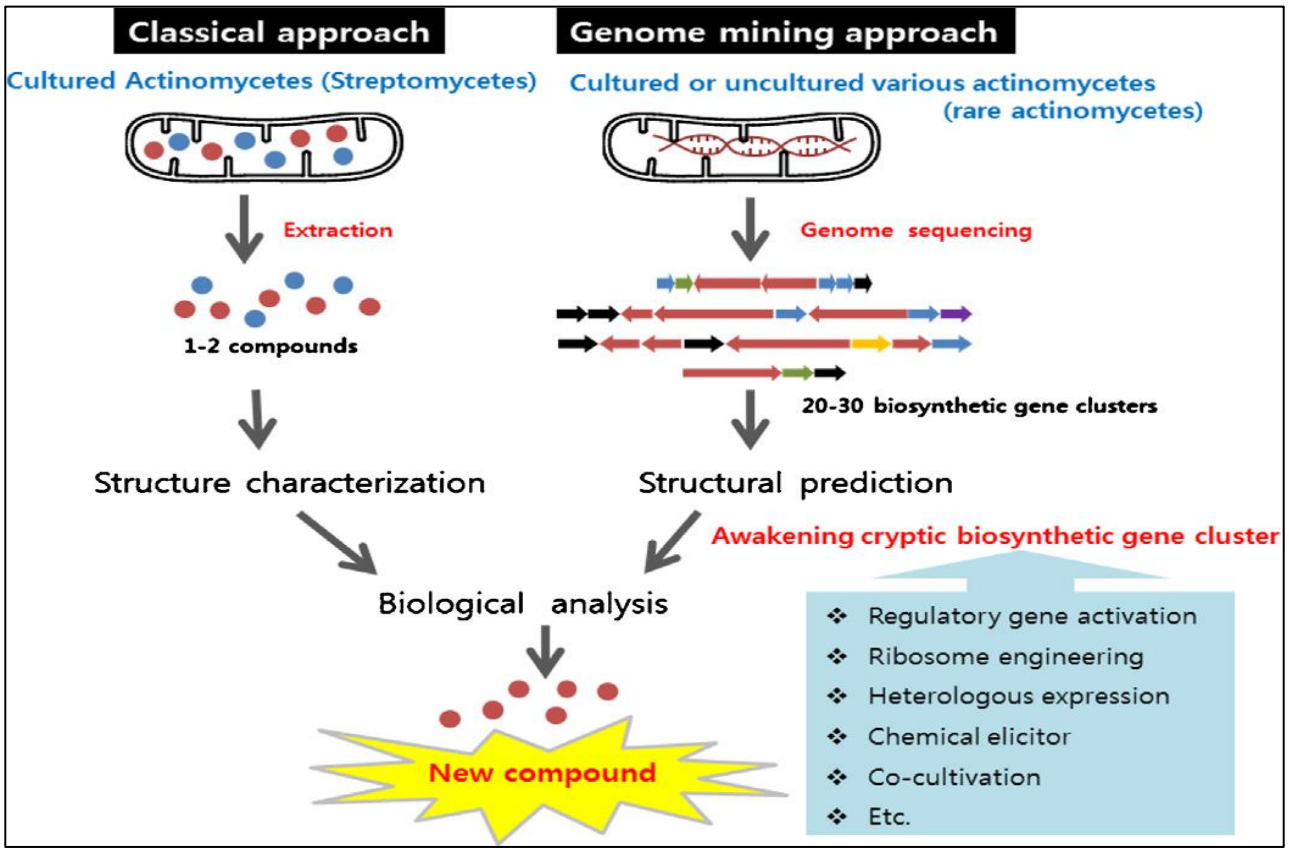


Figure 17: Comparaison of the classical approach based on biological activity screening with the genome-mining approach for the detection of natural products. Taken from Choi *et al.,* 2015.

Generally, a BGC minimally encodes the enzymes, regulators and transporter required for the biosynthesis, processing and export of a specific secondary metabolite. This feature is the landmark targeted to mine genomes for BGCs, as is the peculiar repeated genomic structure of NRPS- or PKS-encoding genes, for instance. Although the mining of NRPS- or PKS-encoding BGCs might sometimes be impaired by errors in the genome itself due to the high number of repeats resulting from the modular architecture of these genes, this problem might however be overcome using third-generation sequencing technologies, such as those developed by Pacific Bioscience and Oxford Nanopore which generate long-reads sequencing data.

To date, genome mining of genomic sequences available in databases is an efficient technique that fulfills the resurgent interest of the research community for natural products. Many automated bio-informatic resources dedicated to scan genomic sequences for specific patterns of BGCs encoding putative secondary metabolites have been developed and provided to researchers (for comprehensive lists and description of genome mining tools, see Medema & Fischbach, 2015; Kim *et al.,* 2017). The most popular resource is antiSMASH (antibiotics and Secondary Metabolites Analysis Shell; antismash.secondarymetabolites.org). It accepts FASTA- or GenBank-formatted genomic sequences as inputs and, in turn, identifies, annotates and analyzes secondary metabolites BGCs. Several improved versions of this resource have been regularly provided, with version 4.0 recently out (Medema *et al.,* 2011; Blin *et al.,* 2017). If focusing more specifically on NRP-type secondary metabolites, NRPSpredictor2 (nrps.informatik.uni-tuebingen.de) stands as a very powerful tool to predict substrate specificity for each detected NRPS module within the BGC, provided the corresponding substrate-specificity conferring signature has previously been attributed to a substrate (Röttig *et al.,* 2011). Additional useful informatic resources are provided as databases for BGCs (e.g. MIBiG [Minimum Information about a Biosynthetic Gene cluster; mibig.secondarymetabolites.org] [Medema *et al.,* 2015]) and for secondary metabolites or fragments of, such as building blocks (e.g. StreptomeDB [pharmaceutical-bioinformatics.de/streptomedb/] [Klementz *et al.,* 2015], NORINE [bioinfo.lifl.fr/norine/] [Pupin *et al.,* 2016]). Interestingly, many of these resources are interconnected, as shown by antiSMASH which compares newly identified BGCs with experimentally-described BGCs stored in the MIBiG repository (Medema, 2018).

In addition to the classical approach based on biological activities and to the genome mining computational approaches, it is interesting to note the recent development of analytical techniques based on mass spectrometry (LC/MS-MS and MALDI-imaging) that establish the metabolic profile of whole extracts. The handful load of data provided are subsequently analyzed by dereplication bio-informatic tools and databases which sort information concerning already known metabolites *vs* new compounds.

All informatic resources which allow computational approaches to discover and identify natural products face now the challenge to handle the millions of data expected to be generated by pangenomic and metagenomic initiatives, with a global potential of over 10^9 natural compounds remaining to be discovered (Stewart, 2012).

I.2. How to awaken silent or poorly-expressed BGCs

As mentioned above, the mining of thousands of entire microbial genomes reveals the huge potential to discover new structural classes of secondary metabolites just by “reading” the genome, preceding the investigation of the corresponding unknown compounds. New papers showing the unexpected metabolic potential of microbes whose genomes have been mined are published almost every second day (for a recent example, see Jackson *et al.,* 2018).

The enthusiasm generated by the unraveling of this yet-undetected reservoir of new secondary metabolites is moderated by the apparent lack of production of the corresponding compounds. Although we cannot exclude that some of the *in silico* newly-discovered BGCs encode the biosynthesis of known molecules, it seems that most of them remain, at least in our current laboratory conditions, “silent” or “orphan” or “cryptic” (all three terms sharing basically the same meaning) (Machado *et al.,* 2017). However, “silent” specifically accounts for a

transcriptional inactivity (though it is sometimes difficult to distinguish between a true non-expressed BGC and a BGC whose level of production of the natural product is below the detection threshold). Hence, several strategies have been designed to activate or "awaken" these "silent" BGCs (for extensive reviews, see Choi *et al.,* 2015; Reen *et al.,* 2015; Zarins-Tutt *et al.,* 2016). Most of these strategies have been tested on Actinomycetes (especially in *Streptomyces* strains) since they represent the source of almost half of the 23000 biologically-active natural compounds isolated to date in microorganisms and also the biggest reservoir for still-to-be discovered secondary metabolites.

- **Varying culture conditions**

  Changes in culture conditions, such as the modifications of temperature or of the carbone source, the use of rich *vs* minimal media, or the increase of precursor/substrate supply, result in various arrays of produced secondary metabolites. This illustrates the OSMAC (One Strain, Many Compounds) approach which shows that a single strain can produce an array of structurally-different compounds in response to internal (metabolic state) and external (environment) variations. Moreover, the expression of BGCs eventually varies upon the growth phase: for example, in *Streptomyces coelicolor*, eight out of 22 BGCs are expressed at higher levels after the transition between the exponential and the stationary phases (Huang J *et al.,* 2001). Alternatively, the addition in the culture medium of "elicitors" or "elicitor libraries" (e.g. rare Earth elements or subinhibitory concentrations [<1% of MIC value] of antibiotics lead to the biosynthesis of previously undetected natural products (Rosen & Seyedsayamdost, 2017).

  Co-culture represents another strategy for awakening silent BGCs. In Nature, bacteria live in multi-species complex communities in which take place cross-talks triggering production of specific bio-active compounds. Co-cultures mimic, at least partly, the natural environment of the targeted microorganism where it is more prone to the need of secondary metabolites compared to laboratory conditions. A nice illustration is given by the study, using high-of-the-art MS techniques, of a strain of *S. coelicolor* grown in five independent co-cultures with, respectively, a different Actinomycete. The *S. coelicolor* strain was shown to produce a different portfolio of metabolites in each interaction, with a total of 227 detected natural products and only 50% of them being already known (Traxler *et al.,* 2013).

- **Genetic manipulations**

  BGCs are transcribed from inducible promoters, under the control of regulators, activators and/or repressors. Their expression is therefore finely tuned in response to environmental changes. Therefore, the down-regulation of repressors, the up-regulation of activators or the activation of global regulators (the latter producing pleiotropic effects on numerous secondary metabolites) are some of the means to modify the expression pattern of the targeted BGCs.

  Finally, probably the most popular strategy to influence the expression of "silent" BGCs is the heterologous expression. It helps when genetic manipulation is not possible in some microorganisms (as is the case for some *Streptomyces* strains). It also allows overexpression of the BGC given it is cloned under the control of inducible promoters. This strategy however faces several putative bottlenecks: indeed, one must find heterologous hosts that allow a correct folding of enzymes, and which respond to the same transcription factors as the natural host. Also, especially for PKSs and NRPSs, their BGCs contain large genes that might be dispatched in more than one cosmid for the heterologous transfer.

Other strategies (such as ribosome engineering) have also been used with success to awaken silent BGCs, but the ones described above represent basically the various techniques I use(d) in my projects (see below).

I.3. General strategy based on genome mining used in three present and future projects to discover new natural products

An almost similar strategy has been used or is planned to be used in the projects "Meta-B", "XanthoNRPS" and "CONTACT" that I developed to detect new secondary metabolites of interest in bacteria from the genera *Xanthomonas* and *Streptomyces*, respectively. The "Meta-B" project is already at an advanced stage whereas both other projects are at preliminary or early stages, respectively. Briefly, the sequences of bacterial genomes, either available on GenBank or provided as draft genome sequences by the large-scale sequencing of collections of bacteria performed in frame of these projects are mined for BGCs using bio-informatic tools such as antiSMASH (Blin *et al.,* 2017) for the presence of NRPSs-encoding genes. In most of draft sequences, NRPS-encoding sequences correspond to fragments of NRPS genes, which are distributed in several independent contigs. The search for NRPS domains will be done on these fragmented and incomplete sequences in order to identify interesting domains. Bacterial strains belonging to the genus *Streptomyces*, and to a lesser extent to the genus *Xanthomonas*, are huge reservoirs for BGCs, and in order to take advantage of our knowledge on the albicidin biosynthesis genes, we decided to use original features of these genes as a filter in the mining process. Indeed, we proposed to first mine by BLAST the collections for the presence within their BGCs of domains providing its structural specificity to albicidin, *i.e.* the *p*ABA-specific and the αANH-like-containing NRPS A domains.

Once identified by antiSMASH (for the genes and NRPS/PKS domains organization) and NRPSpredictor2 (for the substrate specificity determination of NRPS A domains [Röttig *et al.,* 2011]), we aim at performing an *in silico* analysis which allows us to predict at least part of the final compound structural features.

Draft sequences exhibiting NRPS domains of interest will be selected for further sequencing with the PacBio technology and further assembly experiments in order to determine the complete finished sequence of the corresponding promising NRPS BGC. According to the quality of the assembly of draft sequences, cloning of the complete PKS/NRPS loci may be necessary to determine their complete sequence. This cloning will also be required for the eventual transfer of the locus into a heterologous host.

As previously stated, many PKS/NRPS BGCs may be silent under standard laboratory conditions and several methods aim to unlock these cryptic biosynthetic pathways. Some of these methods have been explored to overproduce any unknown small molecule synthesized by a new promising NRPS BGC. The first option, which is quite straightforward to apply, in the overexpression of the transcriptional regulator encoded by the BGC. An overexpression of the PPTase gene could also improve the yield of production of the NRP. The transfer of the complete locus into a heterologous host is also planned to induce a significant increase of the production of the NRP. In case heterologous transfer is chosen, we will perform an heterologous transfer in another species of the same genus as the natural host because: (i) there are strong indications that the promoters are only recognized efficiently by the transcriptional machinery of bacteria within the same genus; and (ii) the unknown non-proteinogenic substrates assembled by the NRPS may not be available in other bacterial genera. As a reminder, the transfer of albicidin biosynthesis system into *X. axonopodis* pv. *vesicatoria* led to a significant increase in the production of this antibiotic (Vivien *et al.,* 2007). No approach will be excluded in these projects, including co-cultures or cultivation in a medium mimicking the natural environment of the bacterium possessing the NRPS BGC.

Various cultivation conditions of temperature, shaking speed and culture media composition will be assayed to optimize the production of NRPs. Gene inactivation (KO mutants *vs* WT) or gene activation (overproducer strain *vs* WT) within the BGC will be supported by HPLC and MS in order to detect changes in the metabolites production profiles and to estimate their yield of production. Predictive *in silico* data could guide isolation by providing insights into physico-chemical properties of the NRPS-synthesized molecule(s) and identifying aa precursors. In order to facilitate detection of unknown peptide products, deuterium (^{2}H)- or ^{15}N- or ^{13}C-labeled aa precursors will be fed to *Xanthomonas* or *Streptomyces* cultures, respectively, which will render

characteristic isotope patterns and molecular mass shifts for an enhanced identification of NRPs. The cultivation of *Xanthomonas* strains for production of NRPs will be carried out mainly by our team. We will also provide overproducer strains as heterologous expression hosts, based on previous experience with such systems and perform isolation of the compound.

Pure compounds will be subjected to high-resolving HPLC-ESI-Orbitrap-MS. MS/MS fragmentation, which renders a maximum of structural information, will help to determine the molecular formula. The subsequent work phase contains the investigation by 2D NMR spectroscopy methods (COSY, TOCSY, HSQC, HMBC, NOESY and ROESY). As experienced previously, the GC-MS method is of considerable value in determining the stereochemistry of aas as well as to confirm the identity of aas, fatty acids and other structural constituents. If sufficient material from prolific producers is available, the Süssmuth team will additionally perform X-ray crystallography, which will deliver, next to a conformation also the information on stereocenters present in the molecules of interest. If required, chemical synthesis will aid in structure elucidation, should there be ambiguities on the structure.

As soon as the main structure of one compound has been elucidated, we will assign the biosynthetic function and propose a likely course of the biosynthesis. Gene inactivation mutants which will be generated in this context will be screened by HPLC-ESI-MS for biosynthesis intermediates.

Small molecules expected to be discovered in frame of these projects should potentially be of interest for applications in human and veterinary medicine, agriculture, or manufacturing. Should a molecule exhibiting a new structure be discovered, experimental studies will be conducted to characterize the function of this molecule and to explore its interest for such applications. Exploration of the function of the molecule will cover a wide field of applications: antibiotic, antifungal, antiviral, chemotherapy, plant disease-control, etc. Availability of the structure will make chemists able to synthesize the molecule and to modify it in order to suppress or enhance certain characteristics such as solubility, efficiency or stability and to facilitate a putative application. Should a molecule offering a new putative application be discovered, it will be patented.

II. Characterization of Meta-B, a new lipopeptide discovered by genome mining in several species of *Xanthomonas* [project "Meta-B"]

II.1. *In silico* analysis of the Meta-B BGC in *Xanthomonas* species

The sequencing, annotation and mining of the 3.8 Mb genome of *X. albilineans* strain GPE PC73 provided new information on the potential of this strain to produce NRPs. In addition to the albicidin BGC, five additional loci encoding NRPSs were found in the bacterial chromosome, covering up to 4% of the genome (Pieretti *et al.,* 2009). The *in silico* analysis of these loci indicates that only one (code name Meta-B) contains, in addition to NRPS genes, several genes involved in transport, regulation or biosynthesis of non-proteinogenic aa building blocks (figure 18). According to our bioinformatic analysis of the Meta-B BGC from *X. albilineans* strain GPE PC73, we predicted that it encodes a 16-module NRPS with the characteristic [C-A-PCP] domain arrangement in all modules (Royer *et al.,* 2013). From the study of the A domains sequences, we deduced the nature of 12 out of the 16 specifically-incorporated aa building blocks, the four others corresponding to two unknown signatures present twice. As we identified a C-starter domain and four dual C/E epimerization domains, the Meta-B molecule is therefore likely to be a lipopeptide containing four D-aas. Other genes within the Meta-B BGC encode an AraC transcriptional regulator, an ABC-transporter, a MbtH protein, an hydroxylase, and a gene involved in the biosynthesis of 2,4-diaminobutyric acid (Dab) (figure 18). Finally, five genes involved in the biosynthesis of the non-proteinogenic substrate 3,5-dihydroxyphenylglycine (Dpg) are present in the Meta-B BGC, suggesting that one of the unknown substrate specificity signatures could correspond to Dpg (figure 18) (Royer *et al.,* 2013).

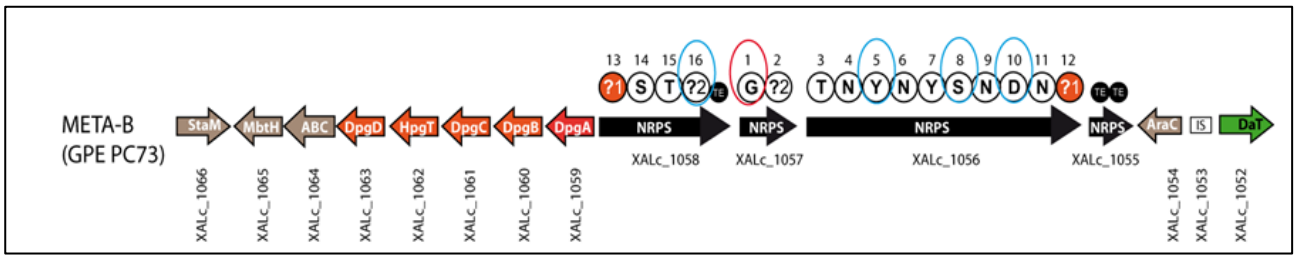


Figure 18. Scheme representation of data yielded from the *in silico* analysis of the Meta-B BGC in *X. albilineans* strain GPE PC73. NRPS-encoding genes are represented by black arrows above which circles, corresponding to the encoded modules, indicate both the substrate specificity (when a known A domain signature sequence was identified) and their respective position within the final assembly line (as shown by the numbers above the circles). ?1 and ?2 symbolize unknown substrates. The modules circled in blue possess a C/E dual condensation/epimerization domain and therefore mediate the epimerization of the substrate incorporated by the upstream module. The module circled in red corresponds to the initiation module and contains a lipo-initiation C-starter domain. Colored arrows represent genes involved in the biosynthesis of building blocks (green: Dab; red: Dpg). Finally, brown arrows are genes encoding the transcription regulator AraC, an ABC transporter, a hydroxylase and a MbtH protein (taken from Sabri, 2016).

Interestingly, unlike the other NRPS BGCs from *X. albilineans*, the Meta-B BGC is present in several strains of two other phylogenetically distant *Xanthomonas* species, namely the rice pathogen *X. oryzae* and the wheat pathogen *X. translucens*, as well as in the strain GPE39 of the *'X. pseudalbilineans'* species, a close relative of *X. albilineans*. Although the general organization of the Meta-B BGC is conserved in all the strains, the Meta-B BGCs from *X. oryzae* and *X. translucens* contain indeed supplementary genes involved either in the biosynthesis of an unknown aa substrate (*i.e.* genes encoding a transaminase, a dehydratase and a hypothetical protein) or in hydroxylation. Nevertheless, the predicted final peptidic sequence of the Meta-B NRP is very different in all three species according to our extensive *in silico* analysis of all the A domain signature sequences (figure 19) (Royer *et al.,* 2013; Sabri, 2016). We therefore decided to investigate further the Meta-B molecule.

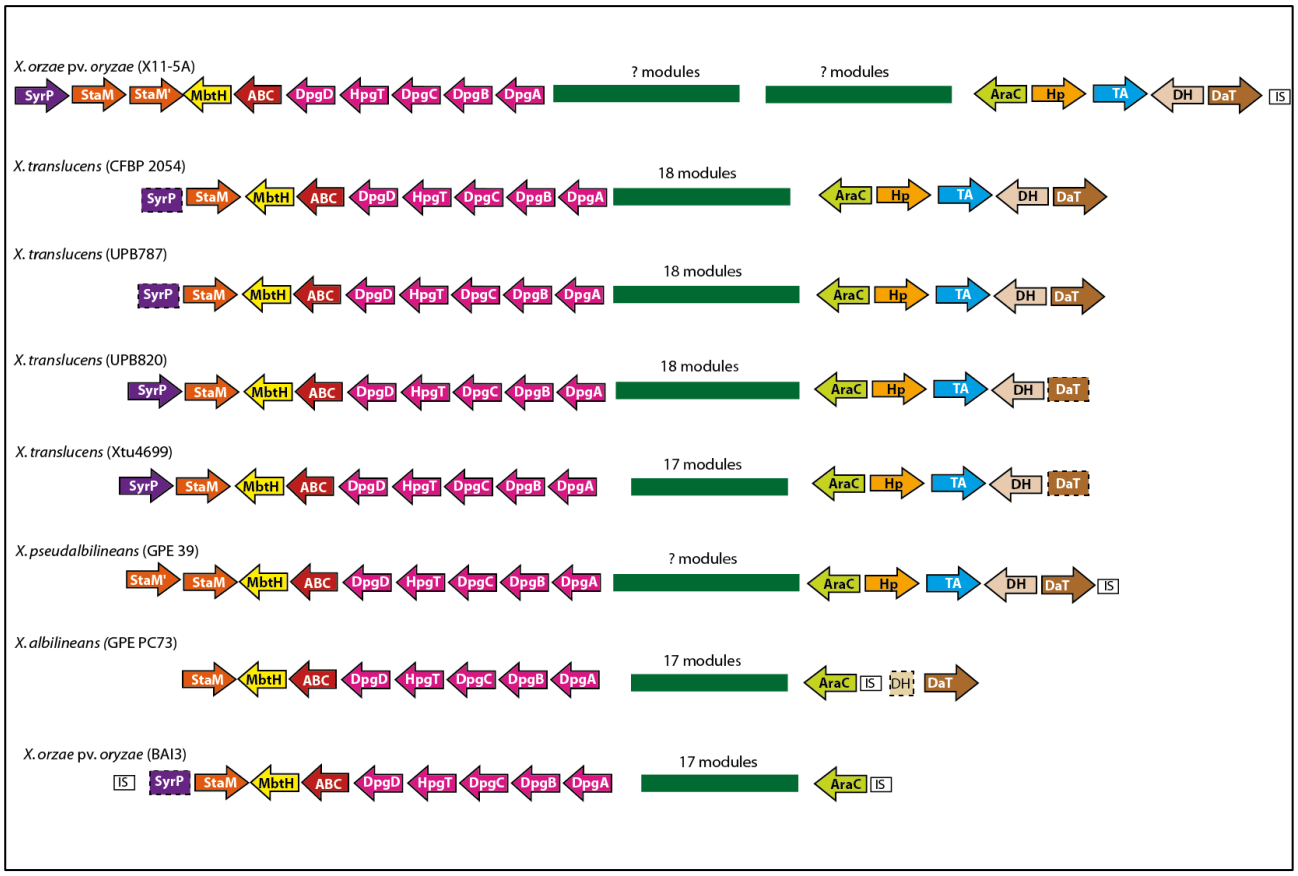


Figure 19: Comparison of META-B BGCs identified in eight *Xanthomonas* strains. A green box represents a Meta-B NRPS system. Each of these systems contains several NRPS genes. The respective number of modules for each system is noted above the green boxes. Each system contains a lipo-initiation C-starter domain and a TE domain. Associated genes are indicated with colored arrows. The length of the arrows is not proportional to the length of the respective genes. Dark orange arrows: β-hydroxylases-encoding genes *staM* and *staM'*. Brown arrows: gene *daT*, responsible for the biosynthesis of Dab (2,4-diaminobutyric acid). Pink arrows: five genes involved in biosynthesis of Dpg (3,5- dihydroxyphenylglycine). Yellow, red and light green arrows: genes encoding a MbtH protein, an ABC transporter and an AraC transcriptional regulator, respectively. Blue, light brown and light orange arrows: genes encoding a transaminase, a dehydrogenase and a hypothetical protein, respectively, which could be involved in the biosynthesis of an unknown non proteinogenic aa. IS stands for insertion sequence. Dotted boxes correspond to pseudogenes. Taken from Sabri, 2016.

II.2. Isolation and characterization of the Meta-B compound

Simultaneously to the *in silico* analysis, I operated the Meta-B purification process. Following our former experience with albicidin, we anticipated a low production yield for Meta-B in *X. albilineans* culture supernatants, as is the case for many secondary metabolites. We therefore engineered an over-producing *X. albilineans* strain using an auto-replicative plasmid containing the Meta-B AraC transcriptional regulator gene under the control of a strong constitutive promoter (plasmid pAraC). Cultures of the pAraC-transformed *X. albilineans* strain GPE PC73 were grown in presence of XAD-7 Amberlite resin. The compounds adsorbed onto the resin were eluted and the corresponding concentrated extracts were submitted to reversed-phase HPLC. The control WT strain bearing the corresponding empty plasmid was treated in the same conditions, allowing differential analysis. Indeed, I observed a supplementary peak in the pAraC chromatogram *vs* the control chromatogram. This result suggests that the WT strain does not produce the Meta-B compound in our laboratory conditions (at least not in detectable quantities), but that the Meta-B BGC has been awaken by the increased level of the AraC transcriptional regulator. I isolated the compound and transferred it at the Technical University of Berlin for MS analysis. A main mass of $[M+2H]^{2+}$ = 1147.51 Da was detected, corresponding to a

compound of exact mass = 2293 Da. MS/MS measurements gave insights into parts of a peptidic sequence being close enough to the *in silico* predicted sequence to confirm that the purified molecule is indeed Meta-B. More precisely, the peptidic sequence between aa3 and aa14 could be determined by calculation of the mass differences between adjacent MS peaks. It displays the following slight modifications compared to the *in silico*-predicted sequence: masses corresponding to aa12 and aa13 (same unknown substrate-specificity signature) are indeed similar and correspond to Dpg, whereas masses of aa4, aa6 and aa9 (*in silico*-predicted as asparagines) are 16 Da higher than expected, leading us to propose that they correspond to hydroxy-asparagines (Hyn) (figure 20A). In connexion to this result, we previously observed that the MS measurement of the Meta-B molecule also yielded a series of minor peaks whose masses potentially correspond to Meta-B analogues in which one, two or the three Hyn are replaced by asparagines (figure 20B). Strikingly, we never observed the form bearing four Hyn, although Meta-B is predicted to contain a total of four asparagines.

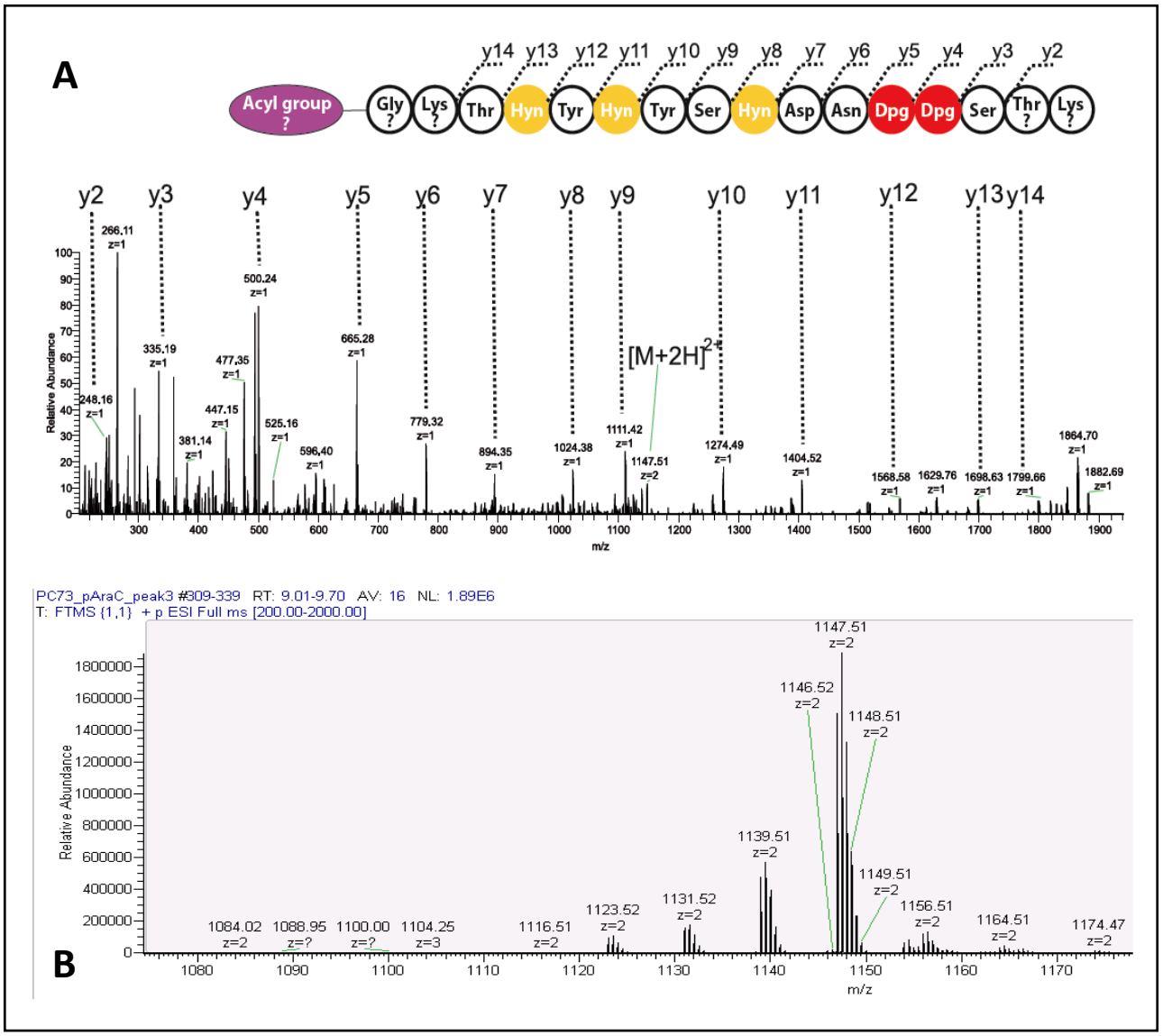


Figure 20. MS analysis of the Meta-B compound. A: MS/MS analysis of Meta-B and attribution of the Roepsdorff peptidic fragments to the *in silico*-predicted sequence. Question marks indicate residues predicted *in silico* but not confirmed by MS; Dpg: 3,5-dihydroxyphenylglycine; Hyn: hydroxy-asparagine. B: MS analysis of Meta-B. Peaks displying the masses $[M+2H]^{2+}$ = 1147.51 Da, 1139.51 Da, 1131.52 Da and 1123.52 Da correspond to the meta-B compound containing three, two, one and zero Hyn, respectively. Notably, no mass peak corresponding to a Meta-B molecule containing four Hyn ($[M+2H]^{2+}$ = 1155.51 Da) has been observed (taken from Sabri, 2016).

Additional sequence features were obtained by MS and MS/MS studies of tryptic digests of Meta-B. Initially designed to get rid of most of the peptidic part in order to facilitate the structure determination of the acyl part of the lipopeptide, they however unexpectedly allowed to identify a 17th aa in the peptidic sequence (between Lys2 and Thr3 in the initially predicted sequence), which has been identified as 2,3-didehydroaminobutyric acid (Dhb). These experiments also confirmed that the predicted Gly1, Lys2, Thr16 and Lys17 are correct. A gas chromatography-MS analysis performed in the Süssmuth's laboratory confirmed the presence of Dhb and attributed the D-*allo* conformation to a threonine, most likely Thr16 since its corresponding module is followed by a module containing a dual C/E epimerization domain. From these data, we deduced that the acyl part of the molecule exhibits a mass of 228 Da. Its structure is currently under investigation.

The discovery of the additional 17th aa somehow contradicted the genomic sequence data of *X. albilineans*. This prompted Souhir Sabri, in frame of her PhD, to re-sequence the corresponding clones from the *X. albilineans* strain GPE PC73 genomic library. Restriction patterns of the corresponding clones showed indeed a 3-kb additional sequence compared to the theoretical one (Sabri, 2016). This "missing" module was most likely over-looked due to an error in the contigs assembly during the genome sequencing since the Dhb module and its contiguous Thr module have very similar sequences. More generally, bacterial genomes are often mis-assembled at the level of NRPS genes because of the repetitive nature of their sequences, indicating that eventhough *in silico* analyses generate many useful data for detection of BGCs and structure predictions of the corresponding secondary metabolites, the biochemical approach leading to isolation and characterization of the encoded molecule is compulsory for full description. Concerning Meta-B, the sequencing of the additional module surprisingly shows that its substrate-specificity signature sequence is meant to activate a threonine. But there are other examples of NRPS modules incorporating Dhb instead of a predicted threonine. This is for instance the case for the syringopeptine and the syringomycine from *Pseudomonas syringae* (Scholz-Schroeder *et al.,* 2001), where Dhb was shown to derive from threonine by a dehydration reaction (Grgurina & Mariotti, 1999).

We also investigated the mechanism(s) of hydroxylation of asparagines. In order to identify candidate genes, Souhir Sabri constructed several mutants respectively affected in the gene *alb08* (encoding a hydroxylase) from the albicidin BGC, in the gene *staM* from the Meta-B BGC and in a cytochrome P450 gene specific to *X. albilineans*. Neither the individual mutants nor the triple mutant had any effect on the hydroxylation status of asparagines in Meta-B, as confirmed by MS analysis of the respective purified Meta-B compounds. Another gene candidate for asparagine hydroxylation is a gene encoding another cytochrome P450 which is conserved in all *Xanthomonas* species, but unfortunately no mutant could be obtained for this gene, suggesting that it could be of vital importance for the bacterium (Sabri, 2016).

II.3. Conclusion

As of today, the peptidic part of Meta-B is completely deciphered but its acyl part is more reluctant to surrender and is still under investigation, as is the biological function of Meta-B. No obvious antibacterial, antifungal or siderophore activities could be evidenced so far for this molecule. Plant-pathogen interactions were tested using Meta-B-deleted strains of *X. albilineans* (PhD of Souhir Sabri) or of *X. oryzae* (PhD of Mélanie Marguerettaz) which were tested in various plants for their colonization properties, but no data were strong enough to conclude about a putative role of Meta-B in the plant colonization by the pathogen.

III. How data on the albicidin biosynthetic pathway help to discover new natural products in bacteria from the genera *Xanthomonas* and *Streptomyces*?

III.1. Are the unique structural features of albicidin unique?

Strikingly, within days after our publication of the structure of β-albicidin appeared on-line, a paper was published describing cystobactamids, a family of DNA gyrase inhibitors produced by the soil myxobacterium *Cystobacter*. Interestingly, they exhibit almost similar structures compared to albicidins, as well as a similar DNA gyrase inhibitor activity (Baumann *et al.,* 2014). Recently, more cystobactamids derivatives were discovered, sometimes only differing by a single hydroxyle group from their albicidin couterparts (Hüttel *et al.,* 2017; Cheng *et al.,* 2017). With no surprise, because of their almost similar structures, albicidins and cystobactamids share almost-identical biosynthetic machineries which we compared in frame of the PhD of Souhir Sabri. The cystobactamid NRPS assembly line is devoid of PKS, but instead exhibits an additional NRPS module predicted to incorporate *p*ABA. Indeed, five out of the six NRPS modules from cystobactamid are predicted to be specific of *p*ABA or *p*ABA derivatives. The sixth module (being actually the third in the assembly line) is non-functional (as is the case for albicidin module NRPS-2) and is predicted to be *trans*-complemented by a stand-alone NRPS module which has a signature sequence dedicated to the incorporation of an asparagine. It also contains the same 342-aa αANH-like insertion found in the module NRPS-2* from albicidin.

Because we hypothesize that it is a way to reveal new promising NRPS BGCs candidates for chemical and functional studies, we used the unique features of albicidin, *i.e.* the NRPS-1 A domain whose specificity is directed towards *p*ABA building blocks, and the NRPS-2* A domain containing a 342-aa αANH-like extension, as targets in a genome-based survey performed with bacterial sequences available in GenBank. Thirteen BLAST hits were obtained for the NRPS-2* A domain and three others for the *p*ABA-specific domain sequence. All hits were analyzed by the antiSMASH software which yielded the domain architecture of the corresponding NRPS/PKS BGCs. The in-built device NRPSpredictor2 determined the aa-specificity-conferring signatures. (Sabri *et al.,* in preparation).

*p*ABA-activating A domains were identified in the genomes of the mango pathogen *Xanthomonas citri* pv. *mangiferaeindicae* strain LMG941 (see figure 21 for full description of the corresponding BGC), the aquatic bacterium *Janthinobacterium* sp. HH01 and the soil bacterium *Actinoplanes utahensis* NRRL 12052. The respective BGCs comprising the NRPS genes encoding these *p*ABA-specific A domains are all different, implying that they encode three different molecules. Interestingly, a *p*ABA synthase encoding gene is present in all three BGCs (Sabri *et al.,* in preparation).

A domains exhibiting the αANH-like 342-aa insertion were identified in 13 genomes of bacteria isolated in environments as different as plant root nodules, human blood, beach sand, a green alga, waste water, a municipal sewage sludge or the gut of a wood-feeding termite, for instance. In all these strains, the newly-discovered BGCs all predictively encode the biosynthesis of very different molecules because (i) the BGCs exhibit a gene composition highly variable, especially regarding the genes involved in tailoring of either the aa-substrates or the final products, (ii) the NRPSs are very different in number and modules specificity, and (iii) the domain architecture of the αANH-specific modules are variable (Sabri *et al.,* in preparation).

[Note: These results are included in a publication which has been drafted since 2016 before the PhD defense of Souhir Sabri. However, we decided to postpone the submission for publication in BMC Genomics because it focuses for a great part on the original αANH-like domain suspected to catalyze the conversion of asparagine into cyanoalanine during albicidin biosynthesis. Indeed, we currently prefer it not to be an eye-catcher since our german collaborators in Berlin actually perform experimental functional studies on this domain, aiming at describing its exact function and mechanism, with these results being of great interest to reinforce the impact of the paper of Souhir Sabri. However, the submission of this paper is planned before June 2019].

These results tell us that this genome mining approach is efficient for the identification of BGCs exhibiting rare structural features which are worth investigating as they might, as albicidin, represent good candidates for drug development. We chose to take advantage of the data on these very unusual features of the biosynthesis of albicidin to develop projects seeking to implement a genome mining approach to uncover new microbial secondary metabolites BGCs. Two collaborative projects are in the pipeline and aim at mining either a collection of *Xanthomonas* bacteria or a collection of *Streptomyces* bacteria for the presence in their genome of genes encoding *p*ABA-specific or αANH-like-containing A domains. Both projects, which involve slightly different approaches and different collaborators, are described below.

III.2. Genome mining of a collection of *Xanthomonas* bacteria for new natural products [project "XanthoNRPS"]

Data obtained in frame of our previous studies on both albicidin and Meta-B, and from our genome mining analysis of available sequences of plant pathogenic bacteria belonging to the *Xanthomonas* genus prompted us to consider a thorough prospection of the biosynthetic potential of NRPS systems in the genus *Xanthomonas* to discover new secondary metabolites potentially of interest for agronomical or medical applications. Hence, we propose a genomics-based tactic supported by (bio)chemical and functional studies. In our previous extensive genome mining study, we showed that bacteria belonging to the *Xanthomonas* genus stand as a promising reservoir for biosynthesis of NRPs (Royer *et al.,* 2013). Besides albicidin, Meta-B and four other putative NRPs already described in *X. albilineans*, we showed the potential of other species, namely *X. oryzae* and *X. translucens*, to produce Meta-B-like molecules. We also identified NRPS genes encoding a 6-module NRP in draft genomes of *X. oryzae* pv. *oryzicola* strain BSL256 and a remnant of the same BGC has been pinpointed in *X. axonopodis* pv. *citri* strain 306 (Royer *et al.,* 2013). Also, in our attempt to "fish" in the bacterial genomes available in GenBank NRPS genes encoding a *p*ABA-specific A domain, we revealed in the genome of the mango pathogen *Xanthomonas citri* pv. *mangiferaeindicae* strain LMG941 a new NRPS/PKS BGC, hereafter referred to as Mango-LMG941, and which is predicted to synthesize a natural product containing *p*ABA and therefore displaying a novel structural feature (figure 21).

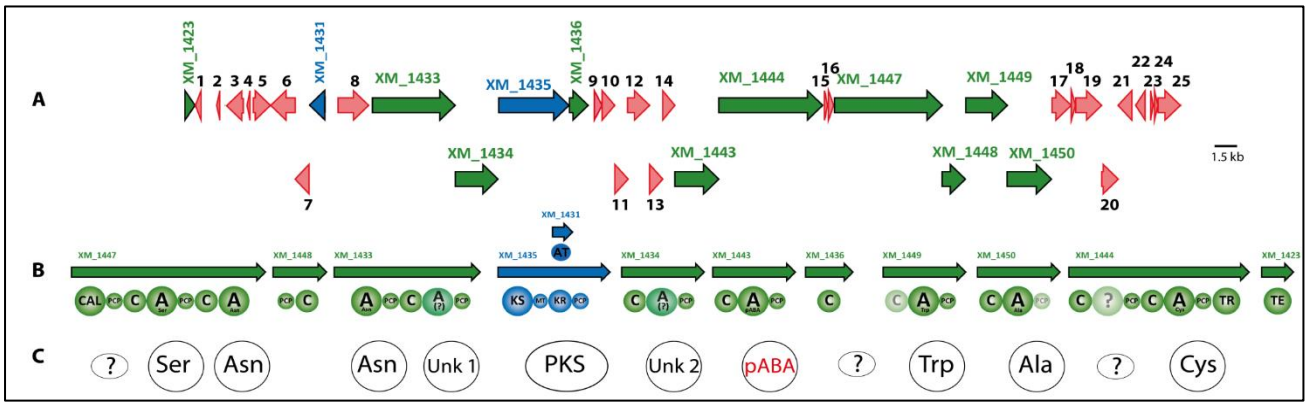


Figure 21. *In silico* analysis of the PKS/NRPS BGC of *X. citri* pv. *mangiferaeindicae* strain LMG941. A: Genetic organization of the Mango-LMG941 BGC. Ten NRPS and two PKS genes (noted as XM_14xx) are shown by green and blue arrows, respectively. Other genes, numbered from 1 to 25, are predicted to encode as follows: 1, 4, 7, 10, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23: hypothetical proteins; 2, 24: transposases; 3: a *p*ABA synthase similar to the albicidin biosynthesis enzyme Alb17; 5: a permease family protein; 6: an ABC transporter; 8: a glycosyltransferase; 9, 11, 13: β-hydroxylases; 15: a MbtH-like protein; 25: a PPTase. B: Predicted arrangement of NRPS modules and domains encoded by the Mango-LMG941 BGC. Nomenclature for domains is as follows: CAL: fatty acyl-AMP ligase; TR: thioreductase, MT: methyltransferase. For all other domains, see figure 13. Pale discs: domains for which the prediction level is weak. The signature of the A domain predicted to be specific to *p*ABA is APKHVANVDK which is close to the signature of albicidin *p*ABA modules (AVKYVANDAK). C: Partially predicted sequence of the PK/NRP encoded by the Mango-LMG941 BGC (3-letter code). Unk: unknown. *p*ABA: para-aminobenzoate. PKS: moiety incorporated by the PKS part of the locus. ?: substrates for which prediction is not possible (low prediction levels for their corresponding NRPS domains).

So far, our tentatives to generate a mutant in the BGC Mango-LMG941 have failed, probaby because this species is reluctant to genetic manipulation. Thus, we propose to transfer this BGC into an heterologous *Xanthomonas* host.

Xanthomonas bacteria appear therefore as a genus offering an important biosynthetic potential however largely under-estimated and under-explored. To explore the *Xanthomonas* biosynthetic potential, we therefore propose to identify new NRPS or hybrid PKS/NRPS BGCs by sequencing and mining approx. 200 strains spanning the overall diversity of the genus *Xanthomonas*. Indeed, the *Xanthomonas* genus encompasses at least 28 species of Gram-negative plant-associated bacteria (Triplett *et al.,* 2015) which cause dramatic damages to hundreds of plants species of ornamental or agronomical interest. *Xanthomonas* bacteria target over 400 plants belonging to both monocotyledons (banana, sugarcane, rice,…) and dicotyledons (tomato, cabbage, citrus,…). New promising BGCs will be investigated to determine the structure and the biological function of the corresponding natural products. The proposed project is a collaborative effort with the teams led by Marie-Agnès Jacques (UMR IRHS, Angers), Jérôme Gouzy (UMR LIPM, Toulouse) and Roderich Süssmuth (Technical University of Berlin). This partnership enlightens complementarity between four teams with high-level competences on different methods and techniques. Our team will provide large quantities of secondary metabolites and develop strong strategies for overproduction and isolation. The Süssmuth team will perform structure elucidation and synthesis of NRPs. At UMR IRHS, the team of Marie-Agnès Jacques is in charge of the French Collection for Plant-associated Bacteria (CIRM-CFBP http://www6.inra.fr/cirm_eng/CFBP-Plant-Associated-Bacteria), whereas the team of Jérôme Gouzy, at UMR LIPM, is in charge of a bioinformatics platform of INRA. The three French teams belong to the French Network on Xanthomonads (FNX) which was created in 2008 in order to achieve projects on genomics, and already published together the annotated sequences of the genomes of *X. albilineans* strain GPE PC73 (Pieretti *et al.,* 2009; Pieretti *et al.,* 2012) and *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R (Darrasse *et al.,* 2013).

We propose to sequence 192 strains of *Xanthomonas* to identify additional promising NRPS loci. In 2015, the genomic sequences available in GenBank included only 23 out of the 28 valid *Xanthomonas* species with an over-representation of some species or some pathovars, *i.e.* 66 sequences of *Xanthomonas axonopodis* pv. *manihotis*, 24 sequences of *X. axonopodis* pv. *phaseoli* and 26 sequences of *X. citri* pv. *citri*. Additionally, uncharacterized species of *Xanthomonas* are not yet sequenced because they are not of agronomical interest. To optimize the chances of identifying new additional NRPS BGCs, the 192 strains of *Xanthomonas* selected for sequencing will be representative of the overall diversity of the genus and will not be limited to agronomically important taxa. These strains will be provided by the French Collection for Plant-associated Bacteria, which holds about 1700 *Xanthomonas* strains distributed in 115 different taxa.

I propose to implement the general genome mining strategy described in §I.3 p57 to screen the genomes of the collection of *Xanthomonas* strains. Gene content of the 192 assembled draft sequences (obtained by NGS sequencing and Shotgun Illumina) will be examined by BLAST, using as target the sequences of *p*ABA-specific or αANH-like-containing A domains. Positive hits will be further analyzed with predictive informatic tools. PacBio resequencing of the draft sequences exhibiting NRPS domains of interest needs to be done in order to confirm the sequence. Heterologous transfer of the full BGC of interest will be performed for eventual awakening of the BGC if silent, or for over-production of the encoded natural product should its production yield be low in its natural producer. Subsequent steps for production, extraction, fractionation, isolation and structral characterization will globally follow the protocole described above.

Expected results should also be of interest for basic research. Because *Xanthomonas* strains are plant-associated bacteria, identifying new molecules in this genus should help to better understand plant-bacteria interactions. Because NRPS systems described to date in *Xanthomonas* differ greatly from those of other genera, expected results should contribute to decipher molecular mechanisms involved in the

fascinating and yet poorly understood non-ribosomal peptide synthesis pathway. The discovery of molecules exhibiting hitherto unknown structures should also be of interest for chemistry research.

This project has been previously submitted for funding at ANR but was not selected. We plan, however, a future re-submission.

III.3. Genome mining of a collection of *Streptomyces* bacteria for new natural products [project "CONTACT"]

Willing to broaden our horizon, our team recently decided to extend to sources other than *Xanthomonas* bacteria its competences for identification, isolation and characterization of new non-ribosomally-synthesized secondary metabolites. In this regard, *Streptomyces* bacteria (actinomycetes) have drawn our attention. *Streptomyces* became particularly popular after the discovery, in the 1940s, of widely used antibiotics such as actinomycin, streptomycin and chloramphenicol (Barka *et al.,* 2016). Indeed, *Streptomyces* represent the most prolific source of bioactive compounds since almost half of the 23000 microbial secondary metabolites and over 50% of the commercially-available antibiotics isolated to date are produced by members of this genus (Demain, 2014). As shown by genome sequencing and mining, *Streptomyces* are true metabolic factories, with an average reservoir of about 20 to 30 BGCs per strain, although it appears that most of the secondary metabolic potential of *Streptomyces* strains remains silent under the normal laboratory conditions (see, for instance, Bentley *et al.* (2002) for the first *Streptomyces* genome sequencing, annotation and mining).

Lastly, we co-developed the project "CONTACT" (mycotoxins and fungal pathogens CONTrol by ACTinomycetes) in collaboration with the team of Sabine Schorr-Galindo at UMR QualiSud (University of Montpellier), which provides expertise on actinomycetes, and the team of Jean-Benoit Morel at UMR BGPI (INRA) whose interest is the study of plant-pathogen interactions (including models of fungal diseases on cereals). Also in the project stands our historical partner Roderich Süssmuth (TUB) who provides his expertise on structural characterization of natural products. As such, the project gathers a pluridisciplinary international consortium of microbiologists, phytopathologists, biochemists, molecular biologists and chemists, all expert in their domains. The project aims at studying the potentialities of actinomycetes from soils for the development of novel biocontrol products to manage cereal fungal diseases and mycotoxins risk, while taking advantage of actinomycetes metabolites production richness and diversity to identify new biomolecules with high agronomical or biopharmaceutical potential. This 30-month project has been funded up to 126 k€ by the Agropolis Foundation (OpenScience call 2016), and has just started in December 2017. A PhD and five master students are expected to take part of the project. Our team will host two master students who I presently co-supervise.

UMR QualiSud houses a collection of nearly 60 *Streptomyces* strains isolated from the soils of vine and cereal cultures in the south of France. Some have been already characterized for their antagonist effect against mycotoxigenic fungi with encouraging results such as a high inhibition rate of moulds growth and toxin production. This project aims at valorizing this collection through two main objectives: (i) to propose a global approach to evaluate the potential of *Streptomyces* isolates for the control of wheat fungal diseases from *in vitro* testing till *in planta* efficacy, and (ii) to discover new secondary metabolites of interest through genome mining of the collection of *Streptomyces* from UMR QualiSud.

For the first part of this project, the most promising strains regarding (i) their antagonist activity against selected toxigenic and phythopathogenic fungi (direct confrontation and cell-free extracts), (ii) their enzymatic activities that reflect a potential to degrade fungi cell wall, but also to grow on poor substrates such as residuals and wastes from agriculture, (iii) their technological properties (growth yield, sporulation, dehydration), (iv) their ability to colonize the rhizosphere, to ensure a sustainable effect (endophyte capacity), and, finally, (v) their potential as plant growth promoters (PGPR) mostly through *in vitro* and molecular method analysis, will be selected for *in planta* testing to confirm their potential to reduce the incidence of fungal contamination and/or

the production of mycotoxins and their potential as biocontrol agents. The strains having the most potent effect *in vitro* (mostly through the cell-free extracts antagonist activities that we will perform) will be fully sequenced and we will mine their genome for BGCs (secondary metabolites production) (see general strategy described in §I.3 p57). Strains for which genome mining will unravel promising new BGCs will be selected to study their active metabolites fractions for their effects and mode of action on the biocontrol of fungal pathogens, and will be further analysed for structure elucidation in the case of unknown molecules.

My main interest in this project, however, lies in the opportunity to have access to a collection of *Streptomyces* strains recently isolated in the field and therefore which have not yet been investigated for their secondary metabolic potential. As such, this collection possesses a metabolic reservoir that calls for exploration. My goal is to screen this collection for the presence of BGCs encoding new compounds displaying unusual structural features. As for the project "XanthoNRPS", we propose to start by investigating the presence of genes encoding NRPSs with either *p*ABA-specific or αANH-like-containing A domains. This work has just started in frame of the internship of a master1 student, Baptiste Durand, who performed a BLAST search against the available genome sequences of *Streptomyces* in GenBank with the342-aa αANH-like sequence as query (and not the whole alb04 A domain containing the αANH-like insertion, as in Sabri *et al.,* in preparation). Positive hits were obtained for ten genomes in which he identified a new gene encoding the αANH-like domain always followed by an epimerization domain. This gene is always present upstream of an asparagine synthase within a NRPS BGC, in which it nevertheless always appears as a stand-alone module. Each of these ten BGCs falls into one of four NRPS system architectures. Interestingly, Baptiste Durand also blasted the GenBank *Streptomyces* genomes for the presence of genes encoding *p*ABA-specific A domains, but he did not find any.

On the basis of the alignment of the ten sequences for the gene "αANH-epimerization", Baptiste Durand designed primers that he used to screen by PCR the collection of *Streptomyces* strains from UMR QualiSud. Interestingly, we observed a positive hit for one of the strains, which conforts this approach. The best match of the 16/23S sequences of this strain is *Strepomyces griseus.* The genome of this strain will be sequenced and the *in silico* analysis of the corresponding BGC should result in structural predictions that will give hints for subsequent isolation and characterization of the encoded molecule.

For the continution of the project, we propose to apply the general strategy described in §I.3 p57 for awakening the BGC of interest and over-producing the secondary metabolite of interest. Production, isolation and structural characterization would also follow the same strategy. In frame of his master, Baptiste Durand also determined conditions for optimal *Streptomyces* manipulation (culture conditions, DNA extraction).

Regarding the outcome of this project, in addition to the expected scientific knowledge about the soil actinomycetes diversity, richness and potentialities that will be characterized for their efficiency to inhibit the growth of fungal cereal pathogens and mycotoxinogens and/or to inhibit *in vitro* and *in planta* the production of toxins itself, and their technological potential (growth on poor agricultural waste, sporulation, production yield), we are optimistic (after our first screening of the collection) to generate knowledge about promising secondary metabolites whose biological activity will be investigated through various biological assays.

Supervision

03-2018 / 08-2018	Agathe Ropars : Master 2 and engineer school: Etude d'un signal de quorum sensing synthétisé par des NRPS (Ecole Supérieure d'Ingénieurs en Agroalimentaire de Bretagne Atlantique (ESIAB)).
12-2017 / 04-2018	Baptiste Durand : Master 1: Recherche de loci codant de nouveaux métabolites secondaires chez les bactéries du genre *Streptomyces* (Master Biotin, Université de Nîmes, Mines d'Alès).
04-2016 / 05-2016	Touhfa Fettouhi : Licence 3: Etude de peptides non-ribosomiques produits par des bactéries (Université de Haute-Alsace).
03-2014 / 08-2014	Alexandre Morisset : Master 1: Etude du rôle du polymorphisme du locus LPS de *Xanthomonas albilineans* dans l'évolution et la pathogénie de cette bactérie responsable de l'échaudure des feuilles de la canne à sucre (Université Montpellier 2).
02-2013 / 07-2016	Souhir Sabri : Thèse (co-direction ADR) soutenue en juillet 2016: La biosynthèse non-ribosomique chez les bactéries du genre *Xanthomonas*. (Montpellier SupAgro / ED GAIA).
01-2012 / 12-2015	Isabelle Pieretti : Thèse (co-supervision; member of the thesis comitee): Génomique comparative et évolutive de *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre. (Montpellier SupAgro / ED Sibaghe).
03-2012 / 06-2012	Cyril Magno : Master 1: Etude de trois loci NRPS identifiés dans le génome de *Xanthomonas albilineans*, bactérie pathogène de la canne à sucre. (M1 Biologie des Plantes Tropicales / Université Montpellier 2 / Montpellier SupAgro).
10-2011 / 12-2011	Julie Arasté : BTS 2: Purification des formes alpha et gamma de l'albicidine. (BTS BioAnalyses et Contrôles / Lycée Mermoz, Montpellier).
05-2011 / 07-2011	Julie Arasté : BTS 1: Purification des formes alpha de l'albicidine. (BTS BioAnalyses et Contrôles / Lycée Mermoz, Montpellier).
10-2010 / 11-2013	Imène Mensi : Thèse (co-supervision): Localisation *in planta* de *Xanthomonas albilineans* et identification de déterminants moléculaires impliqués dans la colonisation épiphyte de sa plante hôte, la canne à sucre (Université Montpellier 2 / ED Sibaghe).
01-2008 / 06-2008	Jérôme Puig : Master 2: Etude du système de sécrétion de type III de *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre (M2 Sciences et Technologies / Université Montpellier 2).
11-2007 / 12-2010	Mélanie Marguerettaz : Thèse (co-supervision; member of the thesis comitee): Rôle du système de sécrétion de type III SPI-1 et des mégaenzymes NRPS dans le cycle de vie de *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre (Montpellier SupAgro / ED Sibaghe).
01-2004 / 05-2006	Patrice Champoiseau : Thèse (co-supervision): *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre: caractérisation et variabilité des bases génétiques du pouvoir pathogène, en Guadeloupe et dans le monde (Université des Antilles et de la Guyane).
03-2004 / 06-2004	Adeline Renier : DEA: Diversité génétique de *Xanthomonas albilineans* et variabilité des enzymes de biosynthèse de l'albicidine, pathotoxine produite par l'agent causal de l'échaudure des feuilles de la canne à sucre (DEA Ressources Phytogénétiques et Interactions Biologiques / Agro Montpellier / Université Montpellier 2).
01-2003 / 10-2005	Eric Vivien : Thèse (co-supervision; member of the thesis comitee): Biosynthèse de l'albicidine, une molécule antibiotique et une pathotoxine de *Xanthomonas albilineans* (Université Montpellier 2).

Funding

* 2009-2012: ANR / DFG (Acronym: ALBILINEANS; ANR-09-BLAN-0413-01; DFG SU 239/11-1): Etude structurale et biologique de petites molécules impliquées dans la pathogénie de *Xanthomonas albilineans*, une bactérie pathogène de la canne à sucre : 213 k€

* 2013-2016: ANR program "Investissements d'avenir" (Acronym: XANTHOMONAPEPTIDES; ANR-10-LABX-0001-01, project n° 1202-013): Deciphering a new family of small molecules produced by *Xanthomonas oryzae*, *Xanthomonas translucens* and *Xanthomonas albilineans*: 130 k€

* 2017-2019: Agropolis Foundation - OpenScience 2016 call (Acronym: CONTACT; 1605-052): Mycotoxins and fungal pathogens control by actinomycetes: 126 k€

* Project submitted in 2018: ANR / DFG PRCI (Acronym: DickeyaVFM): Characterization and structure elucidation of a new quorum sensing system in plant pathogenic bacteria: 503 k€

Collaborations

* Roderich Süssmuth (Institut für Chemie, Technische Universität Berlin, Berlin, Germany)

* Sabine Schorr-Galindo (UMR QualiSud, Université de Montpellier)

* Nicole Cotte-Pattat (UMR MAP, CNRS, Lyon)

* Jean-Benoit Morel (UMR BGPI, INRA, Montpellier)

* FNX "French Network on Xanthomonads:"
UMRs IPME (Montpellier), IRHS (Angers), LIPM (Toulouse), PVBMT (La Réunion)

Valorization

The work performed on pathogenicity of *Xanthomonas albilineans* has been valorized in 17 publications (5 as first/second or last author), 2 reviews (1 as last author), 1 book chapter, 1 patent, 1 article for general audience (first author) and 42 communications (2 invited) in meetings (oral or poster) (not listed, see list of publications):

Vivien E, Megessier S, Pieretti I, **Cociancich S**, Frutos R, Gabriel DW, Rott PC & Royer M **(2005)** *Xanthomonas albilineans* HtpG is required for biosynthesis of the antibiotic and phytotoxin albicidin. *FEMS Microbiol. Lett.* 251, 81-89.

Champoiseau P, Daugrois J-H, Pieretti I, **Cociancich S**, Royer M & Rott P **(2006)** High variation in pathogeny of genetically closely related strains of *Xanthomonas albilineans*, the sugarcane leaf scald pathogen, in Guadeloupe. *Phytopathology* 96, 1081-1091.

Vivien E, Pitorre D, **Cociancich S**, Pieretti I, Gabriel DW, Rott PC & Royer M **(2007)** Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* 51, 1549-1552.

Renier A, Vivien E, **Cociancich S**, Letourmy P, Perrier X, Rott PC & Royer M **(2007)** Substrate specificity-conferring regions of the non ribosomal peptide synthetase adenylation domains involved in albicidin pathotoxin biosynthesis are highly conserved within the species *Xanthomonas albilineans*. *Appl. Environ. Microbiol.* 73, 5523-5530.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, **Cociancich S**, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M & Rott P **(2009)** The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* 10, 616.

Rott P, Marguerettaz M, Fleites L, **Cociancich S**, Girard J-C, Pieretti I, Gabriel DW & Royer M **(2011)** Unravelling pathogenicity of *Xanthomonas albilineans*, the causal agent of sugarcane leaf scald. *Int. Sugar J.* 113, 490-496.

Marguerettaz M, Pieretti I, Gayral P, Puig J, Brin C, **Cociancich S**, Poussier S, Rott P & Royer M **(2011)** Genomic and evolutionary features of the SPI-1 type III secretion system that is present in *Xanthomonas albilineans* but is not essential for xylem colonization and symptom development of sugarcane leaf scald. *Mol. Plant-Microbe Interact.* 24, 246-259.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Couloux A, Darrasse A, Gouzy J, Jacques M-A, Lauber E, Manceau C, Mangenot S, Poussier S, Ségurens B, Szurek B, Verdier V, Arlat M, Gabriel DW, Rott P & **Cociancich S (2012)** Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genomics* 13, 658.

Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin GP, Brin C, Carrere S, Gomez C, Hügelland M, Völler GH, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth RD & **Cociancich S (2013)** Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genomics* 14, 658.

Darrasse A, Carrère S, Barbe V, Boureau T, Arrieta-Ortiz ML, Bonneau S, Briand M, Brin C, **Cociancich S**, Durand K, Fouteau S, Gagnevin L, Guérin F, Guy E, Indiana A, Koebnik R, Lauber E, Munoz A, Noël LD, Pieretti I, Poussier S, Pruvost O, Robène-Soustrade I, Rott P, Royer M, Serres-Giardi L, Szurek B, Van Sluys M-A, Verdier V, Vernière C, Arlat M, Manceau C & Jacques M-A **(2013)** Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of *Xanthomonas*. *BMC Genomics* 14, 761.

Kretz J, Kerwat D, Schubert V, Grätz S, Pesic A, Semsary S, **Cociancich S**, Royer M & Süssmuth RD **(2015)** Total synthesis of albicidin: a lead structure from *Xanthomonas albilineans* for potent antibacterial gyrase inhibitors. *Angew. Chem.-Int. Edit.* 54, 1969-1973.

Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll J, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M & Süssmuth RD **(2015)** The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. Nat. Chem. Biol. 11, 195-197.

Pieretti I, Bolot S, Carrère S, Barbe V, **Cociancich S**, Rott P & Royer M **(2015)** Draft genome sequence of *Xanthomonas sacchari* strain LMG 476. *GenomeA.* 3, e00146-15.

Vieweg L, Kretz J, Pesic A, Kerwat D, Grätz S, Royer M, **Cociancich S**, Mainz A & Süssmuth RD **(2015)** The albicidin resistance factor AlbD is a serine endopeptidase that hydrolyzes unusual oligoaromatic-type peptides. *J. Am. Chem. Soc.* 137, 7608-7611.

Pieretti I, **Cociancich S**, Bolot S, Carrère S, Morisset A, Rott P & Royer M **(2015)** Full genome sequence analysis of two isolates reveals a novel *Xanthomonas* species close to the sugarcane pathogen *Xanthomonas albilineans*. *Genes* 6, 714-733.

Petras D, Kerwat D, Pesic A, Hempel B-F, von Eckardstein L, Semsary S, Arasté J, Marguerettaz M, Royer M, **Cociancich S** & Süssmuth RD **(2016)** The O-carbamoyl-transferase Alb15 is responsible for the modification of albicidin. *ACS Chem. Biol.* 11, 1198-1204.

von Eckardstein L, Petras D, Dang T, **Cociancich S**, Sabri S, Grätz S, Kerwat D, Seidel M, Pesic A, Dorrenstein PC, Royer M, Weston JB & Süssmuth RD **(2017)** Total synthesis and biological assessment of novel albicidins discovered by mass spectrometric networking. *Chem.-Eur. J.* 23, 15316-15321.

Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD & **Cociancich S (2015)** What makes *Xanthomonas albilineans* unique amongst Xanthomonads? *Front. Plant Sci.* 6, 289.

Jacques M-A, Arlat M, Boulanger A, Boureau T, Carrère S, Cesbron S, Chen NWG, **Cociancich S**, Darrasse A, Denancé N, Fischer-Le Saux M, Gagnevin L, Koebnik R, Lauber E, Noël LD, Pieretti I, Portier P, Pruvost O, Rieux A, Robène I, Royer M, Szurek B, Verdier V & Vernière C **(2016)** Using ecology, physiology and genomics to understand host specificity in *Xanthomonas*. *Ann. Rev. Phytopathol.* 54, 163-187.

Royer M, Pieretti I, **Cociancich S** & Rott P **(2018)** Recent progress in understanding three major bacterial diseases of sugarcane: gumming, leaf scald and ratoon stunting. *Achieving sustainable cultivation of sugarcane. Volume 2: breeding, pests and diseases.* (ed. Philippe Rott) pp311-336, Burleigh Dodds Science Publishing, Cambridge, UK.

Süssmuth R, Kretz J, Schubert V, Pesic A, Hügelland M, Royer M, **Cociancich S**, Rott P, Kerwat D & Grätz S **(2014)** Albicidin derivatives, their use and synthesis. PCT WO 2014/125075 A1.

Cociancich S, Royer M & Rott P **(2017)** Un antibiotique découvert dans la canne à sucre. *La Recherche, n°524 (juin 2017)*, pp 62-65.

CONCLUDING REMARKS

As explained in this document, microorganisms engineer an overwhelming number of natural products encompassing a broad range of biological activities and displaying an incredible variety of complex structures. For a large part, this chemical diversity is biosynthetically achieved through the non-ribosomal peptide synthesis—a ribosome-independent enzymatic mechanism that is used by bacteria and fungi to produce bio-active secondary metabolites. Within larger genomic regions called biosynthetic gene clusters (BGCs), non-ribosomal peptide synthetases (NRPSs) genes, along with the related polyketide synthases (PKSs) genes, encode multi-modular enzymatic assembly chains which assemble building blocks in a sequential step-wise fashion into growing non-ribosomal peptides or polyketides, respectively.

Following the decrease of discovery of new metabolites through the classical screening of biological activities, new approaches took advantage of the increase of genome sequencing and of the development of bioinformatic tools dedicated to identify BGCs and to analyze their NRPS genes. In my past works as well as current and future projects on the study of non-ribosomally-synthesized secondary metabolites, I used two approaches whether the goal was to (i) determine the structure of compounds previously discovered through their biological activity, or (ii) determine the structure of new compounds discovered by genome mining.

(i) The work on the antibiotic albicidin shows how the *in silico* analysis of the NRPSs/PKSs-encoding genes provided predictive data which efficiently contributed to its structure determination. Surely, these data were those missing to Robert Birch back in 1985 when he had in his hands all experimental data but did not get the clues brought later by the *in silico* analysis. I also want to acknowledge that, from his side, my colleague Roderich Süssmuth provided also a state-of-the-art analysis of dozens of mass and NMR spectra which, everything taken together, resulted in the deciphering of the long awaited structure of albicidin.

The project on the quorum sensing signal VFM is another example of the power of *in silico* analysis, which allowed us to generate a predictive model for the biosynthesis of VFM (confidential data). Generally-speaking, information incoming from *in silico* analyses are decisive for decoding the final structure of a compound as they can drive the production (e.g. by supplying the culture medium with predicted substrates), the isolation (e.g. by helping for the choice of solvents for the extraction and of chromatographical chemistry for the isolation procedures) and the structural elucidation (e.g. by helping to the attribution of MS/MS fragmentation patterns) of the natural product.

Obviously, *in silico* analysis of a secondary metabolite BGC is only an ingredient of the recipe, together with experimental structural analyses performed on the isolated compound. Sometimes, *in silico* analyses generate incomplete predictions which can be corrected by experimental analytical studies, as it has been the case with the Meta-B molecule for which the presence of an additional building block has been identified by MS compared to the predictive *in silico* analysis. Indeed, the molecular nature of the NRPS often creates misassembled contigs after sequencing of the BGCs.

(ii) The urgent quest for many new biologically-relevant secondary metabolites (e.g. new antibiotics needed to fight emerging multi-drug resistant strains from the ESKAPE group) is facilitated by the exponentially-rising number of available microbial genome sequences. Indeed, genome mining is very useful to unravel among these genome sequences the true metabolic potential of microorganisms which often exceeds by far the number of known metabolites, and shows that some microbial strains (mostly *Streptomyces* bacteria) are huge reservoirs for secondary metabolites. However, many BGCs seem most of the time silent or poorly-expressed in normal laboratory conditions (which rarely mimic the natural environment of the microorganisms). As an

example, we could find, beyond albicidin, five additional BGCs in the genome sequence of *X. albilineans*. The *in silico* analysis of the most promising BGC (Meta-B) predicted the final product to be original. Indeed, after having awaken the Meta-B BGC by over-expressing its transcriptional regulator, we almost obtained its structure but are still empty-handed for its biological function, showing here a limit of the genome mining approach. However, since the Meta-B BGC is conserved in other *Xanthomonas* strains pathogens for monocotyledons, we hypothesize that Meta-B might be involved in some kind of interaction between the plant and the bacterium (a modified ability to colonize the plant, for instance). We expect to perform some experiments within the summer with Meta-B mutants.

In our projects « CONTACT » and « XanthoNRPS », we plan to mine genomes of collections of *Streptomyces* and *Xanthomonas* strains, respectively. With a potential of several hundreds of BGCs, one might ask which to choose to study ? Due to the limited manpower in our team, we decided to limit, in a first round, our screening of both collections to the presence of genes encoding two unique features of albicidin (*i.e.* the *p*ABA-specific and the αANH-like-containing A domains). We performed a preliminary genome mining of bacterial genomes available on GenBank and showed that these features are not unique to albicidin, but still are pretty rare. Nonetheless, five *Streptomyces* strains chosen in frame of the project « CONTACT » for their potential to reduce the incidence of fungal contamination *in planta* or in direct confrontation assays will be sequenced. Genome mining on these genomes might unravel new promising BGCs. Interestingly, according to the Nagoya protocol, the strains from this collection do belong to the project partner since they have been collected in France.

As a final word, I want to thank one more time my colleagues, collaborators and students for their contribution to all these works and results. There are ambitious and difficult projects ahead of me but it is still with the same pleasure that I work everyday with the objective to unlock Nature's hidden secrets…

RÉFÉRENCES

Adeolu M, Alnajar S, Naushad S & Gupta RS (2016) Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* **66,** 5575-5599. DOI: 10.1099/ijsem.0.001485

Admiraal SJ, Walsh CT & Khosla C (2001) The loading module of rifamycin synthetase is an adenylation-thiolation didomain with substrate tolerance for substituted benzoates. *Biochemistry* **40,** 6116-6123. DOI: 10.1021/bi010080z

Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Meier-Kolthoff JP, Klenk HP, Clément C, Ouhdouch Y & van Wezel GP (2016) Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev.* **80,** 1-43. DOI: 10.1128/MMBR.00019-15

Baton LA & Ranford-Cartwright LC (2005) How do malaria ookinetes cross the mosquito midgut wall? *Trends Parasitol.* **21,** 22-8. DOI: 10.1016/j.pt.2004.11.001

Baumann S, Herrmann J, Raju R, Steinmetz H, Mohr KI, Hüttel S, Harmrolfs K, Stadler M & Müller R (2014) Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed. Engl.* **53,**14605-14609. DOI: 10.1002/anie.201409964

Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J & Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417,** 141-147. DOI: 10.1038/417141a

Birch RG & Patil SS (1985a) Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli. J. Gen. Microbiol.* **131**, 1069-1075. DOI: 10.1099/00221287-131-5-1069

Birch RG & Patil SS (1985b) June 1985. Antibiotic and process for the production of thereof. US patent 4,525,354.

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T & Medema MH (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45,** W36-W41. DOI: 10.1093/nar/gkx319

Bloudoff K & Schmeing TM (2017) Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta* **1865,** 1587-1604. DOI: 10.1016/j.bbapap.2017.05.010

Boman HG, Faye I, Gudmundsson GH, Lee JY & Lidholm DA (1991) Cell-free immunity in *Cecropia*. A model system for antibacterial proteins. *Eur. J. Biochem.* **201,** 23-31. DOI: 10.1111/j.1432-1033.1991.tb16252.x

Bonmatin JM, Bonnat JL, Gallet X, Vovelle F, Ptak M, Reichhart JM, Hoffmann JA, Keppi E, Legrain M & Achstetter T (1992) Two-dimensional ^{1}H NMR study of recombinant insect defensin A in water: resonance assignments, secondary structure and global folding *J. Biomol. NMR* **2,** 235-256. DOI: 10.1007/BF01875319

Bulet P, Cociancich S, Dimarcq JL, Lambert J, Reichhart JM, Hoffmann D, Hetru C & Hoffmann JA (1991) Insect immunity. Isolation from a coleopteran insect of a novel inducible antibacterial peptide and of new members of the insect defensin family. *J. Biol. Chem.* **266,** 24520-24525.

Bulet P, Cociancich S, Reuland M, Sauber F, Bischoff R, Hegy G, Van Dorsselaer A, Hetru C & Hoffmann JA (1992) A novel insect defensin mediates the inducible antibacterial activity in larvae of the dragonfly *Aeschna cyanea* (Paleoptera, Odonata). *Eur. J. Biochem.* **209,** 977-984. DOI: 10.1111/j.1432-1033.1992.tb17371.x

Bulet P, Dimarcq JL, Hetru C, Lagueux M, Charlet M, Hegy G, Van Dorsselaer A & Hoffmann JA (1993) A novel inducible antibacterial peptide of *Drosophila* carries an O-glycosylated substitution. *J. Biol. Chem.* **268**, 14893-14897.

Casteels P, Ampe C, Jacobs F, Vaeck M & Tempst P (1989) Apidaecins: antibacterial peptides from honeybees. *EMBO J.* **8,** 2387-2391.

Challis GL, Ravel J & Townsend CA (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7,** 211-224. DOI: 10.1016/S1074-5521(00)00091-0

Champoiseau P (2006) *Xanthomonas albilineans*, l'agent causal de l'échaudure des feuilles de la canne à sucre: caractérisation et variabilité des bases génétiques du pouvoir pathogène, en Guadeloupe et dans le monde. Thèse de l'Université des Antilles et de la Guyane. 187 pages.

Charkowski A, Blanco C, Condemine G, Expert D, Franza T, Hayes C, Hugouvieux-Cotte-Pattat N, López Solanilla E, Low D, Moleleki L, Pirhonen M, Pitman A, Perna N, Reverchon S, Rodríguez Palenzuela P, San Francisco M, Toth I, Tsuyumu S, van der Waals J, van der Wolf J, Van Gijsegem F, Yang CH & Yedidia I (2012) The role of secretion systems and small molecules in soft-rot Enterobacteriaceae pathogenicity. *Annu. Rev. Phytopathol.* **50,** 425-449. DOI: 10.1146/annurev-phyto-081211-173013

Cheng B, Müller R & Trauner D (2017) Total syntheses of cystobactamids and structural confirmation of cystobactamid 919-2. *Angew. Chem. Int. Ed. Engl.* **56,** 12755-12759. DOI: 10.1002/anie.201705387

Choi SS, Kim HJ, Lee HS, Kim P & Kim ES (2015) Genome mining of rare actinomycetes and cryptic pathway awakening. *Process Biochem.* **50,** 1184-1193. DOI: 10.1016/j.procbio.2015.04.008

Cociancich S, Goyffon M, Bontems F, Bulet P, Bouet F, Menez A & Hoffmann J (1993a) Purification and characterization of a scorpion defensin, a 4kDa antibacterial peptide presenting structural similarities with insect defensins and scorpion toxins. *Biochem. Biophys. Res. Commun.* **194,** 17-22. DOI: 10.1006/bbrc.1993.1778

Cociancich S, Ghazi A, Hetru C, Hoffmann JA & Letellier L (1993b) Insect defensin, an inducible antibacterial peptide, forms voltage-dependent channels in *Micrococcus luteus*. *J. Biol. Chem.* **268,** 19239-19245.

Cociancich S (1994a) Isolement et caractérisation de peptides antibactériens inductibles chez les insectes. Thèse de l'Université Louis Pasteur, Strasbourg I. 145 pages.

Cociancich S, Dupont A, Hegy G, Lanot R, Holder F, Hetru C, Hoffmann JA & Bulet P (1994b) Novel inducible antibacterial peptides from a hemipteran insect, the sap-sucking bug *Pyrrhocoris apterus*. *Biochem. J.* **300,** 567-575. DOI: 10.1042/bj3000567

Cociancich S, Bulet P, Hetru C & Hoffmann JA (1994c) The inducible antibacterial peptides of insects. *Parasitol. Today* **10,** 132-139. DOI: 10.1016/0169-4758(94)90260-7

Cociancich SO, Park SS, Fidock DA & Shahabuddin M (1999) Vesicular ATPase-overexpressing cells determine the distribution of malaria parasite oocysts on the midguts of mosquitoes. *J. Biol. Chem.* **274,** 12650-12655. DOI: 10.1074/jbc.274.18.12650

Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll J, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M & Süssmuth RD (2015). The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* **11**, 195-197. DOI: 10.1038/nchembio.1734

Cociancich S, Royer M & Rott P (2017) Un antibiotique découvert dans la canne à sucre. La Recherche, n°524, pp 62-65.

Conti E, Stachelhaus T, Marahiel MA & Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16,** 4174-4183. DOI: 10.1093/emboj/16.14.4174

Cuénot L (1896) Etudes physiologiques sur les orthoptères. *Arch. Biol.* **14,** 293-341.

Darrasse A, Carrere S, Barbe V, Boureau T, Arrieta-Ortiz M, Bonneau S, Briand M, Brin C, Cociancich S, Durand K, Fouteau S, Gagnevin L, Guerin F, Guy E, Indiana A, Koebnik R, Lauber E, Munoz A, Noël L, Pieretti I, Poussier S, Pruvost O, Robène-Soustrade I, Rott P, Royer M, Serres-Giardi L, Szurek B, van Sluys MA, Verdier V, Vernière C, Arlat, M, Manceau C & Jacques MA (2013) Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* **14**, 761. DOI: 10.1186/1471-2164-14-761

Demain AL (2014) Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.* **41,** 185-201. DOI: 10.1007/s10295-013-1325-z

Dimarcq JL, Keppi E, Dunbar B, Lambert J, Reichhart JM, Hoffmann D, Rankine SM, Fothergill JE & Hoffmann JA (1988) Insect immunity. Purification and characterization of a family of novel inducible antibacterial proteins from immunized larvae of the dipteran *Phormia terranovae* and complete amino-acid sequence of the predominant member, diptericin A. *Eur. J. Biochem.* **171,** 17-22. DOI: 10.1111/j.1432-1033.1988.tb13752.x

Drake EJ, Nicolai DA & Gulick AM (2006) Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem. Biol.* **13,** 409-419. DOI: 10.1016/j.chembiol.2006.02.005

Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Allen CL, Skiniotis G, Aldrich CC & Gulick AM (2016) Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529,** 235-238. DOI: 10.1038/nature16163

Du L, Sánchez C, Chen M, Edwards DJ & Shen B (2000) 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.* **7,** 623-642. DOI: 10.1016/S1074-5521(00)00011-9

Feifel SC, Schmiederer T, Hornbogen T, Berg H, Süssmuth RD & Zocher R (2007) *In vitro* synthesis of new enniatins: probing the alpha-D-hydroxy carboxylic acid binding pocket of the multienzyme enniatin synthetase. *ChemBioChem* **8,** 1767-1770. DOI: 10.1002/cbic.200700377

Garcie C, Tronnet S, Garénaux A, McCarthy AJ, Brachmann AO, Pénary M, Houle S, Nougayrède JP, Piel J, Taylor PW, Dozois CM, Genevaux P, Oswald E & Martin P (2016) The bacterial stress-responsive Hsp90 chaperone (HtpG) is required for the production of the genotoxin colibactin and the siderophore yersiniabactin in *Escherichia coli*. *J. Infect. Dis.* **214,** 916-924. DOI: 10.1093/infdis/jiw294

Glaser RW (1918) On the existence of immunity principles in insects. *Psyche* **25,** 39-46.

Graf M, Mardirossian M, Nguyen F, Seefeldt AC, Guichard G, Scocchi M, Innis CA & Wilson DN (2017) Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.* **34,** 702-711. DOI: 10.1039/c7np00020k

Grandclément C, Tannières M, Moréra S, Dessaux Y & Faure D (2016) Quorum quenching: role in nature and applied developments. *FEMS Microbiol. Rev.* **40,** 86-116. DOI: 10.1093/femsre/fuv038

Grätz S, Kerwat D, Kretz J, von Eckardstein L, Semsary S, Seidel M, Kunert M, Weston JB & Süssmuth RD (2016) Synthesis and antimicrobial activity of albicidin derivatives with variations of the central cyanoalanine building block. *ChemMedChem* **11,** 1499-1502. DOI: 10.1002/cmdc.201600163

Grgurina I & Mariotti F (1999) Biosynthetic origin of syringomycin and syringopeptin 22, toxic secondary metabolites of the phytopathogenic bacterium *Pseudomonas syringae* pv. *syringae*. *FEBS Lett.* **462,** 151-154. DOI: 10.1016/S0014-5793(99)01528-8

Han YS, Thompson J, Kafatos FC & Barillas-Mury C (2000) Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. *EMBO J.* **19,** 6030-6040. DOI: 10.1093/emboj/19.22.6030

Hashimi SM, Wall MK, Smith AB, Maxwell A & Birch RG (2007) The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* **51**, 181-187. DOI: 10.1128/aac.00918-06

Helman Y & Chernin L (2015) Silencing the mob: disrupting quorum sensing as a means to fight plant disease. *Mol. Plant. Pathol.* **16,** 316-329. DOI: 10.1111/mpp.12180

Holak TA, Engström A, Kraulis PJ, Lindeberg G, Bennich H, Jones TA, Gronenborn AM & Clore GM (1988) The solution conformation of the antibacterial peptide cecropin A: a nuclear magnetic resonance and dynamical simulated annealing study. *Biochemistry* **27,** 7620-7629. DOI: 10.1021/bi00420a008

Huang G, Zhang L & Birch RG (2001) A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans*. *Microbiology* **147,** 631-462. DOI: 10.1099/00221287-147-3-631

Huang J, Lih CJ, Pan KH & Cohen SN (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes Dev.* **15,** 3183-3192. DOI: 10.1101/gad.943401

Hugouvieux-Cotte-Pattat N, Condemine G, Nasser W & Reverchon S (1996) Regulation of pectinolysis in *Erwinia chrysanthemi*. *Annu. Rev. Microbiol.* **50,** 213-257.

Hugouvieux-Cotte-Pattat N (2016) Metabolism and Virulence Strategies in *Dickeya*-Host Interactions. *Prog. Mol. Biol. Transl. Sci.* **142,** 93-129. DOI: 10.1016/bs.pmbts.2016.05.006

Hultmark D, Engström A, Andersson K, Steiner H, Bennich H & Boman HG (1983) Insect immunity. Attacins, a family of antibacterial proteins from *Hyalophora cecropia*. *EMBO J.* **2,** 571-576.

Hüttel S, Testolin G, Herrmann J, Planke T, Gille F, Moreno M, Stadler M, Brönstrup M, Kirschning A & Müller R. (2017) Discovery and total synthesis of natural cystobactamid derivatives with superior activity against Gram-negative pathogens. *Angew. Chem. Int. Ed. Engl.* **56,** 12760-12764. DOI: 10.1002/anie.201705913

Imler JL (2014) Overview of *Drosophila* immunity: a historical perspective. *Dev. Comp. Immunol.* **42,** 3-15. DOI: 10.1016/j.dci.2013.08.018

Jackson SA, Crossman L, Almeida EL, Margassery LM, Kennedy J & Dobson ADW (2018) Diverse and abundant secondary metabolism biosynthetic gene clusters in the genomes of marine sponge derived *Streptomyces* spp. isolates. *Mar. Drugs* **16,** 67. DOI: 10.3390/md16020067

Kim HU, Blin K, Lee SY & Weber T (2017) Recent development of computational resources for new antibiotics discovery. *Current Opin. Microbiol.* **39,** 113-120. DOI: 10.1016/j.mib.2017.10.027

Klementz D, Döring K, Lucas X, Telukunta KK, Erxleben A, Deubel D, Erber A, Santillana I, Thomas OS, Bechthold A & Günther S (2015) StreptomeDB 2.0--an extended resource of natural products produced by streptomycetes. *Nucleic Acids Res.* **44,** D509-D514. DOI: 10.1093/nar/gkv1319

Kretz J, Kerwat D, Schubert V, Grätz S, Pesic A, Semsary S, Cociancich S, Royer M & Süssmuth RD (2015), Total synthesis of albicidin: a lead structure from *Xanthomonas albilineans* for potent antibacterial gyrase inhibitors. *Angew. Chem. Int. Ed. Engl.* **54**, 1969-1973. DOI: 10.1002/anie.201409584

Lambert J, Keppi E, Dimarcq JL, Wicker C, Reichhart JM, Dunbar B, Lepage P, Van Dorsselaer A, Hoffmann J, Fothergill J & Hoffmann D (1989) Insect immunity: isolation from immune blood of the dipteran *Phormia terranovae* of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides. *Proc. Natl. Acad. Sci. USA* **86,** 262-266. DOI: 10.1073/pnas.86.1.262

Lv M, Mohamed AA, Zhang L, Zhang P, Zhang L (2016) A family of CSαβ defensins and defensin-like peptides from the migratory locust, *Locusta migratoria*, and their expression dynamics during mycosis and nosemosis. *PLoS One* **11,** e0161585. DOI: 10.1371/journal.pone.0161585

Machado H, Tuttle RN & Jensen PR (2017) Omics-based natural product discovery and the lexicon of genome mining. *Curr. Opin. Microbiol.* **39,** 136-142. DOI: 10.1016/j.mib.2017.10.025

Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA, Toth I, Salmond G & Foster GD (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant. Pathol.* **13,** 614-629. DOI: 10.1111/j.1364-3703.2012.00804.x

Matsuyama K & Natori S (1988) Molecular cloning of cDNA for sapecin and unique expression of the sapecin gene during the development of *Sarcophaga peregrina*. *J. Biol. Chem.* **263,** 17117-17121.

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E & Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39,** W339-W346. DOI: 10.1093/nar/gkr466

Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJ, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann HW, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus EP, Barona-Gómez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P, Cheng YQ, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein PC, Doyle S, Driessen AJ, Ehling-Schulz M, Entian KD, Fischbach MA, Gerwick L, Gerwick WH, Gross H, Gust B, Hertweck C, Höfte M, Jensen SE, Ju J, Katz L, Kaysser L, Klassen JL, Keller NP, Kormanec J, Kuipers OP, Kuzuyama T, Kyrpides NC, Kwon HJ, Lautru S, Lavigne R, Lee CY, Linquan B, Liu X, Liu W, Luzhetskyy A, Mahmud T, Mast Y, Méndez C, Metsä-Ketelä M, Micklefield J, Mitchell DA, Moore BS, Moreira LM, Müller R, Neilan BA, Nett M, Nielsen J, O'Gara F, Oikawa H, Osbourn A, Osburne MS, Ostash B, Payne SM, Pernodet JL, Petricek M, Piel J, Ploux O, Raaijmakers JM, Salas JA, Schmitt EK, Scott B, Seipke RF, Shen B, Sherman DH, Sivonen K, Smanski MJ, Sosio M, Stegmann E, Süssmuth RD, Tahlan K, Thomas CM, Tang Y, Truman AW, Viaud M, Walton JD, Walsh CT, Weber T, van Wezel GP, Wilkinson B, Willey JM, Wohlleben W, Wright GD, Ziemert N, Zhang C, Zotchev SB, Breitling R, Takano E & Glöckner FO (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **11,** 625-631. DOI: 10.1038/nchembio.1890

Medema MH & Fischbach MA (2015) Computational approaches to natural product discovery. *Nat. Chem. Biol.* **11,** 639-648. DOI: 10.1038/nchembio.1884

Medema MH (2018) Computational genomics of specialized metabolism: from natural product discovery to microbiome ecology. *mSystems* **3**, e00182-17. DOI: 10.1128/mSystems.00182-17

Metalnikow S (1920) Immunité de la chenille contre divers microbes. *Compt. Rend. Soc. Biol.* **83,** 119-121.

Miyanaga A, Cieślak J, Shinohara Y, Kudo F & Eguchi T (2014) The crystal structure of the adenylation enzyme VinN reveals a unique β-amino acid recognition mechanism. *J. Biol. Chem.* **289,** 31448-31457. DOI: 10.1074/jbc.M114.602326

Nasser W, Dorel C, Wawrzyniak J, Van Gijsegem F, Groleau MC, Déziel E & Reverchon S (2013) Vfm a new quorum sensing system controls the virulence of *Dickeya dadantii*. *Environ. Microbiol.* **15,** 865-880. DOI: 10.1111/1462-2920.12049

Paillot A (1920) L'immunité acquise chez les insectes. *Compt. Rend. Soc. Biol.* **83,** 278-280.

Payne JAE, Schoppet M, Hansen MH & Cryle MJ (2017) Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis. *Mol. Biosyst.* **13,** 9-22. DOI: 10.1039/C6MB00675B

Petras D, Kerwat D, Pesic A, Hempel BF, von Eckardstein L, Semsary S, Arasté J, Marguerettaz M, Royer M, Cociancich S & Süssmuth RD (2016) The O-carbamoyl-transferase Alb15 is responsible for the modification of albicidin. *ACS Chem. Biol.* **11,** 1198-1204. DOI: 10.1021/acschembio.5b01001

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques MA, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M & Rott P (2009) The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* **10**, 616. DOI: 10.1186/1471-2164-10-616

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Couloux A, Darrasse A, Gouzy J, Jacques MA, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M, Gabriel DW, Rott P & Cociancich S (2012) Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genomics* **13**, 658. DOI: 10.1186/1471-2164-13-658

Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD & Cociancich S (2015) What makes *Xanthomonas albilineans* unique amongst xanthomonads? *Front. Plant Sci.* **6,** 289. DOI: 10.3389/fpls.2015.00289

Potrykus M, Hugouvieux-Cotte-Pattat N & Lojkowska E (2017) Interplay of classic Exp and specific Vfm quorum sensing systems on the phenotypic features of *Dickeya solani* strains exhibiting different virulence levels. *Mol. Plant Pathol.* **19,** 1238-1251. DOI: 10.1111/mpp.12614

Pupin M, Esmaeel Q, Flissi A, Dufresne Y, Jacques P & Leclère V (2016) Norine: A powerful resource for novel nonribosomal peptide discovery. *Synth. Syst. Biotechnol.* **1,** 89-94. DOI: 10.1016/j.synbio.2015.11.001

Reen FJ, Romano S, Dobson AD & O'Gara F (2015) The sound of silence: activating silent biosynthetic gene clusters in marine microorganisms. *Mar. Drugs* **13,** 4754-4783. DOI: 10.3390/md13084754

Reverchon S, Bouillant ML, Salmond G & Nasser W (1998) Integration of the quorum-sensing system in the regulatory networks controlling virulence factor synthesis in *Erwinia chrysanthemi*. *Mol. Microbiol.* **29,** 1407-1418. DOI: 10.1046/j.1365-2958.1998.01023.x

Reverchon S & Nasser W (2013) *Dickeya* ecology, environment sensing and regulation of virulence programme. *Environ. Microbiol. Rep.* **5,** 622-636. DOI: 10.1111/1758-2229.12073

Reverchon S, Muskhelisvili G & Nasser W (2016) Virulence program of a bacterial plant pathogen: the *Dickeya* model. *Prog. Mol. Biol. Transl. Sci.* **142,** 51-92. DOI: 10.1016/bs.pmbts.2016.05.005

Rosen PC & Seyedsayamdost MR (2017) Though much is taken, much abides: finding new antibiotics using old ones. *Biochemistry* **56,** 4925-4926. DOI: 10.1021/acs.biochem.7b00782

Rott PC, Costet L, Davis MJ, Frutos R & Gabriel DW (1996) At least two separate gene clusters are involved in albicidin production by *Xanthomonas albilineans. J. Bacteriol.* **178**, 4590-4596. DOI: 10.1128/jb.178.15.4590-4596.1996

Röttig M, Medema MH, Blin K, Weber T, Rausch C & Kohlbacher O (2011) NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39,** W362-W367. DOI: 10.1093/nar/gkr323

Royer M, Costet L, Vivien E, Bes M, Cousin A, Damais A, Pieretti I, Savin A, Megessier S, Viard M, Frutos R, Gabriel DW & Rott PC (2004) Albicidin pathotoxin produced by *Xanthomonas albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. *Mol. Plant Microbe Interact.* **17**, 414-427. DOI: 10.1094/mpmi.2004.17.4.414

Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin GP, Brin C, Carrere S, Gomez C, Hügelland M, Völler GH, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth RD & Cociancich S (2013) Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genomics* **14**, 658. DOI:10.1186/1471-2164-14-658

Sabri S (2016) La biosynthèse non ribosomique chez les bactéries du genre *Xanthomonas*. Thèse de Montpellier SupAgro. 167 pages.

Sabri S, Royer M, Pesic A, Petras D, Mainz A, Süssmuth RD & Cociancich S. Genome-based survey of nonribosomal peptide synthetase genes sharing fascinating features with the albicidin biosynthesis genes. In preparation. To be submitted to BMC Genomics before summer 2018.

Scholz-Schroeder BK, Hutchison ML, Grgurina I & Gross DC (2001) 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.* **14,** 336-348. DOI: 10.1094/MPMI.2001.14.3.336

Schwartz AR, Potnis N, Timilsina S, Wilson M, Patané J, Martins J Jr, Minsavage GV, Dahlbeck D, Akhunova A, Almeida N, Vallad GE, Barak JD, White FF, Miller SA, Ritchie D, Goss E, Bart RS, Setubal JC, Jones JB & Staskawicz BJ (2015) Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* **6,** 535. DOI: 10.3389/fmicb.2015.00535

Shahabuddin M (2002) Do *Plasmodium* ookinetes invade a specific cell type in the mosquito midgut? *Trends Parasitol.* **18,** 157-161.DOI: 10.1016/S1471-4922(01)02219-X

Shahabuddin M & Pimenta PF (1998) *Plasmodium gallinaceum* preferentially invades vesicular ATPase-expressing cells in *Aedes aegypti* midgut. *Proc. Natl. Acad. Sci. USA* **95,** 3385-3389.

Stachelhaus T, Mootz HD & Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6,** 493-505. DOI: 10.1016/S1074-5521(99)80082-9

Steiner H, Hultmark D, Engström A, Bennich H & Boman HG (1981) Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* **292,** 246-248. DOI: 10.1038/292246a0

Stewart EJ (2012) Growing unculturable bacteria. *J. Bacteriol.* **194,** 4151-4160. DOI: 10.1128/JB.00345-12

Süssmuth RD, Kretz J, Schubert V, Pesic A, Hügelland M, Royer M, Cociancich S, Rott P, Kerwat D & Grätz S (2014) Albicidin derivatives, their use and synthesis. US20150376120 A1, PCT/EP2014/052922, December 2015. WO2014125075 A1, PCT/EP2014/052922, August 2014

Süssmuth RD & Mainz A (2017) Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem. Int. Ed. Engl.* **56,** 3770-3821. DOI: 10.1002/anie.201609079

Toth IK, van der Wolf JM, Saddler G, Lojkowska E, Hélias V, Pirhonen M, Tsror (Lahkim) L & Elphinstone JG (2011) *Dickeya* species: an emerging problem for potato production in Europe. *Plant Pathol.* **60,** 385-399. DOI: 10.1111/j.1365-3059.2011.02427.x

Traxler MF, Watrous JD, Alexandrov T, Dorrestein PC & Kolter R (2013) Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *MBio* **4,** e00459-13. DOI: 10.1128/mBio.00459-13

Triplett LR, Verdier V, Campillo T, Van Malderghem C, Cleenwerck I, Maes M, Deblais L, Corral R, Koita O, Cottyn B & Leach JE (2015) Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie Van Leeuwenhoek* **107,** 869-881. doi: 10.1007/s10482-015-0379-5

Vieweg L, Kretz J, Pesic A, Kerwat D, Grätz S, Royer M, Cociancich S, Mainz A & Süssmuth RD (2015) The albicidin resistance factor AlbD is a serine endopeptidase that hydrolyzes unusual oligoaromatic-type peptides. *J. Am. Chem. Soc.* **137,** 7608-7611. DOI: 10.1021/jacs.5b04099

Vivien E, Megessier S, Pieretti I, Cociancich S, Frutos R, Gabriel DW, Rott PC, Royer M (2005) *Xanthomonas albilineans* HtpG is required for biosynthesis of the antibiotic and phytotoxin albicidin. *FEMS Microbiol. Lett.* **251,** 81-89. DOI: 10.1016/j.femsle.2005.07.026

Vivien E, Pitorre D, Cociancich S, Pieretti I, Gabriel DW, Rott PC & Royer M (2007) Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* **51**, 1549-1552. DOI: 10.1128/aac.01450-06

von Eckardstein L, Petras D, Dang T, Cociancich S, Sabri S, Grätz S, Kerwat D, Seidel M, Pesic A, Dorrestein PC, Royer M, Weston JB & Süssmuth RD (2017) Total synthesis and biological assessment of novel albicidins discovered by mass spectrometric networking. *Chemistry* **23,** 15316-15321. DOI: 10.1002/chem.201704074

Washio K, Lim SP, Roongsawang N & Morikawa M (2010) Identification and characterization of the genes responsible for the production of the cyclic lipopeptide arthrofactin by *Pseudomonas* sp. MIS38. *Biosci. Biotechnol. Biochem.* **74,** 992-999. DOI: 10.1271/bbb.90860

Winn M, Fyans JK, Zhuo Y & Micklefield J (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33,** 317-347. DOI: 10.1039/c5np00099h

Zarins-Tutt JS, Barberi TT, Gao H, Mearns-Spragg A, Zhang L, Newman DJ & Goss RJ (2016) Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat. Prod. Rep.* **33,** 54-72. doi: 10.1039/c5np00111k

Zhang L & Birch RG (1997) The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans*. *Proc. Natl. Acad. Sci. USA* **94,** 9984-9989.

RESUME DU DOCUMENT EN FRANÇAIS

De la valeur ajoutée de la génétique pour la découverte
et la caractérisation de nouveaux produits naturels
synthétisés par les voies de synthèse peptidique non ribosomique
et de synthèse polycétidique chez les bactéries associées aux plantes.

PREAMBULE

Ceci est un résumé en français du document anglais. Limité à 15 pages, je n'y inclus pas les figures. En ce qui concerne les valorisations et les encadrements d'étudiants, je renvoie le lecteur vers les tableaux récapitulatifs à la fin de chaque grande partie dans la version anglaise (pp 31, 35 et 68).

INTRODUCTION

Pour tout chercheur, l'Habilitation à Diriger des Recherches est le sésame permettant de diriger officiellement des doctorants. La rédaction d'un mémoire et une soutenance doivent démontrer les compétences scientifiques du candidat, ses qualités d'encadrement d'étudiants ainsi que sa capacité à développer des projets de recherche. Il est donc temps pour moi de faire un bilan de ma carrière et d'imaginer son futur…

Entre 1990 et 1999, au cours de mes DEA, thèse et stages post-doctoraux, j'ai travaillé sur plusieurs projets centrés sur des interactions hôtes/pathogènes (par exemple la réaction immunitaire des insectes en réponse à des agressions bactériennes, le développement de *Plasmodium* dans son vecteur moustique, ou l'activité insecticide d'une protéine de pois). Pour tous ces projets, à une époque où l'on ne jurait que par la biologie moléculaire dans les laboratoires, j'ai développé des méthodes biochimiques qui ont efficacement complété les techniques moléculaires et génétiques, et qui m'ont permis d'isoler et de caractériser la structure des différentes molécules d'intérêt étudiées. J'ai également bénéficié, au cours de cette période, de formations en électrophysiologie et en biologie moléculaire.

En 2003, j'ai rejoint l'équipe dirigée par Philippe Rott à l'UMR BGPI du Cirad. Le modèle d'étude de l'équipe était la bactérie *Xanthomonas albilineans*, l'agent causal de la maladie de l'échaudure des feuilles de la canne à sucre. Une des caractéristiques de *X. albilineans* est la production d'albicidine, une phytotoxine qui possède également une très puissante activité antibactérienne et dont la structure n'avait toujours pas été dévoilée plus de 20 ans après sa découverte. L'albicidine est produite par la voie de synthèse peptidique non-ribosomique. Il s'agit d'une particularité des bactéries et des champignons, et correspond à la biosynthèse enzymatique (indépendante des ribosomes) de métabolites secondaires, qui sont des produits naturels présentant des activités biologiques diverses et possédant des structures en général complexes. Cette voie de biosynthèse fait intervenir de grandes séquences génomiques qui contiennent des gènes codant des méga-enzymes multimodulaires responsables de la biosynthèse du squelette de métabolites peptidiques ou polycétidiques. Ces gènes de biosynthèse sont souvent regroupés avec d'autres gènes impliqués dans la modification, le transport, la régulation ou la résistance à ces métabolites, et forment ainsi des groupements de gènes biosynthétiques. L'analyse *in silico* de l'organisation génomique et de la séquence de ces groupements de gènes fournit des informations utiles pour prédire les voies de biosynthèse et les caractéristiques structurales des métabolites secondaires correspondants. La caractérisation complète de molécules aussi complexes que les métabolites secondaires exige des compétences dans l'analyse *in silico* des domaines fonctionnels en

complément des connaissances obtenues par les méthodes de biochimie et de biologie moléculaire pour permettre leur sur-production et aboutir à la description de leur structure.

Pendant plus de dix ans, j'ai étudié l'albicidine de manière extensive. Ce travail, combinant des données de prédiction obtenues par des analyses *in silico* du groupement de gènes de biosynthèse avec des données expérimentales en biochimie et en biologie moléculaire aboutissant à la description de la structure complète de l'albicidine, est donc une preuve de concept de l'importance de ces données prédictives issues d'analyses *in silico*. Ce travail m'a permis d'acquérir une certaine expérience de ces systèmes de synthèse peptidique non-ribosomique, et je développe actuellement un projet basé sur une approche identique pour l'étude d'une molécule signal de quorum sensing chez les bactéries phytopathogènes du genre *Dickeya*.

Comme cela a été le cas pour l'albicidine, la méthode classique employée jusqu'à la fin du 20ème siècle pour découvrir de nouveaux métabolites secondaires était basée sur le criblage d'activités biologiques spécifiques. Toutefois, il est possible à présent d'utiliser les connaissances obtenues par l'analyse *in silico* du nombre exponentiellement croissant de génomes microbiens disponibles pour détecter la présence de groupements de gènes de biosynthèse de métabolites secondaires encore inconnus. Ces analyses *in silico* permettent également de prédire certaines caractéristiques des produits naturels codés par ces groupements de gènes. Cette approche, dite de minage de génomes ("genome mining") permet notamment de révéler le vrai potentiel métabolique de micro-organismes qui apparait comme totalement sous-estimé par l'approche classique basée sur le criblage des activités biologiques. En effet, certains de ces groupements de gènes de biosynthèse peuvent éventuellement être silencieux ou tout du moins non exprimés dans les conditions habituelles de laboratoire. L'analyse *in silico* de ces nouveaux groupements de gènes de biosynthèse ainsi découverts ne permet qu'un décryptage partiel de la voie de biosynthèse et qu'une prédiction partielle de la structure des produits synthétisés. Le minage de génome doit donc être considéré comme une approche prédictive requise pour déchiffrer la structure de nouveaux produits naturels, et pour déterminer leur voie de biosynthèse complexe.

Je participe actuellement et développe plusieurs projets basés sur une approche de minage de génomes visant à cribler des génomes de bactéries du genre *Xanthomonas* (projets "Meta-B" et "XanthoNRPS") et du genre *Streptomyces* (projet "CONTACT") afin de découvrir de nouveaux métabolites secondaires synthétisés par la voie non-ribosomique de synthèse peptidique et possédant des caractéristiques particulières. Ces projets font appel à des approaches multi-disciplinaires qui impliquent le minage de génomes puis une analyse *in silico* avant la caractérisation expérimentale de la structure des métabolites secondaires. Ces projets devraient produire des connaissances fondamentales et aussi proposer de potentielles applications comme le développement d'antibiotiques ou autres produits intéressant l'industrie de la chimie.

Au cours de ma carrière, j'ai co-encadré plusieurs étudiants en master ou en thèse, et j'ai récemment co-dirigé une thèse sur la synthèse peptidique non ribosomique chez les bactéries du genre *Xanthomonas*.

Ce document représente une synthèse de mes travaux passés et de mes projets actuels et futurs.

PARTIE 1. 1990-1999 – DEA/THÈSE/STAGE POST-DOCTORAL

I. Isolement et caractérisation de peptides antibactériens inductibles chez les insectes

Master/Thèse (UPR Réponse Immunitaire et Développement chez les Insectes – Direction : Jules Hoffmann)
Strasbourg, 1990-1994

Depuis le début du 20ème siècle, on sait que les insectes résistent de manière très efficace aux infections bactériennes grâce à l'induction d'une activité antibactérienne dans l'hémolymphe. Mais ce n'est que 60 ans plus tard que le premier composé antibactérien a été décrit. Il s'agit de la cécropine, un peptide antibactérien inductible isolé chez un Lépidoptère, et dont l'activité est dirigée contre les germes à Gram négatif. Par la suite, d'autres peptides antibactériens ont été isolés chez des Diptères, Lépidoptères et Hyménoptères (insectes récents dans l'évolution). Ces peptides antibactériens ont été classés en quatre familles : les cécropines (anti-Gram négatif, 4 kDa), les défensines d'insectes (anti-Gram positif, 4 kDa, six cystéines engagées dans trois ponts disulfure intra-moléculaires générant une structure avec une hélice α stabilisée sur un feuillet β par deux ponts disulfure [motif CS$\alpha\beta$]), les grands peptides riches en glycine (anti-Gram négatif, 9-22 kDa) et les petits peptides riches en proline (anti-Gram négatif, 2-4 kDa, jusqu'à un tiers de prolines).

Au cours de mon DEA, puis de ma thèse, j'ai isolé une vingtaine de nouveaux peptides antibactériens inductibles chez divers ordres d'insectes (incluant les Odonates, très anciens dans l'évolution). Ces peptides entrent tous dans les quatre familles citées plus haut. La famille la plus abondante que j'ai isolée est celle des défensines d'insectes, y compris chez l'espèce très ancienne *Aeschna cyanea* (Odonate) et, de manière intéressante, chez une espèce de scorpion (Arachnides). J'ai également isolé chez deux espèces de punaises de nombreuses isoformes de petits peptides riches en proline, caractérisés par une association répétée de triplets proline-arginine-proline. Pour l'un d'entre eux, la pyrrhocoricine, j'ai montré qu'une des prolines était O-glycosylée, et que cette substitution était nécessaire à son activité biologique.

Dans le cadre d'une collaboration avec Lucienne Letellier (laboratoire des biomembranes à Orsay), j'ai décrit le mode d'action d'une défensine d'insecte au niveau de sa cible bactérienne grâce à des techniques d'électrophysiologie. J'ai ainsi pu démontrer que la défensine d'insecte provoque un efflux total et rapide (moins de deux minutes) du potassium bactérien et provoque une diminution (mais pas une disparition) du potentiel transmembranaire. Cet efflux est inhibé par des ions divalents. Une expérience de patch-clamp réalisée sur des liposomes géants a permis d'observer des événements d'ouverture et de fermeture de canaux ioniques. Toutes ces données convergent vers la conclusion que les défensines d'insectes forment des canaux voltage-dépendants dans la membrane bactérienne des bactéries à Gram positif.

Au cours de ma thèse, j'ai "encadré" quotidiennement Pascale Fehlbaum, une doctorante arrivée après moi dans l'équipe.

II. Etude de l'interaction de *Plasmodium* avec le moustique

Stage post-doctoral – NIH – Laboratory of Parasitic Diseases – Direction: Louis H. Miller
Bethesda, MD, USA, 1995-1998

A la fin du 20ème siècle, le paludisme est toujours responsable de plus d'un million de morts par an. Aucun vaccin n'est encore disponible, et des parasites résistants aux traitements apparaissent. Le développement de *Plasmodium* consiste en un cycle complexe avec un passage obligatoire chez le moustique vecteur. L'idée de contrôler le développement de *Plasmodium* en manipulant la compétence de son vecteur a fait son chemin, mais requiert davantage de connaissances quant aux interactions *Plasmodium*/moustique. Au sein même du moustique, *Plasmodium* opère une série de transformations (gamètes, zygotes, ookinètes mobiles qui traversent la paroi de l'intestin, oocystes fixés sur la paroi de l'intestin, sporozoïtes qui envahissent les glandes salivaires).

Seules certaines espèces de moustiques transmettent *Plasmodium*. Je me suis intéressé à cette différence de susceptibilité entre les espèces qui pourrait être expliquée par le rôle potentiel de certains facteurs de moustiques sur le développement du parasite. En utilisant le modèle *P. gallinaceum* (pour des raisons évidentes de sécurité), j'ai montré que c'est l'étape de transformation entre zygotes et ookinètes qui a le plus faible taux de réalisation chez des moustiques réfractaires. J'ai analysé par HPLC des extraits totaux de moustiques infectés par *P. gallinaceum*. Les fractions obtenues ont été testées *in vitro* pour leur capacité à inhiber la transformation des zygotes en ookinètes. J'ai détecté une activité positive et ai réussi à purifier la molécule responsable de cette activité. La masse de la molécule a été déterminée ($[M+H]^+$ = 503,2 Da) et une séquence peptidique partielle a pu être obtenue. Malheureusement, de trop faibles quantités de molécule purifiée et la fin de mon stage ont eu raison de la détermination complète de la structure de cette molécule.

Parallèlement, j'ai développé un second projet qui m'a permis de me former aux techniques de biologie moléculaire. L'observation que la répartition des oocystes formés par *Plasmodium* est biaisée vers la partie postérieure de l'intestin du moustique a toujours été expliquée par le fait que le moustique se tient au repos verticalement pendant plusieurs heures après un repas de sang. Or, des moustiques ayant été placés après la prise d'un repas de sang pendant plusieurs heures dans un appareil rotatif qui les empêche de rester en position verticale présentent toujours la même distribution biaisée des oocystes. L'observation de la présence plus importante dans la même partie postérieure de l'intestin de cellules intestinales particulières—les cellules de Ross—qui, notamment, sur-expriment une ATPase vésiculaire, nous a amenés à faire l'hypothèse que les ookinètes traverseraient l'intestin spécifiquement au travers de ces cellules avant de se transformer en oocystes. J'ai donc cloné le gène codant une sous-unité de l'ATPase vésiculaire afin de l'exprimer en système hétérologue et de pouvoir produire un anticorps dirigé contre elle. J'ai ainsi pu observer une co-localisation entre les cellules de Ross et des ookinètes.

Au cours de cette période, j'ai encadré un étudiant de niveau licence lors de son stage d'été, ainsi que trois autres post-doctorants arrivés après moi au laboratoire.

PARTIE 2. 2003-...: TRAVAUX SCIENTIFIQUES A L'UMR BGPI: TRAVAUX PASSES, PROJETS ACTUELS ET FUTURS

Contribution de la génétique à la caractérisation et à la découverte de produits naturels synthétisés par les voies de sythèse peptidique non ribosomique et de synthèse polycétidique chez les bactéries associées aux plantes

Introduction

Les métabolites secondaires microbiens sont des produits naturels qui présentent des activités biologiques importantes (antibiotiques, antifongiques, antiprolifératifs, sidérophores, etc.). La plupart de ces composés sont produits par des mécanismes enzymatiques non ribosomiques qui sont capables de générer des structures infiniment complexes. Pendant des siècles, de très nombreux métabolites secondaires ont été découverts sur la base de cribles réalisés sur les activités biologiques d'intérêt. Mais vers la fin du 20ème siècle, l'intérêt de la communauté scientifique pour ce type de crible a disparu, notamment à cause de très nombreuses re-découvertes de molécules déjà connues. La source de métabolites secondaires semblait tarie. Mais un regain d'intérêt a récemment fait suite à l'augmentation exponentielle de séquences de génomes bactériens disponibles qui montrent en réalité que le potentiel métabolique des micro-organismes a jusqu'à présent été très sous-évalué. Le challenge actuel réside cependant dans l'exploitation de ce potentiel.

En rejoignant l'équipe de Philippe Rott à l'UMR BGPI, je me suis intéressé à ces métabolites secondaires non ribosomiques chez les bactéries associées aux plantes. Bien que biochimiste de formation, j'ai très vite compris qu'une approche biochimique serait insuffisante pour isoler et caractériser ces molécules complexes, mais également qu'il était possible de tirer des informations prédictives de l'analyse *in silico* des gènes de biosynthèse de ces métabolites. Dans ce document, je souhaite donc montrer, au travers de mes travaux passés et de mes projets actuels et futurs, (i) comment les données des analyses *in silico* des gènes de biosynthèse non ribosomique peuvent contribuer, en complément d'approches biochimiques et de biologie moléculaire, à la sur-production et à la caractérisation structurale de produits naturels déjà découverts, et (ii) comment le minage de génomes *[note: dans la suite du texte, je conserverai l'anglicisme "genome mining"]* permet de découvrir de nouveaux métabolites secondaires par le criblage de génomes pour la présence de groupements de gènes de biosynthèse encore inconnus.

Après une decription synthétique des mécanismes de la synthèse peptidique non ribosomique, je détaillerai dans une première partie mon travail de longue haleine de caractérisation structurale de la phytotoxine albicidine produite par la bactérie modèle de l'équipe, *Xanthomonas albilineans*, ainsi qu'un nouveau projet visant à caractériser la structure d'une molécule signal du quorum sensing (QS) jouant un rôle très important dans la pathogénie des bactéries phytopathogènes du genre *Dickeya*. J'y montrerai l'apport important des analyses *in silico*. Dans une seconde partie, je décrirai l'approche par "genome mining" pour découvrir de nouveaux métabolites secondaires. Cette partie sera illustrée par trois projets en cours ou futurs sur un métabolite secondaire découvert dant le génome de *X. albilineans* (Meta-B) et sur le criblage de collections de bactéries du genre *Xanthomonas* (projet "XanthoNRPS") et du genre *Streptomyces* ("Projet CONTACT").

Tous ces travaux et projets font l'objet de collaborations avec plusieurs équipes, et impliquent de nombreux étudiants dont les contributions seront mentionnées dans le document.

Chapitre 1: Contribution des analyses *in silico* de gènes de biosynthèse à la caractérisation de la structure de produits naturels déjà découverts

I. La synthèse peptidique non ribosomique

Tous les organismes vivants utilisent les ribosomes pour synthétiser leurs protéines et peptides. Depuis l'élucidation du code génétique, on peut associer les codons de trois bases de l'ARN messager à l'un des 20 α-L acides aminés (aas) protéinogènes. Toutefois, de nombreuses bactéries et champignons possèdent un système de synthèse peptidique supplémentaire, indépendant des ribosomes, qui fait intervenir de gigantesques enzymes modulaires (appelées NRPS pour Non Ribosomal Peptide Synthetases) catalysant l'incorporation successive par chacun de leurs modules d'une "brique" au sein d'une chaîne peptidique en croissance. Cette brique peut être l'un des 20 aas protéinogènes, mais également l'un des plus de 500 substrats portant un groupement amine et un groupement carboxyl décrits à ce jour (aas non-protéinogènes).

Les NRPSs possèdent une architecture multimodulaire et fonctionnent globalement comme une chaîne d'assemblage dont chaque module est divisé en domaines qui possèdent chacun une fonction bien précise. Chaque module contient au minimum trois domaines "obligatoires" qui sont le domaine d'adénylation (A), le domaine de thiolation (T, souvent appelé PCP) et le domaine de condensation (C) avec une organisation canonique par module du type [C-A-T]. Le domaine A est responsable de l'activation spécifique par adénylation d'un substrat aa (brique). La spécificité est liée à une séquence-signature de dix aas répartis au sein de la séquence peptidique du domaine A, et dont les chaînes latérales pointent toutes vers une poche où elles stabilisent le substrat correspondant. Il existe donc un code qui permet de prédire quel substrat sera incorporé par un domaine A. Une fois activé, le substrat est transféré par thiolation sur un bras pPant préalablement fixé sur le domaine T par une 4'-phosphopanthéteinyl transférase (PPTase) qui est une enzyme requise pour l'activation post-traductionnelle des systèmes NRPS. Ensuite, le domaine C du module suivant catalyse la liaison peptidique entre le substrat porté par le module n et celui porté par le module n-1. Cette opération se poursuit de module en module jusqu'au dernier qui possède un domaine supplémentaire appelé thio-estérase (TE) qui libère le peptide néoformé. La diversité de substrats alliée à l'action d'éventuels domaines optionnels supplémentaires (méthylation, oxydation, réduction, cyclisation, etc.) et celle des produits de gènes présents dans le groupement de gènes biosynthétiques permettant de décorer (hydroxylations, par exemple) le peptide néo-formé participent à l'extraordinaire complexité des structures ainsi synthétisées.

Il existe une autre famille de méga-enzymes ressemblant aux NRPSs: il s'agit des polycétides synthases (PKSs) qui sont globalement basées sur le même modèle, à la différence que leurs modules incorporent des acyl-thioesters. Certains gènes de biosynthèse codent à la fois des modules PKS et des modules NRPS, permettant ainsi la biosynthèse d'hybrides polycétides/peptides.

II. Approches multi-disciplinaires pour la caractérisation de l'albicidine, un antibiotique très puissant découvert en 1985

Une des caractéristiques de *X. albilineans* est la production de l'albicidine. Il s'agit d'une phytotoxine découverte en 1985 par Robert Birch (Université du Queensland, Australie) et qui est responsable de l'apparition de lignes blanches sur les feuilles de canne à sucre (symptômes de la maladie de l'échaudure des feuilles de la canne à sucre). Il a également été montré à cette époque que l'albicidine possède une activité antibiotique très puissante car elle est active à des concentrations nanomolaires contre *E. coli*. En réalité, l'albicidine est un inhibiteur de l'ADN gyrase qui agit par un mécanisme différent des autres inhibiteurs de gyrase actuellement sur le marché, telles les fluoroquinolones. Dans les années 1980, Birch et ses collègues ont réussi à purifier l'albicidine (en fait l'analogue majoritaire, appelé β-albicidine, d'une famille de molécules)

et l'ont analysée par spectrométrie de masse (MS) et par résonance magnétique nucléaire (RMN). S'ils ont pu déterminer la masse à 842 Da, ils n'ont tiré que des informations très fragmentaires des spectres RMN.

A la fin des années 1990 et au début des années 2000, le groupe de Philippe Rott au Cirad a réussi à décrire l'ensemble des gènes de biosynthèse de l'albicidine. Ils sont regroupés dans deux régions du génome de la bactérie. La première région comprend un groupement de 20 gènes dont trois codent des NRPSs et des PKSs. Les autres gènes sont des gènes de modification, de sécrétion, de régulation et de résistance. La seconde région correspond au gène codant la PPTase nécessaire à l'activation du système. L'analyse *in silico* des gènes NRPS/PKS a permis à l'équipe de proposer une structure théorique du squelette de l'albicidine qui serait donc constituée d'une partie polycétidique, suivie d'un substrat inconnu 1, d'une asparagine, du même substrat inconnu 1 et de deux copies d'un même substrat inconnu 2. Par analogie avec le système NRPS de biosynthèse de la rifamycine, il a été proposé, au vu des séquences-signatures des domaines A concernés, que les substrats inconnus 1 et 2 pourraient être des dérivés de l'acide *para*-aminobenzoïque (*p*ABA). Cette hypothèse est confortée par la présence dans le groupement de gènes de biosynthèse de l'albicidine de gènes codant des enzymes impliquées dans le métabolisme du *p*ABA.

Une autre particularité de la voie de biosynthèse de l'albicidine concerne l'asparagine prédite pour être incorporée dans la partie centrale de la molécule. Le domaine A du module correspondant est en réalité non fonctionnel en raison d'une délétion, et il a été proposé que ce domaine A soit *trans*-complémenté par le produit d'un des trois gènes NRPS qui code un seul module formé d'un domaine A prédit pour incorporer une asparagine. De manière surprenante, ce domaine A contient une séquence supplémentaire de plus de 300 aas qui possède un site de fixation de l'ATP et qui présente une homologie avec les membres de la famille des adénosine nucléotide α-hydrolases (α-ANH-like).

J'ai rejoint l'équipe à ce moment-là, et dans le cadre de la thèse d'Eric Vivien, nous avons identifié une troisième région du génome nécessaire pour la biosynthèse de l'albicidine. Cette région correspond au gène codant la protéine chaperone HtpG.

Depuis sa découverte, de par son activité antibiotique puissante, l'albicidine est considérée comme un candidat prometteur pour un développement de médicament. La prédiction de la structure théorique, de par son originalité supposée, renforce cet intérêt pharmaceutique. La caractérisation de la structure de la molécule devenait donc une priorité. Les premiers essais de production et de purification se sont heurtés au rendement extrêmement faible de l'albicidine par *X. albilineans*. Dans le cadre de la thèse d'Eric Vivien, nous avons donc développé un système d'expression hétérologue de l'albicidine en transferrant, à l'aide de deux plasmides, l'ensemble des gènes de biosynthèse dans une souche à croissance rapide de *Xanthomonas axonopodis* pv. *vesicatoria* (on sait aujourd'hui qu'il s'agit en réalité de *X. perforans*) qui possède l'avantage de pousser en milieu minimum, ce qui est idéal pour les étapes de purification. Nous avons ainsi montré que la production d'albicidine par l'hôte hétérologue est multipliée par un facteur de 6 à 10, et que la molécule synthétisée par l'hôte hétérologue se comporte de la même façon que l'albicidine naturelle dans des tests impliquant des souches d'*E. coli* résistantes à l'albicidine et dans des expériences de chromatographie en couche mince.

Pour la purification de l'albicidine, j'ai adapté le protocole historique de Robert Birch en rajoutant une résine dans le milieu de culture minimum utilisé afin d'y adsorber les métabolites sécrétés par les bactéries. J'ai appliqué aux extraits méthanoliques de cette résine le protocole de Robert Birch, et ai finalement (après plusieurs années), réussi à purifier par HPLC 3 mg de l'analogue majoritaire β-albicidine à partir de plus de 500 litres de culture bactérienne. Dans le cadre d'une collaboration initiée en 2005 (et qui dure toujours), les analyses structurales ont été réalisées dans l'équipe de Roderich Süssmuth, chimiste à l'Université Technique de Berlin, spécialisé dans l'étude de produits naturels. L'analyse en spectrométrie de masse indique une masse de 842,26 Da (identique à celle mesurée par Robert Birch presque 30 ans plus tôt). En considérant que l'albicidine n'est composée que d'atomes de C, O, N et H, plusieurs formules moléculaires correspondent à

cette masse. Nous avons donc eu l'idée de purifier de l'albicidine produite par des bactéries ayant poussé dans un milieu supplémenté en ^{15}N (isotope non-radioactif de l'azote, dont l'incorporation de chaque atome dans une molécule rajoute un dalton à la masse de la molécule non marquée). J'ai purifié cette albicidine marquée qui, finalement, accuse une masse de 849,26 Da, soit 6 Da supplémentaires que l'albicidine non marquée, indiquant que la molécule contient six atomes d'azote. Dans ce cas, une seule formule moléculaire était encore physiquement possible. Des analyses supplémentaires en MS/MS et en RMN ont été réalisées. L'interprétation des spectres RMN, notamment, a été rendue difficile par de nombreux signaux superposés, indiquant des répétitions de structures au sein de la molécule. Finalement, en tenant compte de la prédiction de la présence de *p*ABA ou de dérivés du *p*ABA, les pièces du puzzle se sont mises en place, et nous avons enfin pu dévoiler la structure de l'albicidine plus de 30 ans après sa découverte.

L'albicidine est composée de six "briques" liées entre elles par cinq liaison amides (peptidiques). La première "brique" est un dérivé de l'acide *para*-coumarique, et est assemblée par les modules PKS des enzymes de biosynthèse. On trouve ensuite, dans l'ordre, un *p*ABA, une cyano-alanine, un second *p*ABA, puis deux dérivés identiques hydroxylés et méthoxylés du *p*ABA. Il s'agit donc d'une molécule extrêmement originale puisque c'est la première fois que l'on décrit la présence de *p*ABA et de cyanoalanine dans une molécule synthétisée par la voie non ribosomique. Il est intéressant de noter que cette structure est en accord avec la structure théorique proposée dix ans plus tôt, ce qui valide l'approche de prédiction *in silico*, à la différence près que l'albicidine comprend une cyano-alanine à la place de l'asparagine prédite. Mais nous formulons l'hypothèse que c'est bien une asparagine qui est incorporée par le domaine A et qui est ensuite convertie en cyano-alanine par une réaction ATP-dépendante catalysée par le domaine supplémentaire αANH-like.

Dans le cadre des stages d'une étudiante en BTS, Julie Arasté, nous avons purifié trois analogues naturels supplémentaires de l'albicidine dont les masses sont respectivement $[M+H]^+$ = 861,27 Da, 891,27 Da et 886,27 Da. Plus récemment, la réalisation par nos collègues de Berlin d'une expérience de réseau en MS leur a permis de visualiser dans des extraits de surnageants de *X. albilineans* (hôte naturel) et de l'hôte hétérologue sept analogues naturels de l'albicidine supplémentaires. Des analyses en MS/MS ont finalement permis de caractériser les structures de neuf analogues supplémentaires par rapport à la β-albicidine. Les différentes modifications sont une carbamoylation N-terminale, une troncation N-terminale, une méthylation N-terminale, le remplacement de la cyano-alanine par une asparagine, une méthoxylation de la cyano-alanine et une déhydroxylation de l'un ou l'autre des dérivés du *p*ABA. Toutes ces modifications sont peu ou prou compatibles avec les produits des gènes présents dans le groupement de gènes de biosynthèse de l'albicidine.

Nous avons également caractérisé l'enzyme AlbD, isolée chez *Panteoa dispersa*, capable d'abolir l'activité biologique de l'albicidine, et qui avait précédemment été décrite par Robert Birch comme une estérase. Nous avons montré par MS que cette enzyme clive la liaison peptidique située entre le second *p*ABA et le premier dérivé du *p*ABA. A ce titre, cette "estérase" est en réalité une endo-peptidase d'un nouveau type puisque capable de cliver une liaison peptidique entre deux δ-aas aromatiques.

Finalement, les collègues de Berlin ont développé un protocole de synthèse chimique de l'albicidine, protocole qui est protégé par un brevet dont je suis co-auteur. La synthèse chimique leur permet dorénavant de générer des centaines d'analogues structuraux et de les tester pour des propriétés biologiques et/ou physico-chimiques améliorées, dans le but de propulser l'albicidine dans le pipeline de développement d'antibiotiques.

III. Comment la génétique peut aider à la caractérisation de la molécule signal de quorum sensing VFM impliqué dans la pathogénie des bactéries phytopathogènes du genre *Dickeya*.

Notre équipe a été récemment sollicitée par Nicole Cotte-Pattat, de l'UMR MAP à Lyon, pour développer un projet portant sur nouveau système de QS découvert par son équipe il y a quelques années chez *D. dadantii*, et dont la molécule signal est synthétisée par la voie de synthèse peptidique non-ribosomique. Deux ans de tentatives de purification et de caractérisation de la molécule signal VFM ne leur ont pas permis d'élucider sa structure. Le groupement de gènes de biosynthèse est connu et comporte 26 gènes. Il a été transféré dans une souche d'*E. coli* qui permet une production hétérologue du signal. Nous avons soumis, en collaboration avec l'équipe de Lyon et notre collaborateur historique de Berlin, un projet en réponse à l'appel à projets ANR/DFG PRCI 2018, dont je suis le coordinateur pour la partie française (incluant le financement d'une thèse que je dirigerai).

Les bactéries du genre *Dickeya* sont des bactéries pectinolytiques pathogènes de plantes, provoquant des symptômes de macération. Par ailleurs, une espèce émergente en Europe, *D. solani*, cause des dommages importants dans les cultures de pommes de terre. La virulence des bactéries du genre *Dickeya* repose sur un système de QS qui permet une communication très finement régulée entre bactéries de la même espèce, et qui permet la régulation de toute une série de gènes, dont les gènes de virulence, en fonction de la densité bactérienne. Chez la plupart des bactéries, la molécule signal des systèmes de QS est une homo-sérine lactone (HSL). En plus de ce système de QS HSL, les bactéries du genre *Dickeya* possèdent un second système de QS, VFM, dont il a été montré qu'il intervient bien dans la régulation de la virulence. Nous avons réalisé une analyse *in silico* des gènes de biosynthèse de la molécule signal, et avons pu prédire que la molécule signal est un hybride polycétide/peptide, ainsi qu'un certain nombre d'éléments de structure de la molécule mais pour cause de compétition dans ce domaine, je considère ces prédictions comme confidentielles.

De manière intéressante, nous avons réalisé une analyse *in silico* du groupement de gènes de biosynthèse de VFM chez les 49 génomes de *Dickeya* disponibles dans GenBank, et avons mis en évidence un polymorphisme espèce-dépendant dans les signatures conférant la spécificité des domaines A. Donc, chaque espèce de *Dickeya* produit probablement un signal VFM légèrement différent.

En utilisant la même résine XAD-7 utilisée pour extraire l'albicidine, j'ai réussi à extraire le signal VFM à partir de surnageant de l'hôte hétérologue (souche d'*E. coli* possédant le locus Vfm). Un test biologique a permis de confirmer la présence du signal VFM dans ces extraits.

Au cours d'une expérience préliminaire, j'ai tenté de purifier, en me basant sur le protocole de purification de l'albicidine, la molécule signal VFM, mais j'ai retrouvé une faible activité VFM répartie sur de nombreuses fractions HPLC contiguës, indiquant que la chimie de la colonne utilisée n'est sans doute pas appropriée. Il apparait également qu'il va falloir sur-produire la molécule VFM. Nous proposons une stratégie consistant à sur-exprimer le régulateur trancriptionnel de la famille AraC codé par le gène *vfmE* présent dans le groupement de gènes de biosynthèse. Cela permettra l'activation transcriptionnelle de tous les gènes de biosynthèse de VFM. Nous avons déjà utilisé cette approche avec succès dans un autre projet (Meta-B). En ce qui concerne les étapes de purification et de caractérisation structurale, je développerai un protocole de purification HPLC adapté, et les techniques habituelles de MS et RMN seront utilisées par nos collaborateurs à Berlin pour la détermination de la structure de la molécule.

Chapitre 2: Contribution du “genome mining” à la découverte de nouveaux produits naturels

I. Le “genome mining”, une approche différente pour découvrir de nouveaux métabolites secondaires

Depuis des centaines d’années, des produits naturels ont été découverts grâce à leur activité biologique. Cette approche classique de criblages biologiques a permis de découvrir, par exemple, des milliers d’antibiotiques, d’antifongiques etc. Récemment, les nombreuses re-découvertes de produits déjà connus, et l’engouement pour des bibliothèques de molécules chimiques synthétiques pouvant être criblées à haut débit ont amené les chercheurs à se détourner de l’approche classique. Mais, avec l’avènement des séquençages de génomes bactériens de plus en plus massifs, sans parler des études de métagénomique, l’enthousiasme est revenu dans les rangs des chercheurs qui, en analysant ces génomes, font des découvertes inattendues. En effet, il apparait que le potentiel métabolique des bactéries est bien plus important que celui estimé avec les molécules connues, et que le réservoir de molécules nouvelles à découvrir reste énorme. Par exemple, le premier génome séquencé d’un *Streptomyces* fait apparaitre la présence d’au moins 20 groupements de gènes de biosynthèse, bien plus que le nombre de métabolites secondaires connus chez la souche séquencée. Il semble d’ailleurs que, de manière générale, les souches de *Streptomyces* peuvent posséder jusqu’à 30 groupements par génome, et constituent donc de réelles mini-usines naturelles à métabolites secondaires.

Généralement, un groupement de gènes de biosynthèse contient au minimum les gènes codant les enzymes de biosynthèse, les régulateurs et le transporteur nécessaires à la biosynthèse et à la sécrétion du métabolite secondaire. Cette caractéristique est utilisée comme crible lors des analyses de "genome mining". La structure particulière des domaines fonctionnels des NRPSs et des PKSs facilite également le repérage de groupements de gènes de biosynthèse au sein d’un génome entier. Du fait de la présence répétée de séquences similaires (multimodularité des NRPSs/PKSs), des erreurs d’assemblages des données de séquençage du génome sont toujours possibles au niveau de ces gènes, mais ces erreurs devraient être oubliées avec les séquençages de génomes faisant appel à la technologie PacBio.

De nombreuses ressources bioinformatiques ont été développées pour le “genome mining”. La plus populaire est antiSMASH, une plateforme sur le web qui identifie et annote les groupements de gènes de biosynthèse au sein d’un génome. De plus, antiSMASH est couplée à une autre ressource très pratique: NRPSpredictor2 qui permet de prédire la spécificité de substrat des domaines A. Plusieurs bases de données très utiles sont également à disposition: StreptomeDB, MIBig et NORINE, par exemple.

Comme indiqué plus haut, l’augmentation importante et rapide du nombre de génomes séquencés disponibles pour du “genome mining” débouche sur la découverte de groupements de gènes de biosynthèse en bien plus grand nombre que les métabolites secondaires connus. On peut donc formuler l’hypothèse que certains (beaucoup?) de ces groupements de gènes sont silencieux dans les conditions habituelles de laboratoire qui, c’est un fait, ne miment habituellement pas l’environnement naturel des micro-organismes. Plusieurs stratégies ont été développées pour “réveiller” ces groupements de gènes de biosynthèse silencieux. Parmi les plus utilisées, les changements de conditions de culture (température, milieux, sources de carbone, agitation, oxygénation) sont autant de facteurs qui peuvent influer sur la nature des métabolites secondaires produits (concept OSMAC: une souche, plusieurs composés). La co-culture avec d’autres micro-organismes est une autre méthode efficace. La sur-expression d’activateurs transcriptionnels et le transfert de l’ensemble du groupement de gènes de biosynthèse dans un hôte hétérologue sont autant de méthodes génétiques qui ont également fait leur preuve.

II. Caractérisation de Meta-B, un nouveau lipopeptide découvert par “genome mining” dans plusieurs espèces de *Xanthomonas* (projet Meta-B)

Le séquençage du génome de *X. albilineans* nous a permis d’identifier, en plus du groupement de gènes de l’albicidine, cinq autres groupements de gènes de biosynthèse NRPS. Nous avons réalisé une analyse *in silico* de celui qui semblait le plus complet, et donc le plus intéressant. Nous avons pu prédire que le métabolite secondaire correspondant est un lipopeptide de 16 aas, dont 12 ont pu être prédits, et dont certains sont épimérisés. Dans ce groupement de gènes de biosynthèse se trouvent également plusieurs gènes impliqués dans la biosynthèse d’un aa non protéinogène, la 3,5-dihydroxyphénylglycine, ainsi qu’entre autres un gène codant un régulateur transcriptionnel de la famille AraC. De manière intéressante, ce groupement de gènes de biosynthèse est également présent chez deux espèces de *Xanthomonas* également pathogènes de plantes monocotylédones, *X. translucens* et *X. oryzae,* respectivement pathogènes du riz et des céréales. L’organisation du groupement de gènes de biosynthèse est globalement identique dans huit souches appartenant à ces trois espèces, mais l’analyse *in silico* des gènes NRPS indique que les peptides produits sont tous différents : ils possèderaient le même groupement acyl N-terminal mais leur partie peptidique diffère en taille et en composition.

Suite à nos travaux sur l’albicidine, nous avons anticipé un faible taux de production de Meta-B, et avons fabriqué une souche de *X. albilineans* sur-productrice en clonant le gène codant le régulateur AraC dans un plasmide sous le contrôle d’un promoteur fort (le promoteur du gène HtpG). J’ai pu montrer par HPLC que cette manipulation génétique avait “réveillé” le groupement de gènes de biosynthèse et j’ai ainsi isolé la molécule. Les études en MS et MS/MS ont déterminé la masse de Meta-B (2293 Da) et permis de confirmer la séquence peptidique prédite à la différence que trois des asparagines prédites sont en réalité des hydroxy-asparagines. De manière surprenante, des mesures en MS sur des fragments de digestion tryptique de Meta-B indiquent la présence d’un 17ème aa initialement non prédit. Ce résultat a incité Souhir Sabri, dans le cadre de sa thèse, à vérifier la séquence du locus Meta-B en réalisant des profils de restriction de plusieurs clones d'une banque génomique de la souche GPE PC73 de *X. albilineans* choisie pour l'étude du peptide META-B. Le séquençage des extrémités de ces clones a confirmé la présence d'une séquence répétée supplémentaire correspondant au module manquant.

Souhir Sabri, toujours dans le cadre de sa thèse, a également construit trois mutants d’hydroxylases dans le but de comprendre le mécanisme d’hydroxylation des asparagines, mais aucun de ces mutants n’a produit de Meta-B sans hydroxy-asparagines.

A l’heure actuelle, nous connaissons la séquence peptidique de Meta-B, mais sa partie lipidique est toujours en cours d’investigation. Par ailleurs, nous ne connaissons pas non plus la fonction biologique de Meta-B. Souhir Sabri, dans le cadre de sa thèse, a réalisé des tests d’activité antibiotiques, antifongiques et sidérophores, mais tous étaient négatifs. Nous avons prévu de tester des mutants de Meta-B en les inoculant dans des cannes à sucre afin de voir si Meta-B joue un rôle dans les interactions plantes-bactéries. Mais ces expérimentations n'ont pas pû être réalisées dans le cadre de la thèse de Souhir Sabri à cause d'un problème au niveau de la serre canne à sucre, et elles seront réalisées au cours de l’été 2018.

III Comment les données de la voie de biosynthèse de l'albicidine peuvent aider à découvrir de nouveaux produits naturels chez des bactéries des genres *Xanthomonas* et *Streptomyces*?

Nous nous sommes demandé si les caractéristiques structurales uniques de l'albicidine, à savoir la présence du *p*ABA et de la cyanoalanine étaient réellement uniques. En effet, quelques jours après que nous ayions publié la structure de l'albicidine est parue une publication décrivant la famille des cystobactamides, une famille d'inhibiteurs de l'ADN gyrase produits par une bactérie du sol, et qui ont une structure quasi-identique à celle de l'albicidine. Les cystobactamides sont également composés de plusieurs molécules de *p*ABA et d'une molécule de cyanoalanine. Bien qu'elle soit dépourvue des modules PKS, leur chaîne d'assemblage NRPS est quasi-similaire à celle de l'albicidine, avec notamment un domaine A spécifique de l'asparagine, possédant le domaine supplémentaire αANH-like et qui *trans*-complémente un domaine A non fonctionnel dans la chaîne d'assemblage principale.

Toujours dans le cadre de sa thèse, Souhir Sabri à réalisé un "genome mining" sur l'ensemble des génomes bactériens disponibles sur GenBank avec comme requête le domaine A spécifique du *p*ABA ou le domaine A contenant l'insertion αANH-like. Le domaine A spécifique du *p*ABA a été retrouvé dans trois génomes bactériens, dont celui de la bactérie pathogène du manguier *X. citri* pv. *mangiferaeindicae* souche LMG941. Quant au domaine A contenant l'insertion αANH-like, il a été retrouvé dans une vingtaine de génomes de bactéries d'origines très variables. Toutefois, pour toutes ces souches, on retrouve ce domaine A dans quatre types de gènes de biosynthèse qui diffèrent au niveau des autres domaines NRPS présent sur ces gènes.

[Note: Ces résultats sont inclus dans une publication qui est écrite depuis 2016, avant la soutenance de Souhir Sabri. Toutefois, nous avons décidé de retarder la soumission de cette publication à BMC Genomics car elle est centrée pour grande partie sur le domaine αANH-like que nous pensons être responsable de la conversion de l'asparagine en cyanoalanine lors de la biosynthèse de l'albicidine. Or, nous souhaitons ne pas trop attirer l'attention sur ce résultat car nos collègues berlinois réalisent actuellement des études fonctionnelles visant à démontrer expérimentalement la fonction de ce domaine et son mécanisme d'action. Leurs résultats renforceront beaucoup la publication de Souhir Sabri. Nous envisageons la soumission d'ici à juin 2019].

Ces résultats indiquent que les spécificités structurales de l'albicidine ne sont certes pas uniques, mais qu'elles sont tout de même rares. Dans le cadre de deux autres projets, nous avons envisagé de cribler deux collections de bactéries du genre *Xanthomonas*, d'une part, et du genre *Streptomyces,* d'autre part, avec comme premières requêtes, le domaine A spécifique du *p*ABA et le domaine αANH-like.

III. 1 "Genome mining" d'une collection de bactéries du genre *Xanthomonas* (projet XanthoNRPS)

Nos résultats précédents nous ont incités à proposer de cribler une collection d'environ 200 souches représentatives de la diversité des *Xanthomonas*, un genre bactérien regroupant 28 espèces phytopathogènes. Pour rappel, six groupements de gènes de biosynthèse ont été trouvés chez *X. albilineans*, dont l'albicidine et Meta-B, ce dernier étant également présent chez deux autres espèces (*X. oryzae* et *X. translucens*). Enfin, un groupement de gènes de biosynthèse gouvernant l'incorporation de *p*ABA a été identifié chez le *Xanthomonas* pathogène du manguier. J'ai proposé un projet en collaboration avec les UMR IRHS à Angers (Marie-Agnès Jacques, collection française des bactéries phytopathogènes) et LIPM à Toulouse (Jérôme Gouzy, plateforme bioinformatique), ainsi qu'avec notre partenaire Roderich Süssmuth à Berlin visant à séquencer les 200 souches de *Xanthomonas* et à miner ces génomes, dans un premier temps pour la présence de domaines NRPS spécifiques du *p*ABA ou possédant le domaine αANH-like.

Pour les hits positifs, nous réaliserons une analyse *in silico* qui permettra de prédire des caractéristiques structurales. Les loci les plus prometteurs seront sélectionnés et les métabolites secondaires seront isolés en utilisant dans un premier temps le protocole utilisé avec succès pour l'albicidine et Meta-B.

Ce projet a été soumis à l'ANR en 2016, mais n'a pas été financé. Nous envisageons une prochaine re-soumission.

III. 2 "Genome mining d'une collection de bactéries du genre *Streptomyces* (projet CONTACT)

Les souches de *Streptomyces* représentent sans doute le plus grand réservoir de métabolites secondaires à ce jour. J'ai co-développé un projet avec Sabine Galindo (UMR Qualisud à Montpellier, expertise en *Streptomyces*) et Jean-Benoit Morel (UMR BGPI, INRA à Montpellier, spécialisé dans l'étude des interactions plantes-bactéries) afin de cribler une collection de 60 souches du genre *Streptomyces* prélevées dans des sols près de Montpellier. Le projet a pour but d'évaluer le potentiel de ces souches pour une application en biocontrôle. Mais nous proposons également de cribler par "genome mining" cette collection non encore décrite pour la présence des domaines A spécifiques du *p*ABA et αANH-like.

Ce projet est financé par la fondation Agropolis et vient tout juste de démarrer. Dans ce cadre, notre équipe accueille deux étudiants en master que j'encadre. Dans le cadre de son stage, Baptiste Durand a, dans un premier temps, criblé les génomes de *Streptomyces* disponibles sur GenBank avec le domaine αANH-like pour requête. Dix retours positifs ont été obtenus, ils correspondent tous à un même gène dans lequel le domaine αANH-like est toujours associé à un domaine d'épimérisation. De plus, ce gène se situe toujours en amont d'un gène codant une asparagine synthase. Ce gène est retrouvé dans quatre types de groupements de gènes différents, certains des dix génomes de GenBank possédant le même groupement. Ceci pourrait indiquer que ce gène est impliqué dans la biosynthèse de quatre molécules différentes. L'alignement des dix séquences de ce gène a permis à Baptiste Durand de dessiner des amorces PCR pour cribler les ADN extraits de la collection de *Streptomyces* de l'UMR QualiSud. Un résultat positif vient d'être obtenu.

Au total le génome de cinq souches de cette collection de *Streptomyces* seront séquencés prochainement avec les techniques PacBio et Illumina. La souche identifiée par PCR comme possédant le gène codant le domaine αANH-like sera séquencée. Ce séquençage permettra d'annoter le groupement de gènes correspondant et de proposer des stratégies pour étudier le métabolite codé par ce groupement. Les quatre autres souches seront sélectionnées sur la base des résultats de biocontrôle. Les génomes des cinq souches séquencées seront analysés dans le but d'identifier d'autres groupements de gènes prometteurs, c'est-à-dire des groupements de gènes responsables de la biosynthèse de molécules encore inconnues.

CONCLUSION

Les micro-organismes synthétisent d'inombrables produits naturels avec des activités biologiques diverses et présentant une infinité de structures. Pour une grande part, cette diversité chimique est due aux mécanismes de biosynthèse non-ribosomique permettant la synthèse de métabolites secondaires par les bactéries et les champignons. Au sein de groupements de gènes de biosynthèse, les enzymes NRPSs et PKSs fonctionnent comme des chaines d'assemblage qui incorporent séquentiellement des briques pour former des peptides non ribosomaux ou des polycétides.

Suite au déclin de la découverte de nouveaux métabolites secondaires par l'approche classique de criblage d'activité biologiques, le séquençage de très nombreux génomes et le développement d'outils bioinformatiques dédiés à l'analyse de ces groupements de gènes de biosynthèse ont permis de développer de nouvelles approches pour découvrir de nouveaux métabolites secondaires. Au cours de mes travaux, j'ai utilisé deux approches, selon que le but est de déterminer la structure de composés déjà connus ou bien de déterminer la structure de nouveaux composés découverts par "genome mining".

Le travail sur l'albicidine a montré l'importance des analyses *in silico* dans la determination de la structure, au même titre que les minutieuses analyses de dizaines de spectres RMN réalisées par mon collègue Roderich Süssmuth.

Le projet sur la molécule signal VFM de QS est un autre exemple pour lequel l'analyse *in silico* de gènes NRPS a permis de réaliser des prédictions pour les voies de biosynthèse très originales de cette molécule (données confidentielles).

Bien sûr, les analyses *in silico* ne font pas tout, et elles doivent être complétées par des analyses expérimentales qui permettent parfois de corriger l'une ou l'autre erreur de prédiction.

Le "genome mining" permet de répondre au besoin urgent de découvrir de nouvelles molécules actives (de nouveaux antibiotiques, par exemple). En effet, il permet de révéler le vrai potentiel métabolique des génomes, fusse-t-il silencieux. Plusieurs méthodes de réveil de ces groupements de gènes silencieux existent, comme par exemple la sur-expression d'un régulateur transcriptionnel qui a bien fonctionné pour le lipopeptide Meta-B identifié chez plusieurs espèces de *Xanthomonas* par "genome mining". Si l'on a, depuis, quasiment caractérisé la structure de la molécule, sa fonction biologique reste toujours un mystère, mais sa présence dans le génome de trois espèces pathogènes de monocotylédones laisse penser à une possible fonction dans les interactions plantes-pathogènes. Des expérimentations pour tester cette hypothèse sont prévues dans l'été.

Dans les projets "XanthoNRPS" et "CONTACT", nous prévoyons de miner une collection de souches de *Xanthomonas* et une collection de souches de *Streptomyces*. Cette dernière espèce, notamment, constitue un énorme réservoir de métabolites secondaires. Afin de filtrer la quantité importante de groupements de gènes de biosynthèse découverts lors du "génome mining", nous prévoyons, dans un premier temps, de ne miner que pour la présence des domaines très particuliers identifiés chez l'albicidine (domaine spécifique du *p*ABA et domaine αANH-like).

Pour conclure, je souhaiterais remercier encore une fois mes collègues, collaborateurs et étudiants pour leurs contributions multiples à ce travail. J'ai devant moi des projets ambitieux et difficiles mais c'est toujours avec le même plaisir que je travaillerai chaque jour à décrypter les secrets de la Nature…

ANNEX: FIVE REPRESENTATIVE PUBLICATIONS

Cociancich S, Dupont A, Hégy G, Lanot R, Holder F, Hetru C, Hoffmann JA & Bulet P (1994) **NOVEL INDUCIBLE ANTIBACTERIAL PEPTIDES FROM A HEMIPTERAN INSECT, THE SAP-SUCKING BUG *PYRRHOCORIS APTERUS*.** *Biochemical Journal* **300**, 567-575.

Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin GP, Brin C, Carrere S, Gomez C, Hügelland M, Völler GH, Noëll J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Süssmuth RD & Cociancich S (2013) **GENOME MINING REVEALS THE GENUS *XANTHOMONAS* TO BE A PROMISING RESERVOIR FOR NEW BIOACTIVE NON-RIBOSOMALLY SYNTHESIZED PEPTIDES.** *BMC Genomics* **14**, 658. Doi: 10.1186/1471-2164-14-658

Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll J, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M & Süssmuth RD (2015) **THE GYRASE INHIBITOR ALBICIDIN CONSISTS OF *P*-AMINOBENZOIC ACIDS AND CYANOALANINE.** *Nature Chemical Biology* **11**, 195-197. Doi: 10.1038/nchembio.1734

Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD & Cociancich S (2015) **WHAT MAKES *XANTHOMONAS ALBILINEANS* UNIQUE AMONGST XANTHOMONADS?** *Frontiers in Plant Science* **6**, 289. Doi: 10.3389/fpls.2015.00289

Petras D, Kerwat D, Pesic A, Hempel B-F, von Eckardstein L, Semsary S, Arasté J, Marguerettaz M, Royer M, Cociancich S & Süssmuth RD (2016) **THE *O*-CARBAMOYL-TRANSFERASE ALB15 IS RESPONSIBLE FOR THE MODIFICATION OF ALBICIDIN.** *ACS Chemical Biology* **11**, 1198-1204.

Novel inducible antibacterial peptides from a hemipteran insect, the sap-sucking bug *Pyrrhocoris apterus*

Stéphane COCIANCICH,* Annick DUPONT,† Gilbert HEGY,† René LANOT,* François HOLDER,* Charles HETRU,* Jules A. HOFFMANN* and Philippe BULET*‡

*Unité Propre de Recherche du CNRS 'Réponse immunitaire et développement chez les Insectes', Institut de Biologie Moléculaire et Cellulaire, 15 rue Descartes, 67084 Strasbourg Cédex, and †Laboratoire de Spectrométrie de Masse Bio-Organique de l'Université Louis Pasteur, Unité de Recherche Associée 31 au CNRS 'Chimie Organique des Substances Naturelles', 1 rue Blaise Pascal, 67008 Strasbourg Cédex, France

Insects belonging to the recent orders of the endopterygote clade (Lepidoptera, Diptera, Hymenoptera and Coleoptera) respond to bacterial challenge by the rapid and transient synthesis of a battery of potent antibacterial peptides which are secreted into their haemolymph. Here we present the first report on inducible antibacterial molecules in the sap-sucking bug *Pyrrhocoris apterus*, a representative species of the Hemiptera, which predated the Endopterygotes by at least 50 million years in evolution. We have isolated and characterized from immune blood of this species three novel peptides or polypeptides: (*i*) a 43-residue cysteine-rich anti-(Gram-positive bacteria) peptide which is a new member of the family of insect defensins; (*ii*) a 20-residue proline-rich peptide carrying an *O*-glycosylated substitution (*N*-acetylgalactosamine), active against Gram-negative bacteria; (*iii*) a 133-residue glycine-rich polypeptide also active against Gram-negative bacteria. The proline-rich peptide shows high sequence similarities with drosocin, an *O*-glycosylated antibacterial peptide from *Drosophila*, and also with the N-terminal domain of diptericin, an inducible 9 kDa antibacterial peptide from members of the order Diptera, whereas the glycine-rich peptide has similarities with the glycine-rich domain of diptericin. We discuss the evolutionary aspects of these findings.

INTRODUCTION

A series of pioneering studies had established by 1930 that insects can be protected against the injection of normally lethal doses of bacteria by the inoculation of a low dose of bacteria (reviewed in [1]). The induction of this protection, generally referred to as immunization, is a complex and as yet incompletely understood process involving both cellular and humoral components. In higher insects, i.e. essentially in the endopterygote clade, the induced protection is probably to a large extent explained by the rapid synthesis of a battery of potent antibacterial peptides (reviewed in [2] and [3]). This huge clade of insects, which contains more species than the rest of the living world together, is characterized by a far-reaching metamorphosis. The overwhelming species richness in the Endopterygota results from the diversification found in four of its constituent orders: the Coleoptera, the Hymenoptera, the Diptera and the Lepidoptera. Representative species belonging to these four orders have been investigated for the presence of antibacterial peptides, and several families of active inducible molecules have been fully or partially characterized to date (reviewed in [3]). For two of these families, our information is relatively detailed. These are: (*i*) the cecropins (reviewed in [2,4]), which form two α-helices and are active against both Gram-negative and Gram-positive bacteria; (*ii*) the insect defensins, which have six cysteine residues engaged in three intramolecular disulphide bridges ([5,6]; reviewed in [7]). These peptides, which are primarily active against Gram-positive cells, consist of three distinct domains: an N-terminal flexible loop, a central amphipathic α-helix and a C-terminal β-sheet; the α-helix is linked to the β-sheet via two disulphide bridges [8,9]. The other inducible antibacterial peptides of insects have only been characterized at the level of their amino acid sequences (and/or the nucleotide sequences of the corresponding genes). For convenience, and pending more detailed studies, these peptides can be grouped as follows: (*iii*) glycine-rich peptides, namely attacins (Lepidoptera) [10] and the related sarcotoxins II (Diptera) [11], coleoptericin (Coleoptera) [12], diptericin (Diptera) [13] and hymenoptaecin (Hymenoptera) [14]; (*iv*) proline-rich peptides, namely apidaecins (Hymenoptera) [15], abaecin (Hymenoptera) [16] and drosocin (Diptera) [17]. The latter molecules are predominantly active against Gram-negative cells. A remarkable feature of drosocin and diptericin is the presence of *O*-glycosylated substitutions which are necessary for the biological activity of these peptides ([17]; P. Bulet, unpublished work).

The methodologies used to isolate inducible antibacterial peptides from higher insects have failed so far to demonstrate the presence of similar molecules in exopterygote insects, such as members of the Orthoptera and Dictyoptera (lower Neoptera), which have appeared 100 million years before the Endopterygotes [18]. Although such insects reportedly build up a protection against lethal doses of bacteria when initially challenged by low doses, they presumably rely on a different protective mechanism, possibly on enhancement of phagocytosis.

We are interested in the phylogenetic aspects of the immune response in insects and have addressed in this study the Hemiptera, a major insect group which, although more recent in evolution than the Orthoptera and Dictyoptera, belongs to the exopterygote clade and has not adopted metamorphosis. Earlier studies with two hemipteran species, the sap-sucking bug *Oncopeltus fasciatus* [19] and the blood-sucking bug *Rhodnius prolixus* [20], had shown that a bacterial challenge induces the appearance of antibacterial activity in the haemolymph of these insects.

We now report the isolation from the bug *Pyrrhocoris apterus*

Abbreviations used: m.i.c., minimum inhibitory concentration; c.f.u., colony-forming unit; TFA, trifluoroacetic acid; h.p.g.p.c., high-performance gel-permeation chromatography.

‡ To whom correspondence should be addressed.

of three inducible antibacterial peptides. One of these peptides is a 43-residue new member of the widespread family of insect defensins which are active against Gram-positive bacteria (reviewed in [7]). A second peptide is a proline-rich molecule of 20 amino acids which carries an *O*-glycosylated substitution and is related to drosocin, a peptide which has recently been isolated from the fruitfly, *Drosophila* [17]. This new glycopeptide, which we propose to name 'pyrrhocoricin' (from *Pyrrhocoris apterus*), is active against Gram-negative bacteria. Finally, we have isolated and sequenced a 133-residue polypeptide which is active against Gram-negative bacteria and has some sequence similarity to diptericin, an antibacterial peptide of dipteran insects [13]. It is rich in glycine residues (15%) and we propose the name of hemiptericin (from Hemiptera) for this novel polypeptide.

MATERIALS AND METHODS

Insects, immunization and haemolymph collection

Adults (1000 individuals) of *Pyrrhocoris apterus* (Hemiptera) were collected in the field and received a 2 μl injection containing 2500 cells of *Micrococcus luteus* (Gram-positive strain) and 2500 cells of *Escherichia coli* 1106 (Gram-negative strain). After various time intervals, the haemolymph (about 2 μl per animal) was recovered by sectioning an antenna and gently squeezing the body. The haemolymph was pooled in a precooled polypropylene tube in the presence of the proteinase inhibitor aprotinin (Sigma A-6279; final concn. 10 μg/ml of haemolymph), and of phenylthiourea (final concn. 1 μg/ml of haemolymph) to prevent melanization. The haemolymph was centrifuged at 13000 *g* for 1 h at 4 °C and directly used for the purification of antibacterial peptides.

Bacterial strains and medium

The bacterial strains were gifts from the following colleagues: *E. coli* D31 (streptomycin-resistant), *Serratia marcescens* Db11 and *Enterobacter cloacae* β12 from H. G. Boman (Department of Microbiology, University of Stockholm, Stockholm, Sweden); *E. coli* D22 (an Env A1 mutant with a defection in the outer membrane) from P. L. Boquet (Centre d'Etudes Nucléaires, Saclay, France); *Bacillus megaterium* and *B. subtilis* QB935 from J. Millet and A. Klier (Pasteur Institute, Paris, France); *Pseudomonas aeruginosa* A.T.C.C. 82118, *Alcaligenes faecalis*, *Salmonella typhimurium*, *Staphylococcus aureus*, *Staph. saprophyticus*, *Aerococcus viridans* (= *Gaffkya homarii*), *Listeria monocytogenes*, *Klebsiella pneumoniae* and *Pediococcus acidilactici* from H. Monteil (Institute of Bacteriology, University of Strasbourg, Strasbourg, France); *Micrococcus luteus* A270 and *B. thuringiensis* were from the Pasteur Institute Collection, Paris, France; *E. coli* 1106 was from T. Achstetter (Transgène, Strasbourg, France); *Erwinia carotovora carotovora* (CFBP n° 2141) and *Xanthomonas campestris* pv. *orizae* (CFBP n° 2532) were from INRA (Angers, France).

All strains were grown on Luria–Bertani's rich nutrient medium [Bactotrypton (1%)/yeast extract (0.5%)/NaCl (1%; w/v)].

Chemicals

Acetonitrile, h.p.l.c. grade, was obtained from Farmitalia Carlo Erba (Milano, Italy). Trifluoroacetic acid (TFA), sequencer grade, was obtained from Pierce (Rockford, IL, U.S.A.). Deionized water was produced via a tandem of MilliRO and MilliQ systems (Millipore).

Antibacterial assays and determination of the minimal inhibitory concentration (m.i.c.)

Antibacterial activity was monitored during the different purification steps by a plate-growth-inhibition assay as described in [6].

The method used for the determination of the m.i.c. has been described in [14]. Briefly, two-fold serial dilutions of *Pyrrhocoris* defensin and pyrrhocoricin were prepared in deionized water and 10 μl aliquots were placed in microtitre plates. Control antibiotic peptides MSI-94 (a broad-spectrum linear amphipathic magainin) and PGLa (a naturally occurring antibiotic peptide from frog) were generously given by M. A. Zasloff (Magainin Pharmaceuticals, Plymouth Meeting, Philadelphia, PA, U.S.A.) The mixture was completed by addition of 100 μl of a bacterial suspension ($A_{600} = 0.001$) in Luria-Bertani's rich nutrient medium. Final concentrations ranged from 0.01 to 5 μM for *Pyrrhocoris* defensin, and from 1 to 10 μM for pyrrhocoricin. The microtitre plates were shaken at 25 °C for 24 h. The m.i.c. values of the antibacterial peptides were expressed according to Casteels et al. [14], as an interval, $a-b$, a being the highest concentration tested at which bacteria are growing and b the lowest concentration that inhibits the growth of the cells.

Bactericidal assay

The methodology used has already been described in [21]: purified anti-(Gram-positive bacteria) peptide (10 μl) was incubated in microtitre plates in the presence of 90 μl of an exponential-phase culture of *M. luteus* (starting $A_{600} = 0.15$) in Luria–Bertani's rich nutrient medium. Aliquots were removed at different time intervals and plated on nutrient agar; the number of colony-forming units (c.f.u.) was determined after an overnight incubation at 37 °C.

The bactericidal activity of the purified small-sized anti-(Gram-negative bacteria) peptide was determined according to the methodology used for drosocin [17]. Briefly, the purified peptide (10 μl) was incubated in microtitre plates in the presence of 90 μl of an exponential-phase culture of *E. coli* D22 (starting $A_{600} = 0.002$) in phosphate-buffered saline (130 mM). The *E. coli* D22 strain, which has a permeable outer membrane, makes it more sensitive to the antibacterial peptides and therefore suitable when small amounts of antibacterial peptides are available. The activity was detected as described above.

Purification of the antibacterial peptides

Step 1: Sep-Pak pre-purification

The cell-free haemolymph was acidified (0.05% TFA final concn.) and loaded on to a Sep-Pak C_{18} cartridge (Waters Associates). After washing with 5 ml of acidified water (0.05% TFA), stepwise elution was performed with increasing proportions of acetonitrile (20, 50 and 80%) in acidified water. The fractions were concentrated in a vacuum centrifuge (Savant) to remove the organic solvent and TFA and were reconstituted with MilliQ water before monitoring the antibacterial activity by the plate-growth-inhibition assay on *M. luteus* (Gram-positive strain), *E. coli* D22 and *E. coli* D31 (Gram-negative strains).

Step II: size-exclusion chromatography

The first step of purification was a size fractionation by a high-performance gel-permeation chromatography system (h.p.g.p.c.) consisting of serially linked Beckman SEC 3000 and SEC 2000 columns (300 mm × 7.5 mm; Beckman). As the bulk of antibacterial activity was found during the solid-phase extraction on to the Sep-Pak C_{18} cartridge to be present in the 50% elution fraction (see above), only this fraction was applied on the columns and eluted under isocratic conditions with 30% acetonitrile in acidified water at a flow rate of 0.5 ml/min. The column effluent was detected by its u.v. absorption at 225 nm, and the presence of antibacterial activity was monitored as described above.

Step III: final purification

Different conditions were used for the final purification of the active compounds. For the large-sized antibacterial peptide, the active fraction was applied on an Aquapore OD300 C_{18} column (220 mm × 4.6 mm; Brownlee Associates) developed with a linear gradient of 2–52% acetonitrile in acidified water over 90 min at a flow rate of 1 ml/min. The column effluent was monitored by u.v. absorption at 225 nm, and the antibacterial activity was detected as described above. This peptide was further purified by an additional step; the same column was used with a linear gradient of 15–35% acetonitrile in acidified water over 90 min at a flow rate of 2 ml/min. Concerning the small-sized antibacterial peptides, the active fraction was applied to an Aquapore OD300 C_{18} column (220 mm × 4.6 mm, Brownlee Associates). The elution was performed with a linear gradient of 2–52% acetonitrile in acidified water over 90 min at a flow rate of 1 ml/min. The presence of antibacterial activity in the different fractions was detected as described above.

All h.p.l.c. purifications were performed using a Beckman Gold h.p.l.c. system equipped with a Beckman 168 photodiode-array detector.

Enzymic and chemical cleavages

The large-sized antibacterial polypeptide was subjected to digestion with arginyl endopeptidase (Sequencing grade; Takara, Kyoto, Japan), which specifically cleaves the peptide bond on the carboxy side of arginine residues (and also of lysine residues under our conditions), and endoproteinase Glu-C (sequencing grade; Takara), which cleaves the peptide bond on the carboxy side of glutamic acid residues. Digestions were carried out on separate aliquots of 2 nmol of the polypeptide under standard conditions. After a 15 h incubation at 37 °C, the reactions were stopped by adding 20 μl of acidified water. The peptidic fragments were separated on an Aquapore RP300 C_8 column (220 mm × 4.6 mm; Brownlee Associates) developed with a linear gradient of 2–80% acetonitrile in acidified water over 120 min at a flow rate of 0.8 ml/min.

For the chemical cleavage, 5 nmol of the polypeptide were treated with 1000-fold excess of CNBr (Fluka, Buchs, Switzerland) during 20 h at room temperature. The different fragments were separated by chromatography as described above, except for the flow rate, which was 1 ml/min.

Reduction and alkylation

A 1 nmol portion of the purified anti-(Gram-positive bacteria) peptide was dissolved in 40 μl of 0.5 M Tris/HCl/2 mM EDTA, pH 7.5, containing 6 M guanidinium chloride, to which 2 μl of 2.2 M dithiothreitol were added. The sample was flushed with nitrogen and incubated at 45 °C for 1 h. Freshly distilled 4-vinylpyridine (2 μl) was added and incubated for 10 min at 45 °C in the dark. The pyridylethylated peptide was separated by reversed-phase h.p.l.c. prior to microsequencing.

Microsequence analysis

Automated Edman degradation of the peptides and detection of the phenylthiohydantoin derivatives were performed on a pulse-liquid automatic sequenator (Applied Biosystems; model 473 A).

M.s.

The peptides were dissolved in water/methanol (50:50, v/v) containing 1% acetic acid and analysed on a VG BioTech BioQ mass spectrometer.

Identification of the carbohydrate substitution

The analysis of the carbohydrate was done on 1 nmol of the glycopeptide under the conditions described previously [17]. Briefly, the glycopeptide was incubated at 80 °C for 4 h with 1 M HCl in anhydrous methanol and derivatized with *N*-methyl-*N*-trimethylsilyltrifluoroacetamide in anhydrous pyridine and trimethylchlorosilane [22]. The component was injected in a Fisons MD800 g.c.–m.s. spectrometer equipped with a JW DB5 fused-silica capillary column (30 m long; 0.32 mm inner diameter; 0.1 μm film thickness). The on-column injection's mode was preferred to the Ros one in order to improve the sensitivity in the case of very volatile compounds. The sample was injected at 50 °C, and the temperature was programmed from 60 to 200 °C at a rate of 3 °C/min. Helium was used as carrier gas (1.2 ml/min), and 2–50 ng of material (including the silylating reactives) were injected. The mass spectrometer was operated in electron-impact mode. M.s. conditions were as follows: electron energy, 70 eV; trap current, 150 μA; mass range scanned, from m/z 50 to m/z 650; scan speed, 1 s. The identification of the component was performed by comparing its retention time and its fragmentogram with that of standard sugars treated according to the same procedure.

RESULTS

Isolation and identification of antibacterial peptides

Haemolymph from untreated specimens of *Pyrrhocoris apterus* is devoid of antibacterial activity. However, the injection of a low dose of bacteria induces, within 12–24 h, the appearance of a strong activity directed against both Gram-positive and Gram-negative bacteria. This activity is maintained for several days (results not shown).

To isolate and characterize the molecules responsible for this activity, we have challenged 1000 adult insects by injection of a low dose of bacteria and collected their haemolymph on ice after 24 h. A total volume of 2.2 ml of haemolymph was obtained and, after removal of the blood cells by centrifugation, the supernatant was directly applied to a Sep-Pak C_{18} cartridge for solid-phase extraction. The antibacterial activity was recovered by elution with 50% acetonitrile in acidified water. The active fractions were pooled and applied to size-exclusion h.p.l.c. columns developed under isocratic conditions with 30% acetonitrile in acidified water. Absorbance was monitored at 225 nm, and aliquots of the fractions were tested by the plate-growth-

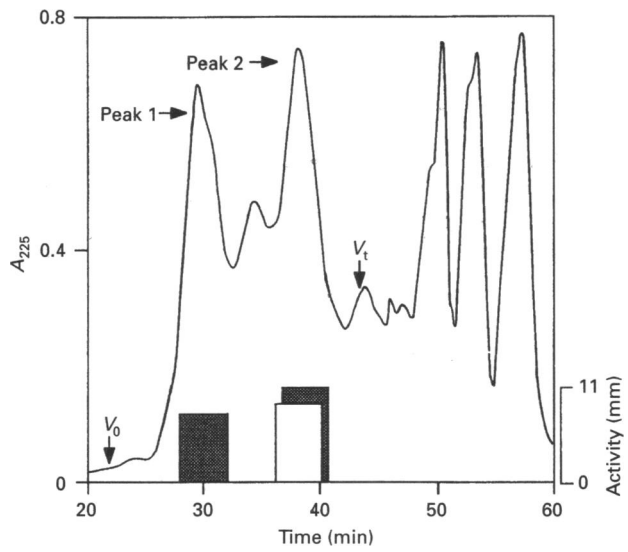


Figure 1 H.p.g.p.c. of immune haemolymph of *P. apterus*

All experimental details were as described in the Materials and methods section. Briefly, active material from the immune haemolymph of 1000 adults of *P. apterus* was pre-purified by solid-phase extraction, and the 50% acetonitrile fraction was analysed on two serial size-exclusion-chromatography columns. Antibacterial activity was detected in two fractions (plate-growth-inhibition assay). Peak 1 was active against Gram-negative bacteria (shaded column), whereas peak 2 exhibited both anti-Gram-positive (white column) and -negative activities (shaded column). The antibacterial activity is expressed in diameter (mm) of the growth-inhibition zones. V_0 is the void volume of the column, and V_t is its total volume.

inhibition assay against *M. luteus* and *E. coli* (Figure 1). Anti-(*E. coli*) activity was detected in absorption peak 1 corresponding to compounds with molecular masses ranging from 10 to 30 kDa, and antibacterial activity directed against both *E. coli* and *M. luteus* was observed in absorption peak 2, which contained the molecules with molecular masses lower than 10 kDa. We did not detect antibacterial activities in the other fractions.

The molecules present in peak 1 were subsequently subjected to C_{18} reversed-phase h.p.l.c., eluted with a 2–52% gradient of acetonitrile in acidified water during 90 min at a flow rate of 1 ml/min and re-chromatographed on the same column with a less steep gradient (15–35% of acetonitrile in acidified water) (over 90 min with an increased flow rate of 2 ml/min). Anti-bacterial activity was monitored on aliquots of the fractions as above. A single anti-(Gram-negative bacteria) compound, appearing pure on the h.p.l.c. chromatogram (results not shown), was recovered and submitted to chemical characterization.

Assuming that the molecule is a polypeptide, we subjected it to Edman degradation (starting material: 240 pmol) and obtained a first sequence of 56 N-terminal residues with an initial yield of 52% and a repetitive yield of 94.5%. We next performed, on separate aliquots of native peptide, enzymic cleavages with arginyl endopeptidase and endoproteinase Glu-C, as well as a chemical treatment with CNBr. The fragments obtained were purified by h.p.l.c. and subjected to Edman degradation. The combined sequences (Figure 2a) lead to the conclusion that the peptide is a 133-residue glycine-rich (15%) polypeptide (see also Figure 4 below). The mass calculated from the amino acid sequence (14744.9 Da) is in excellent agreement with the molecular mass measured by m.s. (14743.8 ± 3.5 Da).

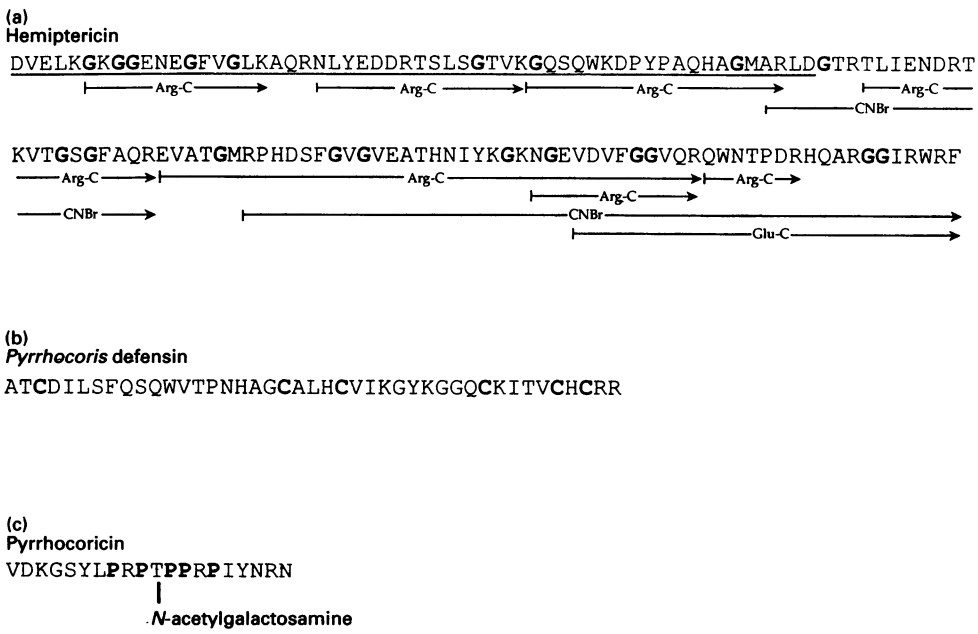


Figure 2 Primary structures of the three antibacterial peptides isolated from the immune haemolymph of *P. apterus*

The residues which are in **bold** allowed us to classify those molecules into their respective families. (**a**) Hemiptericin as a glycine-rich peptide. The 56 N-terminal residues obtained after direct Edman degradation of the native peptide are underlined. The complete amino acid sequence was obtained after enzymic and chemical treatments; arrows indicate the peptidic fragments which were obtained and sequenced after various cleavage treatments of the polypeptide; abbreviations: Arg-C, arginyl endopeptidase; Glu-C, endoproteinase Glu-C. (**b**) *Pyrrhocoris* defensin as an insect defensin. (**c**) Pyrrhocoricin as a proline-rich peptide. Pyrrhocoricin is glycosylated on threonine-11 with an *N*-acetylgalactosamine residue.

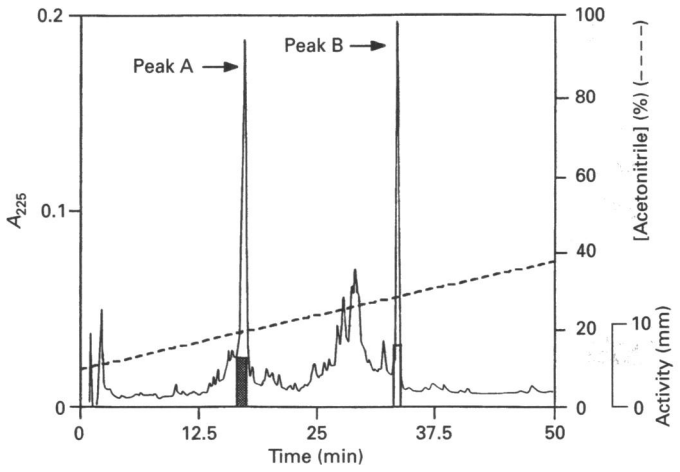


Figure 3 Final purifications of pyrrhocoricin and *Pyrrhocoris* defensin

Peak 2 obtained after the size-exclusion chromatography step was subjected to a reversed-phase h.p.l.c. as described in the Materials and methods section. Two distinct peaks (A and B) contained antibacterial activity. Peak A was active against Gram-negative bacteria (shaded column) and peak B against Gram-positive bacteria (white column). Antibacterial activity is expressed in diameter (mm) of the growth-inhibition zones.

The molecules that were eluted in absorption peak 2 on Figure 1 (corresponding to low-molecular-mass compounds) were subjected to C_{18} reversed-phase h.p.l.c. with a gradient of 2 to 52% of acetonitrile in acidified water during 90 min at a flow rate of 1 ml/min. Two peaks of absorption share an antibacterial activity (Figure 3): peak A was active against *E. coli*, and peak B, which was more hydrophobic, was active against *M. luteus*. The active compounds present in the two peaks appeared sufficiently pure for chemical characterization.

We have first submitted the hydrophobic compound of peak B to Edman degradation and obtained the following sequence of the 15 N-terminal residues: Ala-Thr-Xaa-Asp-Ile-Leu-Ser-Phe-Gln-Ser-Gln-Trp-Val-Thr-Pro. In this sequence, residue 3 could not be assigned. It was immediately apparent that this sequence had a high similarity to that of various members of the family of insect defensins (reviewed in [7]). These peptides have six cysteine residues engaged in three intramolecular disulphide bridges and, assuming that the newly-isolated peptide has similar characteristics, 1 nmol of this molecule was submitted to reduction and alkylation to allow correct identification of the cysteine residues. From 100 pmol of the modified peptide, a sequence of 43 amino acids was subsequently obtained by Edman degradation with an initial yield of 55% and a repetitive yield of 96.1%. This sequence, presented in Figure 2(b) (see also Figure 4), confirms that the peptide is a novel member of the insect defensin family. The calculated mass of the peptide is 4728.5 Da, assuming that the six cysteine residues are engaged in three intramolecular disulphide bridges, as is the case in other defensins (from the fleshflies *Phormia* [23] and *Sarcophaga* [24], and from the dragonfly *Aeschna* [21]). This mass is in excellent agreement with the experimental mass of 4729.3 ± 0.3 Da determined by m.s.

We next analysed the less hydrophobic compound (peak A, Figure 3). Edman degradation on 350 pmol yielded a sequence of 20 amino acids (initial yield of 53%; repetitive yield of 91.5%). Residue 20 gave a signal which strongly suggested that it corresponds to the C-terminal amino acid, as the following cycles gave no signals. From this primary structure, this peptide appears to be particularly rich in proline residues (25%) (Figure 2c; see also Figure 4). The identification of residue 11 was ambiguous,

Pyrrhocoris defensin	**ATCDILSF**QSQWVTP**NHA**G**CALHCVIKG**Y**KGG**Q**C**KIT-**VC**H**CR**R
Phormia defensin	**ATCDLLS**----GTGI**NH**SA**CAAHCLLRG**N**RGG**Y**C**NGKG**VC**V**CR**N
Zophobas defensin	F**TCDVL**G**F**EIAGTKL**N**S**A**ACG**AHCLA**LG**RRGG**Y**C**NSKS**VC**V**CR**

Pyrrhocoricin	VD**K**--GS**YLPRPT**-P**PRPI**YN**R**N
Drosocin	G**K**PRP-**YSPRPT**SH**PRPI**--**R**V
Apidaecin	GNNRPV**YIP**Q----**PRP**PHP**R**I
Proline-rich domain of diptericin	DE**K**PK---**L**IL**PT**--**PAP**PNLPQ...

Hemiptericin	DVELKGKGGENEGFVGLKAQRNLYE**DDR**TS**L**SGTVKGQSQWKD**P**Y
Diptericin	**DEK**PK**L**IL----------**P**T
	PAQHAGMARLDGTRTLIE**N**DRTK**VTGSGF**A**QR**EVATGMRPH**DSFG**
	PAPP--------------**N**LPQL**VGGGGG**-**NR**K--------**DGFG**
	VGVEATHN-**I**YKGK**NG**--EV**D**V**F**G**GV**-**QR**QWNT**P**---D**R**HQA**R**-**G**-**G**IRWR**F**
	VSVDA-**HQKVW**TSD**NG**RHS**I**G**VTPGYSQ**H-LGG**P**YGNS**R**PDY**RIGAG**YSYN**F**

Figure 4 Sequence comparison of the antibacterial peptides from *Pyrrhocoris apterus* with antibacterial peptides from the dipteran *Phormia terranovae* (defensin, diptericin) and *Drosophila melanogaster* (drosocin), the hymenopteran *Apis mellifera* (apidaecin) and the coleopteran *Zophobas atratus* (defensin)

Gaps are introduced so that the amino acid sequences of the peptides belonging to each family are maximally aligned. In each case, identical residues and conservative replacements between the *Pyrrhocoris* antibacterial peptide and at least one of the other are in **bold**. The residues which are underlined carry an *O*-glycosylated substitution.

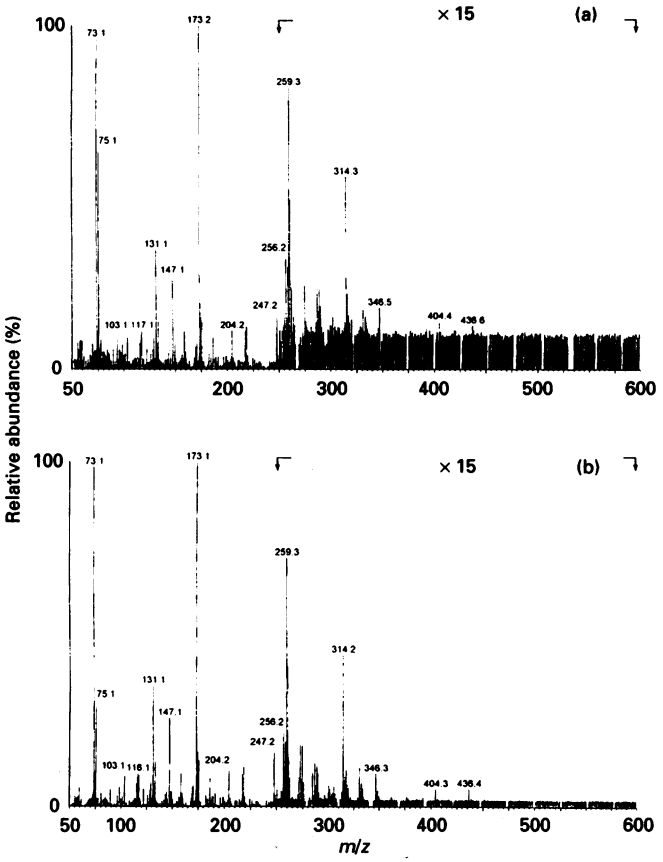


Figure 5 Identification of the carbohydrate substitution of pyrrhocoricin by g.c.–m.s.

Pyrrhocoricin was submitted to methanolysis and pertrimethylsilylation. One major peak appeared in the chromatographic zone corresponding to carbohydrates, whose retention time and fragmentogram were identical with those of the standard *N*-acetylgalactosamine. (**a**) Electron-impact spectrum of the carbohydrate of pyrrhocoricin; (**b**) electron-impact spectrum of standard *N*-acetylgalactosamine.

and none of the signals obtained in Edman degradation corresponded to a conventional phenylthiohydantoin amino acid derivative. Interestingly, these signals were found to be identical with those observed in our recent identification and characterization of drosocin, a 19-residue proline-rich inducible antibacterial peptide isolated from *Drosophila*, which has the striking particularity of carrying an *O*-glycosylated substitution on a threonine residue in position 11 [17]. In drosocin, threonine-11 is located within a pseudo-consensus site for *O*-glycosylation [25–27], which is also the case for the residue 11 in this newly isolated peptide. We have consequently assumed, as a working hypothesis, that residue 11 corresponds to a threonine carrying an *O*-glycosylated substitution. The determination of the molecular mass by m.s. gave a value of 2543.3 ± 0.1 Da, which corresponds to the mass calculated for a peptide in which threonine in position 11 is substituted by an *N*-acetylhexosamine residue (2340.3 Da + 203 Da = 2543.3 Da). To confirm the presence of a sugar residue in the native peptide, we next submitted 1 nmol of purified peptide to methanolysis, which cleaves the carbohydrates from the peptide backbone. The resulting mixture was pertrimethylsilylated and analysed by g.c.–m.s. One carbohydrate was detected and its retention time and fragmentation spectrum were found to be identical with those of *N*-acetylgalactosamine (Figure 5).

From the data obtained above we conclude that the proline-rich 20-residue peptide active against *E. coli* carries an *N*-acetylgalactosamine on a threonine residue in position 11 (Figure 2c; see also Figure 4).

We have recently isolated from a new extract of immune haemolymph of *P. apterus* a minor form of this peptide carrying, in addition, a galactose on the *N*-acetylgalactosamine, as is the case in drosocin [17].

Studies on activity spectra and mode of action

Pyrrhocoris defensin

The purified *Pyrrhocoris* defensin was tested by the liquid-growth-inhibition assay against various bacterial strains (Table 1). The peptide has a marked activity (m.i.c. < 2.5 μM) against *M. luteus*, *B. megaterium*, *A. viridans*, *Staph. aureus*, and *Staph. saprophyticus*, a moderate activity against *P. acidilactici* and *B. subtilis* QB935 and has no activity against *L. monocytogenes* and *B. thuringiensis*. Except for the D22 form of *E. coli* (shortened lipopolysaccharide), none of the Gram-negative strains tested were affected. Interestingly, it appears that this insect defensin is active against the Gram-positive strains tested in the same range as magainins (MSI-94 and PGLa). These results on the activity spectrum of *Pyrrhocoris* defensin are similar to data obtained with defensin A from *P. terranovae* ([6]; see also [21] for the same conditions of bioassay).

To determine the mode of action of *Pyrrhocoris* defensin, this molecule was tested in the liquid-growth-inhibition assay at a concentration of 1 μM against *M. luteus* (Table 2). Less than 1 min contact with the peptide was found to be sufficient to kill almost all bacteria. This result, indicating that the peptide is bactericidal, is comparable with that observed for the antibacterial activity of insect defensin A from *P. terranovae* [6].

Pyrrhocoricin

The antibacterial spectrum of the purified proline-rich *O*-glycosylated peptide from *P. apterus* was also determined using the liquid-growth-inhibition assay (Table 1). Pyrrhocoricin inhibited bacterial growth of the highly sensitive *E. coli* strain D22 at a concentration below 1 μM, which is significantly lower than the m.i.c. values of magainins. The Gram-negative strains *E. coli* 1106, *Ps. aeruginosa* and *Ent. cloacae* β12 were found to be sensitive to the proline-rich peptide of *P. apterus* at an m.i.c. of 5–10 μM. No activity could be detected against the other Gram-negative strains tested. Interestingly, this highly sensitive test allowed us to observe antibacterial activity against two Gram-positive strains: *M. luteus* and *B. megaterium* (m.i.c. 5–10 μM and 1–2.5 μM respectively), in contrast with the less sensitive plate-growth-inhibition assay that we used during the isolation of this peptide (see above).

Finally, as shown in Table 3, purified proline-rich peptide was found to be bactericidal on *E. coli* D22 after a 24 h incubation with the peptide. The kinetics illustrated in Table 3 are similar to those observed with drosocin in *Drosophila* [17]. Note that these kinetics contrast with those of *Pyrrhocoris* defensin, which is bactericidal in less than 1 min of contact with the bacteria (see below).

We were unable to perform similar experiments with the anti-(Gram-negative bacteria) 133-residue polypeptide isolated in this study, as it was necessary to use all the purified material in the determination of the sequence.

Table 1 Activity spectrum of the insect defensin homologue and of the proline-rich *O*-glycosylated peptide isolated from the immune haemolymph of *P. apterus*

The m.i.c.s. of *Pyrrhocoris* defensin and pyrrhocoricin to inhibit the growth of several representative strains are expressed in final concentrations (μM) and compared with those of control antibiotics (MSI-94, PGLa). Various concentrations of the antibacterial molecules were tested using the liquid-growth-inhibition assay described in the Materials and methods section to determine the m.i.c. Abbreviation: n.d., not determined.

Bacterial strain	Gram-positive or Gram-negative	M.i.c. (μM)			
		Pyrrhocoris defensin	MSI-94	PGLa	Pyrrhocoricin
M. luteus	+	0.15–0.3	0.9–1.8	0.3–0.6	5–10
B. megaterium	+	0.15–0.3	0.22–0.45	0.3–0.6	1–2.5
A. viridans	+	0.3–0.6	0.9–1.8	0.6–1.15	n.d.
Staph. aureus	+	0.62–1.25	1.8–3.6	1.15–2.3	> 10
Staph. saprophyticus	+	1.25–2.5	0.9–1.8	0.6–1.15	n.d.
B. subtilis QB935	+	2.5–5	3.6–7.2	4.6–9.2	n.d.
P. acidilactici	+	2.5–5	0.9–1.8	2.3–4.6	n.d.
B. thuringiensis	+	> 5	> 7.2	> 9.2	n.d.
L. monocytogenes	+	> 5	1.8–3.6	1.15–2.3	n.d.
E. coli D22	–	2.5–5	1.8–3.6	1.15–2.3	< 1
E. coli 1106	–	n.d.	1.8–3.6	1.15–2.3	5–10
E. cloacae β12	–	> 5	1.8–3.6	> 9.2	5–10
Ps. aeruginosa A.T.C.C. 82118	–	> 5	1.8–3.6	> 9.2	5–10
S. typhimurium	–	> 5	1.8–3.6	> 9.2	> 10
K. pneumoniae	–	> 5	> 7.2	> 9.2	> 10
S. marcescens Db 11	–	> 5	> 7.2	> 9.2	> 10
A. faecalis	–	> 5	> 7.2	> 9.2	> 10
X. campestris pv *orizae*	–	n.d.	0.9–1.8	> 9.2	> 10
E. carotovora carotovora	–	n.d.	0.9–1.8	1.15–2.3	> 10

Table 2 Bactericidal effect of *Pyrrhocoris* defensin on *M. luteus*

To 90 μl of an exponential-growth-phase culture of *M. luteus* (A_{600} = 0.15) were added 100 pmol of the antibacterial peptide (final concn. 1 μM) or distilled water (control) at zero time. Aliquots were removed at various time intervals and were plated on nutrient agar to determine the number of c.f.u. after an overnight incubation at 37 °C. Results are expressed in 10^6 c.f.u./ml.

Time	C.f.u.	
	1 μM	Control
0 s	39	39
30 s	0	39
0.5 h	0	40
1 h	0	47
3 h	0	65
8 h	0	73
11 h	0	84

Table 3 Bactericidal effect of the proline-rich *O*-glycosylated peptide from *P. apterus* (pyrrhocoricin) on *E. coli* D22

To 90 μl of an exponential-growth-phase culture of *E. coli* D22 (A_{600} = 0.002) was added 1 nmol of the pyrrhocoricin (final concn. 10 μM) or distilled water (control) at zero time. Aliquots were removed at various time intervals and were plated on nutrient agar to determine the number of c.f.u. after an overnight incubation at 37 °C. Results are expressed in 10^6 c.f.u./ml.

Time	C.f.u.	
	10 μM	Control
0 s	1.2	1.4
1 min	1.2	1.4
0.5 h	1.2	1.6
2 h	1.2	1.5
6 h	0.6	4.6
24 h	0	16.2

DISCUSSION

Our results provide the first insight into the structures of molecules responsible for the induced antibacterial activity of a hemipteran insect. Three distinct molecules are induced: a novel member of the insect defensin family, a novel proline-rich *O*-glycosylated peptide structurally similar to the recently characterized drosocin of members of the Diptera [17], and a novel glycine-rich polypeptide which shows some sequence similarities with diptericin isolated from dipterans [13]. Interestingly, no cecropins were found in the haemolymph of stimulated *P. apterus*.

The novel peptides characterized in bacteria-challenged *P. apterus* call for a series of comments.

(i) *Pyrrhocoris* defensin

Insect defensins appear as a widespread family of inducible anti-(Gram-positive bacteria) peptides. They are present in members of the Diptera, Hymenoptera, Coleoptera and Odonata, but are absent from members of the Lepidoptera (reviewed in [7]). The *Pyrrhocoris* defensin is closest to defensin A of *Phormia terranovae*, which has been characterized at the level of its three-dimensional structure. The *Pyrrhocoris* and *Phormia* defensins have the same relative positions of their six cysteine residues (allowing for a four-residue gap), and 60% of the residues are identical, taking into account conservative replacements (Figure 4). Assuming that the *Pyrrhocoris* defensin has the same three-

dimensional structure as that of *Phormia*, we observed that the major difference between the two peptides resides within the putative N-terminal loop which is longer by four residues in *Pyrrhocoris*. Interestingly, the only other insect order in which defensins have an extended putative N-terminal loop is the Coleoptera (Figure 4), which are the most ancient order within the endopterygote clade. Those two insect defensins have in common 51% of their residues (including conservative replacements).

(ii) The proline-rich *O*-glycosylated peptide pyrrhocoricin

This short novel peptide is interesting in many respects. It is, first of all, extremely hydrophilic: in addition to the polar sugar moiety, 50% of the amino acid residues are charged and/or polar, and only three out of 20 residues are hydrophobic. Of the charged residues, four are basic and one acidic, which explains the strongly cationic character of the peptide (pI 10.3). Pyrrhocoricin shows sequence similarities with drosocin (55%) (Figure 4). They have also in common their short size (20 and 19 residues respectively), the relative richness in proline residues (25 and 33% respectively), their cationic character (pI 10.3 and 12.1 respectively) and the presence at the same position (residue threonine-11) of an *O*-glycosylated substitution (*N*-acetylgalactosamine for pyrrhocoricin and a disaccharide, *N*-acetylgalactosamine-galactose, for drosocin). The presence of minor amounts of a more complex pyrrhocoricin (carrying an *N*-acetylgalactosamine-galactose substitution) underscores the possibility that these substitutions are more complex and have partially been trimmed down during our isolation procedure, which relied on acidic conditions. Pyrrhocoricin shows also some sequence similarities with 18-residue cationic proline-rich antibacterial peptides isolated from honey bees, the apidaecins [15] (Figure 4). A major difference, however, pertains to the absence of *O*-glycosylated substitution in apidaecins, which lack threonine or serine residues. Another common characteristic of these proline-rich antibacterial peptides is the organization of most of their proline residues in proline-arginine-proline triplets (two in pyrrhocoricin, three in drosocin and one in apidaecins).

Interestingly, pyrrhocoricin has some sequence similarity with the N-terminal proline-rich domain of diptericin, a 9 kDa inducible antibacterial peptide from *P. terranovae* [13]. This is illustrated in Figure 4. Unpublished data from this laboratory have recently shown that, in the N-terminal sequence of diptericin, the threonine residue in position 10 also carries an *O*-glycosylated substitution (P. Bulet, unpublished work). Like the other proline-rich short cationic peptides characterized so far, pyrrhocoricin is active against Gram-negative bacteria, but also affects viability of two out of a small number of tested Gram-positive bacterial strains (Table 1). Its exact mode of action remains to be established; as is the case for drosocin, pyrrhocoricin is bactericidal, but its action is relatively slow (Table 3).

(iii) The 133-residue glycine-rich polypeptide hemipticin

This novel polypeptide is remarkable for the exceptionally large number of charged residues: one amino acid out of three is charged, either positively (24 residues) or negatively (20 residues). As the balance between positively and negatively charged residues is nearly even, the overall charge at pH 7 is only +3.6 (pI 9.4). In addition, the charges appear evenly distributed all along the polypeptide. Computer-assisted analysis has not allowed the definition of any precise secondary structure (e.g. α-helix, β-sheet). The molecular-mass determination yielded a value in excellent agreement with the mass calculated on the basis of the amino acid sequence, which rules out the existence of any substitution in the polypeptide we have isolated.

Hemipticin shows significant sequence similarity to the glycine-rich central and C-terminal domain of diptericin. The overall similarity is 29%, and better than 42% if conservative replacements are taken into account (Figure 4). Hemipticin shares with the recently isolated 93-residue hymenoptaecin from hymenopterans the relative richness in glycine residues and in charged residues (one out of three); however, it does not show significant sequence similarity to this polypeptide, nor with any of the other glycine-rich inducible polypeptides from members of the Lepidoptera (attacins) or Diptera (sarcotoxins II), although the latter have some sequence similarity with the glycine-rich domain of diptericin [28].

Even if we have only partial information on the activity spectrum and the mode of action of hemipticin, it is clear that the broad-spectrum antibacterial activity in immune haemolymph of the bug *P. apterus* may be the result of: (*i*) individual actions of anti-(Gram-positive bacteria) molecules (insect defensin) and anti-(Gram-negative bacteria) peptide (pyrrhocoricin) and polypeptide (hemipticin); (*ii*) differences in the kinetics of their mode of actions (slow for pyrrhocoricin, and extremely fast for *Pyrrhocoris* defensin); (*iii*) the functional relevance of possible synergistic mechanisms between these different antibacterial peptides in the haemolymph. Such a synergism has been suggested to occur between attacin, cecropin and lysozyme in bacteria-challenged silk-moths [29] and between hymenoptaecin, apidaecins and lysozyme in the honey bee [14].

As stated in the Introduction, our information on inducible antibacterial peptides in insects is limited to the predominant orders of the endopterygote clade {with the exception of one report on an insect defensin in the order Odonata (dragonflies) [21]}. These orders evolved after the greatest mass extinction of all time, the one between the Permian and Triassic periods 250 million years ago, which wiped out 65% of all insect groups [18]. In contrast, the Hemiptera, to which *P. apterus* belongs, were already present in the early Permian. From the phylogenetic point of view, our data show that the facet of the insect immune response which is represented by the secretion into the blood of induced antibacterial peptides, is relatively ancient within the class of insecta. They also show that three distinct classes of molecules predominate (minor classes may have escaped our detection methods): the insect defensins active against Gram-positive bacteria, short proline-rich peptides and large glycine-rich polypeptides, the latter being active essentially on Gram-negative cells. Insect defensins are present in all the orders belonging to the endopterygote clade, with the remarkable exception of the Lepidoptera (reviewed in [7]). The proline-rich peptides, whether glycosylated or not, had so far only been detected in two endopterygote orders, the Hymenoptera [15,16] and the Diptera [17]. They are absent from Lepidoptera. All orders belonging to the Endopterygota contain glycine-rich polypeptides (reviewed in [3]). An interesting result in this context is the observation that the proline-rich *O*-glycosylated 20-residue pyrrhocoricin shows significant sequence similarities with the N-terminal domain of diptericin from Diptera, and that a marked sequence similarity exists between the glycine-rich 133-residue hemipticin and the large glycine-rich domain of the same diptericin. This observation lends some credit to the hypothesis that, in higher Diptera, diptericin could have evolved from the association of gene domains encoding proline-rich and glycine-rich antibacterial peptides [28]. We were unable to detect the presence of cecropin analogues in *P. apterus*. So far, these 4 kDa

anti-(Gram-positive bacteria) and anti-(Gram-negative bacteria) peptides have only been isolated in Lepidoptera (reviewed in [2]) and Diptera [30], and it may be that these peptides have only appeared after the separation of the dipteran–lepidopteran branch from the endopterygotes in the early Triassic.

We thank Rhône Poulenc Rorer for financial support of G.H. through the program BioAvenir. We also thank Mr. W. Hertzog and his pupils (CM_2, Illkirch-Graffenstaden) for their enthusiastic help in collecting several hundred adults of *P. apterus* in the field.

REFERENCES

1 Metalnikow, S. (1927) in Monographie de l'Institut Pasteur (Masson, et al., eds.), l'Institut Pasteur, Paris
2 Boman, H. G., Faye, I., Gudmunsson, G. H., Lee, J. Y. and Lidholm, D. A. (1991) Eur. J. Biochem. **201**, 23–31
3 Cociancich, S., Bulet, P., Hetru, C. and Hoffmann, J. A. (1994) Parasitol. Today **10**, 132–139
4 Steiner, H., Hultmark, D., Engström, A., Bennich, H. and Boman, H. G. (1981) Nature (London) **292**, 246–248
5 Matsuyama, K. and Natori, S. (1988) J. Biol. Chem. **263**, 17112–17116
6 Lambert, J., Keppi, E., Dimarcq, J. L., Wicker, C., Reichhart, J. M., Dunbar, B., Lepage, P., van Dorsselaer, A., Hoffmann, J. A., Fothergill, J. E. and Hoffmann, D. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 262–266
7 Hoffmann, J. A. and Hetru, C. (1992) Immunol. Today **13**, 411–415
8 Bonmatin, J. M., Bonnat, J. L., Gallet, X., Vovelle, F., Ptak, M., Reichhart, J. M., Hoffmann, J. A., Keppi, E., Legrain, M. and Achstetter, T. (1992) J. Biomol. NMR **2**, 235–256
9 Hanzawa, H., Shimada, I., Kuzuhara, T., Komano, H., Kohda, D., Inagaki, F., Natori, S. and Arata, Y. (1990) FEBS Lett. **269**, 413–420
10 Hultmark, D., Engström, A., Andersson, K., Steiner, H. and Boman, H. G. (1983) EMBO J. **2**, 571–576
11 Ando, K. and Natori, S. (1988) Biochemistry **27**, 1715–1721
12 Bulet, P., Cociancich, S., Dimarcq, J. L., Lambert, J., Reichhart, J. M., Hoffmann, D., Hetru, C. and Hoffmann, J. A. (1991) J. Biol. Chem. **266**, 24520–24525
13 Dimarcq, J. L., Keppi, E., Dunbar, B., Lambert, J., Reichhart, J. M., Hoffmann, D., Rankine, S. M., Fothergill, J. E. and Hoffmann, J. A. (1988) Eur. J. Biochem. **171**, 17–22
14 Casteels, P., Ampe, C., Jacobs, F. and Tempst, P. (1993) J. Biol. Chem. **268**, 7044–7054
15 Casteels, P., Ampe, C., Jacobs, F., Vaeck, M. and Tempst, P. (1989) EMBO J. **8**, 2387–2391
16 Casteels, P., Ampe, C., Rivière, L., Damme, J. V., Elicone, C., Fleming, M., Jacobs, F. and Tempst, P. (1990) Eur. J. Biochem. **187**, 381–386
17 Bulet, P., Dimarcq, J. L., Hetru, C., Lagueux, M., Charlet, M., Hégy, G., van Dorsselaer, A. and Hoffmann, J. A. (1993) J. Biol. Chem. **268**, 14893–14897
18 Labandeira, C. C. and Sepkoski, J. J., Jr. (1993) Science **261**, 310–315
19 Gingrich, R. E. (1964) J. Insect Physiol. **10**, 179–194
20 Azambuja, P., Freitas, C. C. and Garcia, E. S. (1986) J. Insect Physiol. **32**, 807–812
21 Bulet, P., Cociancich, S., Reuland, M., Sauber, F., Bischoff, R., Hégy, G., van Dorsselaer, A., Hetru, C. and Hoffmann, J. A. (1992) Eur. J. Biochem. **209**, 977–984
22 DeJongh, D. D., Radford, T., Hribar, J. D., Hanessian, S., Bieber, M., Dawson, G. and Sweely, C. C. (1969) J. Am. Chem. Soc. **91**, 1728–1740
23 Lepage, P., Bitsch, F., Roecklin, D., Keppi, E., Dimarcq, J. L., Reichhart, J. M., Hoffmann, J. A., Roitsch, C. and van Dorsselaer, A. (1991) Eur. J. Biochem. **196**, 735–742
24 Kuzuhara, T., Nakajima, Y., Matsuyama, K. and Natori, S. (1990) J. Biochem. (Tokyo) **107**, 514–518
25 Wilson, I. B. H., Gavel, Y. and von Heijne, G. (1991) Biochem. J. **275**, 529–534
26 Gooley, A. A., Classon, B. J., Marschalek, R. and Williams, K. L. (1991) Biochem. Biophys. Res. Commun. **178**, 1194–1201
27 O'Connell, B. C., Hagen, F. K. and Tabak, L. A. (1992) J. Biol. Chem. **267**, 25010–25018
28 Wicker, C., Reichhart, J. M., Hoffmann, D., Hultmark, D., Samakovlis, C. and Hoffmann, J. A. (1990) J. Biol. Chem. **265**, 22493–22498
29 Engström, P., Carlsson, A., Engström, A., Tao, Z. and Bennich, H. (1984) EMBO J. **3**, 3347–3351
30 Matsumoto, N., Okada, M., Takahashi, H., Ming, Q. X., Nakajima, Y., Nakanishi, Y., Komano, H. and Natori, S. (1986) Biochem. J. **239**, 717–722

Received 12 November 1993/19 January 1994; accepted 25 January 1994

BMC Genomics

RESEARCH ARTICLE Open Access

Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides

Monique Royer[1], Ralf Koebnik[2], Mélanie Marguerettaz[1], Valérie Barbe[3], Guillaume P Robin[2], Chrystelle Brin[4], Sébastien Carrere[5], Camila Gomez[1], Manuela Hügelland[6], Ginka H Völler[6], Julie Noëll[1], Isabelle Pieretti[1], Saskia Rausch[6], Valérie Verdier[2], Stéphane Poussier[7], Philippe Rott[1], Roderich D Süssmuth[6] and Stéphane Cociancich[1*]

Abstract

Background: Various bacteria can use non-ribosomal peptide synthesis (NRPS) to produce peptides or other small molecules. Conserved features within the NRPS machinery allow the type, and sometimes even the structure, of the synthesized polypeptide to be predicted. Thus, bacterial genome mining *via in silico* analyses of NRPS genes offers an attractive opportunity to uncover new bioactive non-ribosomally synthesized peptides. *Xanthomonas* is a large genus of Gram-negative bacteria that cause disease in hundreds of plant species. To date, the only known small molecule synthesized by NRPS in this genus is albicidin produced by *Xanthomonas albilineans*. This study aims to estimate the biosynthetic potential of *Xanthomonas* spp. by *in silico* analyses of NRPS genes with unknown function recently identified in the sequenced genomes of *X. albilineans* and related species of *Xanthomonas*.

Results: We performed *in silico* analyses of NRPS genes present in all published genome sequences of *Xanthomonas* spp., as well as in unpublished draft genome sequences of *Xanthomonas oryzae* pv. *oryzae* strain BAI3 and *Xanthomonas* spp. strain XaS3. These two latter strains, together with *X. albilineans* strain GPE PC73 and *X. oryzae* pv. *oryzae* strains X8-1A and X11-5A, possess novel NRPS gene clusters and share related NRPS-associated genes such as those required for the biosynthesis of non-proteinogenic amino acids or the secretion of peptides. *In silico* prediction of peptide structures according to NRPS architecture suggests eight different peptides, each specific to its producing strain. Interestingly, these eight peptides cannot be assigned to any known gene cluster or related to known compounds from natural product databases. PCR screening of a collection of 94 plant pathogenic bacteria indicates that these novel NRPS gene clusters are specific to the genus *Xanthomonas* and are also present in *Xanthomonas translucens* and *X. oryzae* pv. *oryzicola*. Further genome mining revealed other novel NRPS genes specific to *X. oryzae* pv. *oryzicola* or *Xanthomonas sacchari*.

Conclusions: This study revealed the significant potential of the genus *Xanthomonas* to produce new non-ribosomally synthesized peptides. Interestingly, this biosynthetic potential seems to be specific to strains of *Xanthomonas* associated with monocotyledonous plants, suggesting a putative involvement of non-ribosomally synthesized peptides in plant-bacteria interactions.

* Correspondence: stephane.cociancich@cirad.fr
[1]CIRAD, UMR BGPI, Montpellier Cedex 5, F-34398, France
Full list of author information is available at the end of the article

Background

Various bacteria, most prominently belonging to the orders *Bacillales*, *Pseudomonadales* or *Actinomycetales*, use non-ribosomal peptide synthesis (NRPS) to produce peptides or other small molecules. These molecules exhibit broad structural diversity and display biological activities that range from adaptation to unfavorable environments, communication or competition with other microorganisms in their natural habitat, or even to action as virulence factors (for review [1,2]). Non-ribosomal peptide synthetases (NRPSs) are megaenzymes, usually with a multimodular structure, that catalyze the non-ribosomal assembly of peptides from proteinogenic and non-proteinogenic amino acids [3,4]. A basic module consists of an adenylation domain (A-domain) responsible for amino acid activation; a peptidyl carrier protein domain (PCP-domain)—usually adjacent to the A-domain—for thioesterification of the activated amino acid; a condensation domain (C-domain), which performs transpeptidation between the upstream and downstream peptidyl and amino acyl thioesters to elongate the growing peptide chain. In addition to this basic subset of core domains, each NRPS system also has a chain-terminating thioesterase domain (TE-domain) that is responsible for detachment of the mature polypeptide. Typically, NRPS initiation modules lack a C-domain [3]. The exceptions are initiation modules of NRPS involved in biosynthesis of cyclic lipopeptides [5]. C-domains present in these initiation modules catalyze *N*-acylation of the first amino acid with a fatty acid (with a β-hydroxy-carboxylic acid). Phylogenetic analyses of C-domains in NRPSs showed that C-domains of initiation modules, termed starter C-domains, segregate in a separate phylogenetic clade distant from the C-domains of elongation NRPS modules [6]. In addition to the essential domains mentioned above (A, PCP, C and TE), optional auxiliary domains such as epimerization (E) or heterocyclization (Cy) domains can be found within some NRPS modules. Commonly, epimerization domains are located C-terminally of PCP-domains and perform epimerization of the last amino acid of the adjacent peptidyl-thioester. Interestingly, biochemical data showed that C-domains involved in assembly of arthrofactin not only catalyze transpeptidation between the upstream peptidyl thioester and the downstream amino acyl thioester, but also catalyze epimerization of the last amino acid from the upstream peptidyl thioester from L into D configuration [7]. These C-domains have been termed dual C/E domains [7]. Phylogenetic analyses of C-domains in NRPSs also showed that dual C/E domains segregate in a separate phylogenetic clade [6]. Because of their wide structural complexity and diversity, natural products synthesized by NRPSs constitute a nearly inexhaustible source of new small molecules that might yield lead compounds in drug discovery [8].

The A-domains are the gatekeepers of biosynthesis of the polypeptide due to the specificity of substrate binding pockets of NRPSs for their cognate substrates. Compared to the 22 proteinogenic amino acids used in ribosomal protein synthesis, the utilization of hundreds of different non-proteinogenic amino acids has been described for NRPS. Each substrate binding pocket of NRPSs is specific for its amino acid substrate, and predictive models based on domain arrangement and on the sequence of modules have been deduced. The predictive power of these models has been refined from sequence analyses of NRPS A-domains with known specificity combined with examples of crystal structures of A-domains [9-11], thus identifying amino acid residues crucial for A-domain specificity. These models postulate specificity-conferring signatures for NRPS A-domains consisting of ten critical amino acid residues putatively involved in amino acid or aryl acid substrate specificity. The number of NRPS modules and their domain organization within the enzymes determine the structure of the final peptide product. Using *in silico* analyses of the NRPS genes, these conserved features within the NRPS machinery allow prediction of the type, and sometimes even the structure, of the synthesized polypeptide. It is therefore possible to investigate the biosynthetic potential of a given bacterium by analysis of the architecture of its NRPS gene clusters.

The transfer of a phosphopantetheinyl group to PCP-domains is required for posttranslational activation of NRPSs. Inactive apo-NRPSs are converted to their active holo-forms by transfer of the 4′-phosphopantetheinyl (P-pant) moiety of coenzyme A to the sidechain hydroxyl of a serine residue in the PCP-domains. The P-pant moiety serves to tether covalently the growing polypeptide being assembled by NRPSs. Transfer of the P-pant moiety from coenzyme A to a serine residue is catalyzed by 4′-phosphopantetheinyl transferase (PPTase). PPTases are involved not only in posttranslational activation of NRPS but also in posttranslational activation of fatty acid and polyketide synthases that do not contain PCP-domains but instead have acyl carrier protein (ACP) domains. All PPTases are recognized by a common signature sequence (V/I)G(I/V)D …x40-45… (F/W)(S/C/T)xKE(S/A)xxK, but overall sequence similarities are low [12-15]. Carrier protein specificity (ACP or PCP) has been determined experimentally for some PPTases. For example, in *Escherichia coli*, the EntD PPTase involved in biosynthesis of the siderophore enterobactin was shown experimentally to be active only on PCP-domains and not on ACP-domains [12,14].

The *Xanthomonadaceae* are a family of Gram-negative bacteria belonging to the order *Xanthomonadales* in the gamma subdivision of the *Proteobacteria* [16]. Members of this family are typically characterized as environmental organisms, and occupy diverse ecological niches such

as soil and water, as well as plant tissues. Many *Xanthomonadaceae*, especially species from the genera *Xanthomonas* and *Xylella*, cause plant diseases, and only one, *Stenotrophomonas maltophilia*, is known to be an opportunistic human pathogen. The genus *Xanthomonas* consists of 27 plant-associated species, many of which cause important diseases of crops or ornamental plants. Individual species comprise multiple pathovars, characterized by distinctive host specificity or mode of infection. Collectively, members of this genus cause diseases on at least 120 monocotyledonous and 260 dicotyledonous crop plants.

Xanthomonas albilineans is a systemic, xylem-invading pathogen that causes leaf scald—a lethal disease of sugarcane (interspecific hybrids of *Saccharum* spp.) [17]. Leaf scald symptoms vary from a single, white, narrow, sharply defined stripe to complete wilting and necrosis of infected leaves, leading to plant death. The only pathogenicity factor of *X. albilineans* to be extensively studied to date is albicidin—a secreted non-ribosomally synthesized peptide with phytotoxic and antibiotic properties [18]. Albicidin is a potent DNA gyrase inhibitor with a novel mode of action [19]. Albicidin targets chloroplastic DNA gyrase A, inhibits chloroplast DNA replication and blocks chloroplast differentiation, resulting in the white foliar stripe symptoms [18]. Albicidin also targets bacterial DNA gyrase A and, as a consequence, exhibits a potent antibiotic activity against a wide range of Gram-positive and Gram-negative bacteria [20]. This antibiotic activity may help *X. albilineans* to combat rival microorganisms during sugarcane invasion. The complete albicidin biosynthesis gene cluster, called XALB1, was cloned and sequenced from *X. albilineans* strain Xa23R1 [21]. XALB1 encodes three large NRPS genes and also resistance, regulatory and tailoring genes [21-24]. A PPTase gene, annotated as *xabA* or *albXXI*, was shown to be required for albicidin biosynthesis in *X. albilineans* strains LS155 and Xa23R1, respectively [21,25]. The *E. coli entD* gene restored the biosynthesis of albicidin in a *xabA* knockout mutant of *X. albilineans* strain LS155 [25], demonstrating that *xabA* (or *albXXI*) has the same PPTase activity as *entD*, which is active only on PCP-domains. Preliminary analyses by nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) did not allow the determination of the structure of albicidin [20,26]. However, these studies showed that albicidin contains about 38 carbon atoms and has an estimated molecular mass of 842 Da. *In silico* analyses of XALB1 provided further insights into the structure of this pathotoxin, and suggested that NRPSs involved in biosynthesis of albicidin incorporate as yet unknown non-proteinogenic substrates [21]. Thus, the accumulated data obtained for albicidin and its biosynthesis gene cluster suggest that it is a potent DNA gyrase inhibitor with a novel mode of action, and therefore might constitute a lead structure for novel antibiotics.

The genome of *X. albilineans* strain GPE PC73 was recently sequenced [27]. This genome contains, in addition to three NRPS genes encoding albicidin biosynthesis [21], seven novel NRPS genes that share similarities with NRPS genes present in the genomes of two recently sequenced American *Xanthomonas oryzae* pv. *oryzae* strains [28]. These novel NRPS genes cannot be assigned to any structure or function since there are no orthologous genes in other bacteria. Interestingly, *X. oryzae* possesses an ortholog of the PPTase gene *xabA* (or *albXXI*) which is hereafter referred to as XaPPTase because of its occurence in two species of *Xanthomonas*.

X. oryzae pathovars are the causal agents of two important diseases of rice: bacterial leaf blight caused by *X. oryzae* pv. *oryzae*, and bacterial leaf streak caused by *X. oryzae* pv. *oryzicola*. The XaPPTase gene is found in all six genome sequences of *X. oryzae* published to date (the completed genome sequences of the three Asian *X. oryzae* pv. *oryzae* strains MAFF 311018, KACC10331 and PXO99A [29-31]; the non annotated draft sequences of the genomes of the two American *X. oryzae* pv. *oryzae* strains X11-5A and X8-1A [28]; and the completed sequence of the genome of *X. oryzae* pv. *oryzicola* strain BLS256 [32]). The presence of the XaPPTase gene in all six genomes indicates that all strains of *X. oryzae* likely possess NRPS genes and produce non-ribosomally synthesized peptides.

The objective of this study was to estimate the biosynthetic potential of *X. albilineans* and *X. oryzae* by *in silico* analyses of their uncharacterized NRPS gene clusters and to identify other bacterial candidates for extended genome mining by PCR screening of a collection of 94 plant pathogenic bacteria for the presence of the XaPPTase gene and other NRPS-associated genes. To complete available published genomic data, we included in our analyses unpublished draft genome sequences of the African *X. oryzae* pv. *oryzae* strain BAI3 isolated in Burkina Faso [33] and the *Xanthomonas* spp. strain XaS3 isolated in Guadeloupe (French Caribbean island). To date, no genome sequence of an African *X. oryzae* pv. *oryzae* strain is available in public databases but a draft sequence of the genome of African strain AXO1947 of *X. oryzae* pv. *oryzae* was used recently to identify candidate type III secretion system effector genes [34]. *Xanthomonas* spp. strain XaS3 was isolated from the surface of a sugarcane leaf and was characterized initially as a member of the species *X. albilineans* [35]. However, recent MLSA (multi locus sequence analyses) analyses showed that this strain, which does not possess the albicidin biosynthesis gene cluster, is close to *X. albilineans* but belongs to a separate phylogenetic clade (I. Pieretti, unpublished data).

Results and discussion

Features of the XaPPTase genes in *X. albilineans* and *X. oryzae*

In strains LS155, Xa23R1 and GPE PC73 of *X. albilineans*, as well as in the *Xanthomonas* spp. strain XaS3, the XaPPTase gene is located between the *rpsF* gene encoding the 30S ribosomal protein S6 (XALc_1735) and a gene encoding an iron-sulfur cluster assembly protein (XALc_1737). These latter two genes are conserved and contiguous in other sequenced species of *Xanthomonas*. In all sequenced strains of *X. oryzae*, including strain BAI3, the XaPPTase gene is located in a region containing several tRNA genes and phage-related sequences (Figure 1). In strain BLS256 of *X. oryzae* pv. *oryzicola*, this region also contains two large NRPS genes that are partially conserved and located in the same region in strain 306 of *Xanthomonas axonopodis* pv. *citri* (Figure 1). However, the XaPPTase gene is not conserved in strain 306. The region between the tRNA genes does not contain any genes in several other species of *Xanthomonas* (Figure 1), suggesting that XaPPTase and/or NRPS genes were acquired by lateral gene transfer.

The XaPPTase genes from *X. albilineans* and *X. oryzae*, which are located in two different genomic regions, share only 51% amino acid similarity, although their PPTase signature sequences are very similar (Table 1). The reciprocal best BLAST hit in GenBank for the XaPPTase gene of *X. albilineans* is the XaPPTase gene from *X. oryzae*. This

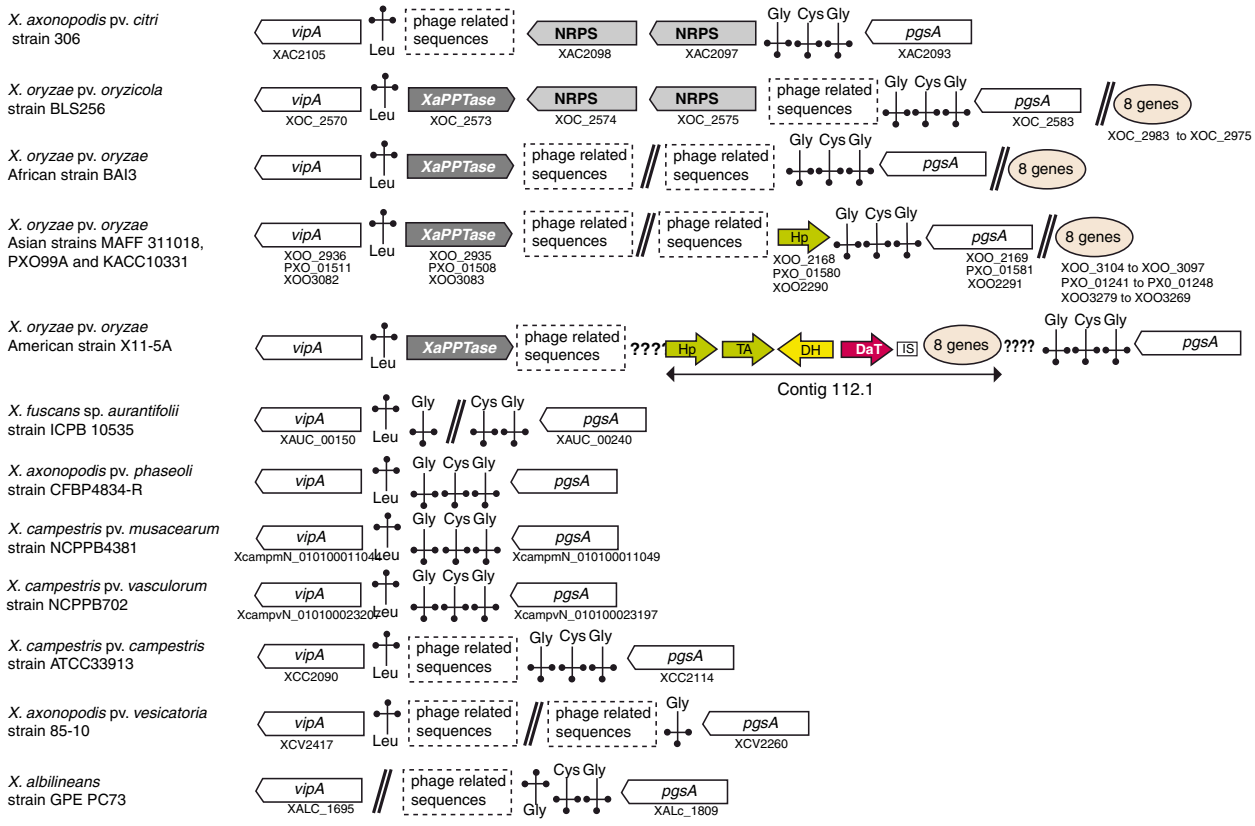


Figure 1 Physical map of the genomic region containing the XaPPTase gene in *Xanthomonas oryzae* strains and of the corresponding region in other sequenced species of *Xanthomonas*. White arrows: gene *vipA* (probable UDP-glucose/GDP-mannose dehydrogenase gene) and gene *pgsA* (probable CDP-diacylglycerol-glycerol-3-phosphate-3-phosphatidyltransferase gene). Dark grey arrows: XaPPTase gene (encoding a 4′-phosphopantetheinyl transferase active on peptidyl carrier protein domains). Light grey arrows: NRPS (nonribosomal peptide synthesis) genes. Coloured arrows correspond to genes described in Figures 3 and 4. Hp: hypothetical protein gene (ortholog of accession Dd703_3065 of strain Ech703 of *Dickeya dadantii*). TA: transaminase gene (ortholog of accession Dd703_3064 of strain Ech703 of *D. dadantii*). DH: lactate dehydrogenase gene. DaT: gene involved in biosynthesis of 2,4-diaminobutyric acid (Dab). Length of arrows is not proportional to the length of genes. Salmon-coloured oval circle: Eight genes present in contig 112.1 of *X. oryzae* strain X11-5A that are conserved in other strains of *X. oryzae* and are not predicted to be involved in NRPS biosynthesis. •⁞•: tRNA. Orientation of the tag indicates the orientation of the tRNA gene in the genomic regions. Amino acid specificity of each tRNA is indicated above or below each tag according to the orientation of the tRNA gene. •⁞•: undetermined sequence located between contigs. Two contigs separated by this tag may be located in two different genomic regions (they are not necessarily contiguous). //: this tag indicates that corresponding genomic regions are not contiguous. Data on strain CFBP4834-R of *X. axonopodis* pv. *phaseoli* were obtained from an unpublished finished genome sequence (M.-A. Jacques, personal communication).

indicates that these genes might have been derived from a common ancestor. Signature sequences of all XaPPTase genes are very similar to the signature sequence of *entD* gene of *E. coli* (Table 1). *X. albilineans* and *X. oryzae* possess another PPTase gene, *hetI*, which is conserved in all sequenced species of *Xanthomonas*. Signature sequences of *hetI* of *X. albilineans* and *X. oryzae* are both similar to the signature sequence of a PPTase of *Mycobacterium tuberculosis* that was shown experimentally to be active only on ACP-domains of fatty acid and polyketide synthases and not on PCP-domains [15] (Table 1).

XaPPTase proteins of strains LS155 and Xa23R1 of *X. albilineans* are 100% identical, but they are only 94% and 83% identical to XaPPTase proteins of strains GPE

Table 1 Similarity between XaPPTase and other PPTases involved in NRPS and fatty acid biosynthesis in bacteria (from [25])

Pathways	Proteins	Organisms [a]	Experimentally-determined specificities (A/P)[b]	Domain I	Spacing between domains I and II (in aa)	Domain II	Overall amino acid identities / similarities with XabA
Albicidin and unknown	XabA	*X. albilineans* str. LS155	P[c]	G**VGID**LERP	–(x)39–	**FS**A**KES**LF**K**AAY	-
	AlbXXI	*X. albilineans* str. Xa23R1	?	G**VGID**LERP	–(x)39–	**FS**A**KES**LF**K**AAY	100% /-
	XaPPTase	*X. albilineans* str. GPE PC73	?	G**VGID**LERP	–(x)39–	**FS**A**KES**LF**K**AAY	94% / 95%
	XaPPTase	*Xanthomonas spp.* str. XaS3	?	G**VGID**LERM	–(x)39–	**FS**A**KES**LF**K**AAY	83% / 87%
Unknown	XaPPTase	All sequenced *X. oryzae* strains*	?	G**IGID**LEHL	–(x)38–	**FS**A**KES**LF**K**ASF	40% / 51%
Unknown	XaPPTase	*X. sacchari* str. NCPPB4393	?	G**IG**L**D**VERV	–(x)38–	**FS**A**KES**FY**K**AAA	41% / 55%
Unknown	BBta_3710	*Bradyrhizobium* spp. BTAi	?	AL**G**L**D**IEDV	–(x)35–	**FS**A**KEA**YY**K**CQY	25% / 36%
Enterobactin	EntD	*E. coli* str. K12 substr. MG1655	P[d]	P**IGID**IEEI	–(x)36–	**FS**A**KES**AF**K**ASE	23% / 31%
Mycobactin	PptT	*M. tuberculosis* str. CSU93	P[e]	S**VGID**AEPH	–(x)35–	**FC**A**KEA**TY**K**AWF	25% / 34%
Gramicidin	Gsp	*Bacillus brevis* str. ATCC 9999	P[d]	P**VGID**IERI	–(x)35–	**WT**I**KES**YI**K**AIG	14% / 21%
Surfactin	Sfp	*Bacillus subtilis* str. RB14	A/P[f]	P**IGID**IEKT	–(x)35–	**WS**M**KES**FI**K**QEG	17% / 25%
Fatty acids	AcpS	*E. coli* str. K12 substr. MG1655	A[d]	GL**G**T**D**IVEI	–(x)40–	**F**AV**KEA**AA**K**AFG	9% / 14%
Fatty acids	AcpT	*E. coli* str. K12 substr. MG1655	?	E**VG**C**D**IEVI	–(x)34–	**WT**R**KEA**IV**K**QRG	13% / 22%
Fatty acids	AcpS	*M. tuberculosis* str. CSU93	A[e]	G**VGID**LVSI	–(x)41–	**W**AA**KEA**VI**K**AWS	11% / 17%
Unknown	HetI	*X. albilineans* str. GPE PC73	?	RL**GVD**IERQ	–(x)37–	**WC**A**KEA**LL**K**AHG	25% / 31%
Unknown	HetI	*Xanthomonas spp.* str. XaS3	?	RL**GVD**IERQ	–(x)37–	**WC**A**KEA**LL**K**AHG	22% / 28%
Unknown	HetI	All sequenced *X. oryzae* strains	?	RL**GVD**LERI	–(x)37 --	**WC**A**KEA**LL**K**AYG	20% / 25%
Consensus				(V/I)G(I/V)D**		(F/W)(S/C/TxKE(S/A)xxK**	

[a] str. = strain; substr. = substrain.
[b] A, specific for acyl carrier protein (ACP) domains; P, specific for peptidyl carrier protein (PCP) domains.
[c] [25].
[d] [12].
[e] [15].
[f] [36].
*: XaPPTase gene from strain PXO99A was mis-annotated (correct start codon is upstream of the one deposited in GenBank).
**: Consensus signature sequences of PPTase domains as defined by [14,15].

PC73 and XaS3, respectively. However, XaPPTase signatures are 100% identical in strains LS155, Xa23R1 and GPE PC73, differing only in one amino acid residue in strain XaS3 (Table 1).

Genome mining reveals three novel NRPS loci in the genome of *X. albilineans* strain GPE PC73

In addition to the three NRPS genes characterized previously as being involved in the biosynthesis of albicidin [21], the chromosome of strain GPE PC73 possesses seven large multimodular NRPS genes clustered in three loci, hereafter referred to as META-A, META-B and META-C, respectively. These non-characterized NRPS genes share no similarity with NRPS genes from the biosynthesis gene cluster of albicidin.

Interestingly, the loci META-A and META-C encode NRPSs only, and no NRPS-associated proteins, *e.g.* tailoring enzymes, in the proximity of these gene clusters. META-A is adjacent to the albicidin NRPS gene cluster XALB1, from which it is separated by the terminus of replication. META-A, which is flanked by IS elements, includes two large NRPS genes: XALc_1551 (22,629 bp) and XALc_1550 (16,965 bp). These two genes encode a NRPS system consisting of 12 modules each containing the characteristic domain arrangement C-A-PCP. A TE-domain is located at the C-terminus of XALc_1550 (Figure 2; Additional file 1). META-C, which is not flanked by IS elements, contains a single NRPS gene, XALc_0772 (23,289 bp). This gene encodes a NRPS system consisting of seven modules with a C-A-PCP domain arrangement, followed by a chain-terminating TE-domain (Figure 2; Additional file 1). Phylogenetic analysis of C-domains revealed that the C-domains of the first modules of XALc_1551 and XALc_0772 cluster in a separate clade, distinct from that of the starter C-domains identified by Rausch et al. [6] in numerous bacteria (Additional file 2). This indicates that these two starter C-domains, which share 81% amino acid similarity, are unusual. The reciprocal best BLAST hit in GenBank for both starter C-domains of META-A and META-C is the N-terminal C-domain of a NRPS gene belonging to the plant-associated bacterium *Bradyrhizobium* spp. strain BTAi [37].

The META-B biosynthesis gene cluster of strain GPE PC73 contains four NRPS genes: XALc_1058 (13,686 bp),

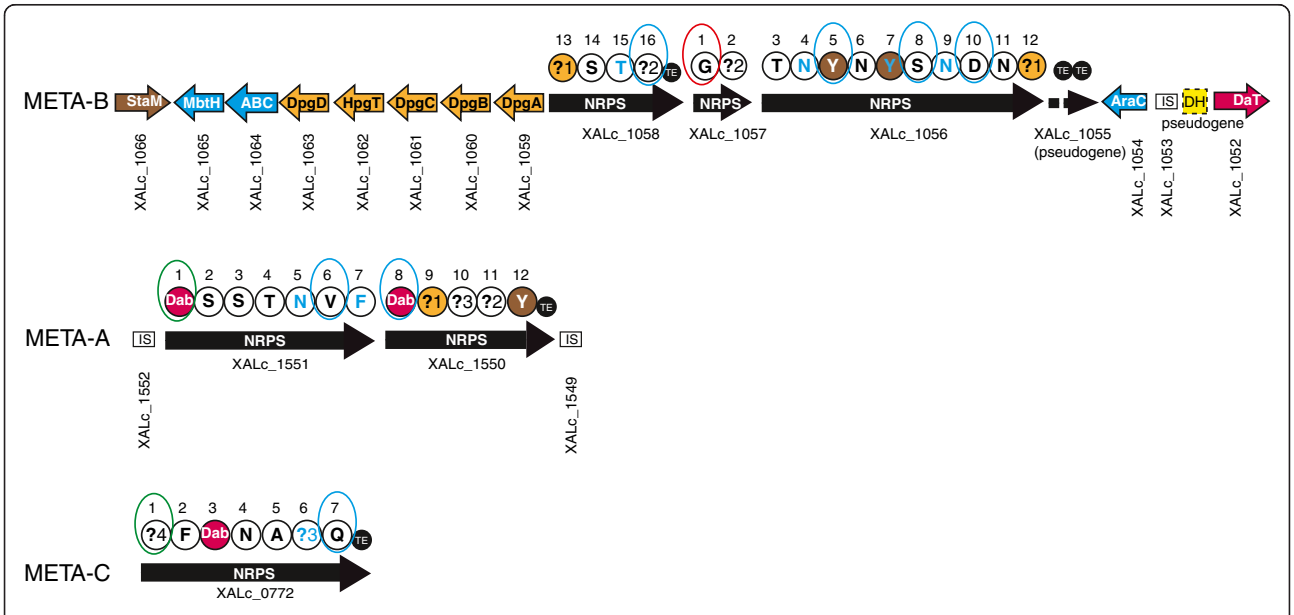


Figure 2 Representation of the NRPS loci META-A, META-B and META-C of *X. albilineans* strain GPE PC73. Brown arrow: *staM* gene predicted to be required for β–hydroxylation of an unidentified amino acid. Purple arrow: *daT* gene predicted to be required for biosynthesis of Dab (2,4-diamino butyric acid). Orange arrows: genes predicted to be required for biosynthesis of Dpg (3,5-dihydroxyphenylglycine). Dark blue arrows: other genes conserved in all META-B loci (ABC transporter, MbtH-like protein and transcriptional regulator AraC). Black arrows: NRPS genes. Length of arrows is not proportional to the length of genes. NRPS modules are represented by circles. Large circles indicate complete NRPS modules (containing domains C, A and PCP). The amino acid predicted to be assembled by the corresponding module is indicated within each circle. ?X: unknown amino acid specific to the NRPS signature UnknownX (Additional file 1). Brown circles: NRPS modules predicted to be specific for tyrosine. Purple circles: NRPS modules predicted to be specific to Dab. Orange circles: NRPS modules predicted to be specific to Dpg. The sequential order of amino acid incorporation from the N- to the C-terminus is illustrated by the number above each circle. Small black circles represent TE domains. The starter module circled in red exhibits a C-domain belonging to the same phylogenetic clade as starter C-domains identified by Rausch et al. [6] (Additional file 2). Starter modules circled in green indicate a C-domain belonging to the distant phylogenetic clade containing starter C-domains of loci META-A and META-C (Additional file 2). Elongation modules circled in blue indicate a C-domain belonging to the same phylogenetic clade as dual C/E domains identified by Rausch et al. [6] (Additional file 2). Amino acids predicted to be epimerized by these domains are in blue. IS: insertion sequence. Dotted box: pseudogene.

XALc_1057 (6,378 bp), XALc_1056 (32,124 bp) and XALc_1055 (1,824 bp) (Figure 2; Additional file 1). These four genes encode a NRPS system consisting of a total of 16 modules with a continuous C-A-PCP domain arrangement. Two TE-domains were identified in META-B: (i) a single TE-domain at the C-terminus of the NRPS XALc_1058, and (ii) one double TE-domain on its own encoded by a separate gene (XALc_1055) (Figure 2). Like the domain arrangement of META-A and META-C, only essential elongation domains (A, PCP, C and TE) were identified in META-B. Phylogenetic analysis showed that the C-domain of the first module of XALc_1057 belongs to the same clade as starter C-domains identified by Rausch et al. [6], indicating that it also corresponds to a starter C-domain (Additional file 2). Therefore, the first module of XALc_1057 was assigned as the initiation module of META-B (Figure 2). The last module of XALc_1058, which harbors a TE domain, was assigned as the termination module (Figure 2). XALc_1055 is the only gene available in GenBank to encode a 'stand-alone' double TE domain, indicating that this gene is not associated to any NRPS system and should be considered as a pseudogene.

In addition to these four NRPS genes, the META-B locus encodes an ABC transporter (XALc_1064), an AraC transcriptional regulator (XALc_1054) and a MbtH-like protein (XALc_1065). The ABC transporter likely is the corresponding transporter for secretion of small molecule(s) synthesized by META-B. The AraC protein may be required for specific transcriptional regulation of genes present in META-B. Finally, the MbtH-like protein shares similarities with proteins involved in biosyntheses of non-ribosomal peptides recently described in other bacteria [38,39]. Some genes in the META-B gene cluster are predicted to be required for the biosyntheses of three non-proteinogenic amino acids: an unidentified β-hydroxy-amino acid, 2,4-diamino butyric acid (Dab), and 3,5-dihydroxyphenyl-glycine (Dpg). XALc_1066 shares 52% amino acid similarity with *staM*, which was assigned in *Streptomyces toyocaensis* as the gene responsible for β-hydroxylation of amino acids, preferably tyrosine [40]. The protein encoded by XALc_1052 is 52% similar to a diaminobutyrate transaminase characterized in *Acinetobacter baumannii* as catalyzing the conversion of aspartate semialdehyde (an intermediate in the biosynthesis of various amino acids) to Dab [41]. Genes predicted to be required for biosynthesis of Dpg share high similarities with genes previously characterized in *Amycolatopsis balhimycina* [42] and *S. toyocaensis* [40], respectively (Table 2). Based on these similarities, XALc_1059 is predicted to encode a polyketide synthase (DpgA) that may generate 3,5-dihydroxyphenylacetyl-CoA from four molecules of malonyl-CoA, with the assistance of XALc_1060 and XALc_1063 (which encode proteins DpgB and DpgD, respectively, and which may exhibit dehydratase activity). XALc_1061 is predicted to encode dioxygenase DpgC, which may convert 3,5-dihydroxyphenylacetyl-CoA into 3,5-dihydroxyphenyl-glyoxylate, and subsequently generate Dpg after transamination by XALc_1062 (HpgT, hydroxyphenyl-glycine transaminase).

The identity of several amino acids recognized and activated by the NRPSs of META-A, META-B or META-C is predicted on the basis of the analysis of specificity-conferring signatures of A-domains (Figure 2; Additional file 1). Interestingly, signatures specific to Dab were identified in both loci META-A and META-C, but remarkably not in the locus of META-B, which encodes the gene required for biosynthesis of Dab (XALc_1052). This finding suggests the occurrence of cross-talk between the three NRPS loci. Four unknown substrate specificity-conferring signatures were

Table 2 Similarities between genes predicted to be required for biosynthesis of Dpg in strain GPE PC73 of *Xanthomonas albilineans* and genes previously characterized from strain DSM 5908 of *Amycolatopsis balhimycina* [42] and from strain NRRL15009 of *Streptomyces toyocaensis* [40]

Genes of strain GPE PC73	Genes of strain DSM 5908 of *A. balhimycina*	Overall amino acid identities / similarities with genes of strain DSM 5908	Genes of strain NRRL15009 of *S. toyocaensis*	Overall amino acid identities / similarities with genes of strain NRRL15009	Functions
XALc_1059	DpgA (CAC48378)	55% / 69%	DpgA (AAM80548)	54% / 67%	Type III chalcone synthase, generates 3,5-dihydroxyphenylacetyl-CoA from four malonyl-CoA
XALc_1060	DpgB (CAC48379)	31% / 46%	DpgB (AAM80547)	29% / 45%	Belongs to crotonase/Enoyl-CoA hydratase superfamily, enhances DpgA activity
XALc_1061	DpgC (CAC48380)	48% / 62%	DpgC (AAM80546)	48% / 62%	Metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase, converts 3,5-dihydroxyphenylacetyl-CoA into 3,5-dihydroxyphenylglyoxylate*
XALc_1062	HpgT (CAC48367)	43% / 58%	HpgT (AAM80549)	42% / 58%	Transaminase, generates 3,5-dihydroxyphenyl-glycine from 3,5-dihydroxyphenylglyoxylate
XALc_1063	DpgD (CAC48381)	62% / 74%	DpgD (AAM80545)	61% / 74%	Belongs to crotonase/Enoyl-CoA hydratase superfamily, enhances DpgA activity

* This function was characterized for biosynthesis of vancomycin [43].

identified in the NRPSs of META-A, META-B or META-C (Unknown1 to Unknown4). Phylogenetic analysis of C-domains also showed that C-domains of two modules of META-A, four modules of META-B and one module of META-C segregate in the same clade as dual C/E domains identified by Rausch et al. [6] in numerous bacteria (Additional file 2). By analogy to previously identified dual C/E domains from arthrofactin NRPS, these latter dual C/E domains are predicted to follow modules that assemble amino acids with D-configuration (Figure 2; Additional file 1).

Genome mining reveals short NRPS genes in the genome of strains GPE PC73 and XaS3

Two short NRPS genes are also present on the chromosome of strain GPE PC73 in two additional loci. These two short NRPS genes, XALc_0364 (4,047 bp) and XALc_1145 (4,011 bp), each encode only one NRPS module, with a C-A-PCP-TE domain arrangement, and overlap at their 3′ end with a glycosyltransferase gene (overlapping sequence is 4 bp and 8 bp in length for XALc_0364 and XALc_1145, respectively). XALc_0364 and XALc_1145 share 66% amino acid similarity but do not share the same substrate-specificity conferring signature (Additional file 3). XALc_0364 harbors a signature specific to Gly, and XALc_1145 harbors an unknown signature that differs from unknown signatures identified in other NRPS loci of strain GPE PC73. Their respective overlapping glycosyltransferase genes (XALc_0365 and XALc_1144) share 62% amino acid similarity (Additional file 3). The C-domains of these two short NRPS genes segregate in the same phylogenetic clade as starter C-domains identified by Rausch et al. [6]. Genes similar to the short NRPS genes of strain GPE PC73, *i.e.* encoding only one NRPS module with a C-A-PCP-TE domain arrangement and overlapping a glycosyltransferase gene at their 3′-end, were found in the genome of the marine hydrocarbonoclastic bacterium *Alcanivorax borkumensis* strain SK2 (Additional file 3). Interestingly, in this strain, a XaPPTase-like gene (ABO_1782) is contiguous with the glycosyltransferase gene (ABO_1783) and the short NRPS gene (ABO_1784), confirming that a PPTase specific to PCP-domains is required for the unknown function of these unusually small NRPS systems. Similar short NRPS genes were found in the genome of two other species of *Xanthomonas*, which nevertheless do not possess the XaPPTase gene (*X. campestris* pv. *campestris* and *X. axonopodis* pv. *citri*; Additional file 3). This indicates that these genes were present in the common ancestor of *Xanthomonas* but may have conserved their function only in species that conserved the XaPPTase gene. Short NRPS genes were not found in any sequenced strains of *X. oryzae*, including strain BAI3. However, orthologs of these genes were found in *Xanthomonas* spp. strain XaS3. The only short NRPS gene present in strain XaS3 shares 94% and 67% amino acid similarity with the two short NRPS genes XALc_0364 and XALc_1145 of strain GPE PC73, respectively, and exhibits the same Gly-specificity conferring signature as in XALc_0364 (Additional file 3). The overlapping glycosyltransferase gene of the short NRPS gene of strain XaS3 shares 92% and 62% amino acid similarity with the two glycosyltransferase genes XALc_0365 and XALc_1144, respectively, of strain GPE PC73 (Additional file 3).

Genome mining uncovers a META-B-like NRPS gene cluster in strains X11-5A, BAI3 and XaS3

Genes similar to those from the gene cluster META-B of strain GPE PC73 were found in the recently published [28] draft genome sequences of the two *X. oryzae* pv. *oryzae* strains X11-5A and X8-1A, and in unpublished draft genome sequences of *Xanthomonas* spp. strain XaS3 and *X. oryzae* pv. *oryzae* strain BAI3. In these draft sequences, fragments of NRPS genes were distributed in several independent contigs. BLAST analyses indicated that the META-B gene cluster was very similar in strains X11-5A and X8-1A. We therefore analyzed only strain X11-5A (Figure 3). For strains X11-5A and XaS3, *in silico* analyses were performed directly on the draft sequence by analyzing each contig independently. For strain BAI3, we performed additional cloning and sequencing experiments to assemble the contigs and determine the complete sequence of an 82-kb region containing META-B.

All NRPS-associated genes found in the META-B gene cluster of strain GPE PC73 are conserved in strains BAI3, X11-5A and XaS3 (Figures 3 and 4; Table 3), except for XALc_1052 (the *daT* gene required for biosynthesis of Dab), which is absent in strain BAI3. Interestingly, the META-B loci of strains X11-5A and XaS3 both contain four additional genes that may be required for the biosynthesis of additional unknown non-proteinogenic amino acids (Table 3). These genes encode a protein sharing 58% amino acid similarity with the protein encoded by the *staM* gene of *S. toyocaensis* (*StaM'*), a transaminase (TA), a lactate dehydrogenase (DH) and a hypothetical protein (HP), respectively. This hypothetical protein has a superfamily domain that is found in a variety of structurally related metalloproteins, including type I extradiol dioxygenases, glyoxalase I and a group of antibiotic resistance proteins. Orthologs of this hypothetical protein (Dd703_3065) and of the transaminase (Dd703_3064) are present in a NRPS gene cluster in the genome of the phytopathogenic strain Ech703 of *Dickeya dadantii*, supporting the assumption that both genes are required for biosynthesis of non-ribosomally synthesized peptides. Orthologs of the hypothetical protein Dd703_3065 are also present in the three Asian *X. oryzae* pv. *oryzae* strains MAFF 311018, PXO99A and KACC10331 in the region containing the XaPPTase gene

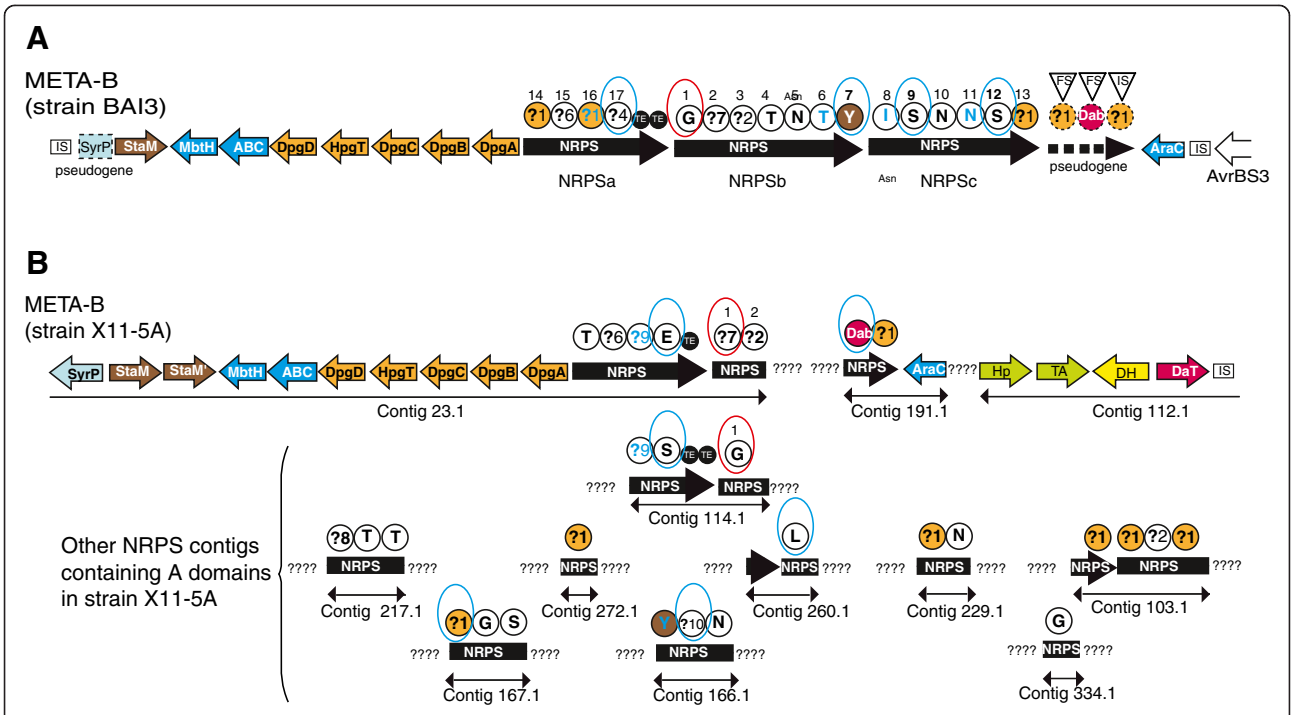


Figure 3 Representation of NRPS sequences of strains BAI3 and X11-5A of *X. oryzae* pv. *oryzae*. A. Representation of the NRPS locus META-B of *X. oryzae* pv. *oryzae* strain BAI3. **B.** Representation of contigs of *X. oryzae* pv. *oryzae* strain X11-5A that contain a nucleotide sequence encoding at least one A domain and/or a nucleotide sequence encoding non-NRPS genes similar to genes found in the locus META-B of strain *X. albilineans* GPE PC73. ???? : undetermined sequence located between contigs. Two contigs separated by this tag may be located in two different genomic regions (they are not necessarily contiguous). Brown arrows: *staM* and *staM'* genes predicted to be required for β-hydroxylation of two unidentified amino acids, respectively. Purple arrow: *daT* gene predicted to be required for biosynthesis of Dab (2,4-diamino butyric acid). Orange arrows: genes predicted to be required for biosynthesis of Dpg (3,5-dihydroxyphenyl-glycine). Dark blue arrows: other genes conserved in all META-B loci (ABC transporter, MbtH-like protein and transcriptional regulator AraC). Green arrows and yellow arrow: genes conserved in strains X11-5A and XaS3 predicted to be required for biosynthesis of an unknown non-proteinogenic amino acid (Hp: hypothetical protein gene, TA: transaminase gene, DH: lactate dehydrogenase gene). Light blue arrow: *syrP* gene predicted to be required for β-hydroxylation of an unknown amino acid. Black arrows: NRPS genes. Length of arrows is not proportional to the length of genes. All information regarding characteristics and color-code for NRPS modules are as detailed in Figure 2. IS: insertion sequence. FS: frameshift mutation. Dotted box or dotted arrow: pseudogene. Dotted circle: non-functional module.

(Figure 1). This suggests that, in the common ancestor of *X. oryzae* strains, META-B may have been already present in the same genomic region as the XaPPTase gene. Contig 112.1 of strain X11-5A, which contains the *daT* gene and the three additional genes mentioned above, also contains eight genes that are not predicted to be required for NRPS on the basis of their sequence. These eight genes are conserved in all other *X. oryzae* strains but are present elsewhere in the genome (Figure 1). This finding supports the conclusion that recombination events had shaped the META-B locus in the ancestor of strain X11-5A. Remnants of the gene encoding the lactate dehydrogenase DH were found in strain GPE PC73 between genes XALc_1052 and XALc_1053, confirming its ancestral origin (Figure 2). In strain X11-5A, the META-B locus contains a *syrP* gene sharing 44% amino acid similarity with the *syrP* gene of *Pseudomonas syringae* pv. *syringae* strain B728a, which was shown to be required for β-hydroxylation of the aspartyl residue in the phytotoxin syringomycin E [44]. No aspartyl residue is predicted to be assembled by META-B in strain X11-5A, suggesting that the *syrP* gene of strain X11-5A may be required for β-hydroxylation of another residue. In strain BAI3, the *syrP* gene should be considered as a pseudogene because it contains a non-sense mutation (Figure 3). The *syrP* gene is also not conserved in the META-B loci of strains GPE PC73 or XaS3. However, in strain GPE PC73, the albicidin biosynthesis gene cluster XALB1 does possess a *syrP* gene (XALc_1524), which shares 42% and 49% amino acid similarity with the *syrP* gene of *P. syringae* pv. *syringae* and the *syrP* gene of strain X11-5A, respectively. The *syrP* gene XALc_1524 may be involved in β-hydroxylation of residues that are assembled by XALB1, but also by META-A, META-B and META-C.

C-domains very similar to the starter C-domain of the locus META-B of strain GPE PC73 were found in strains XaS3, X11-5A and BAI-3, confirming that these three strains possess a NRPS gene cluster similar to META-B. Additionally, a C-domain very similar to the starter

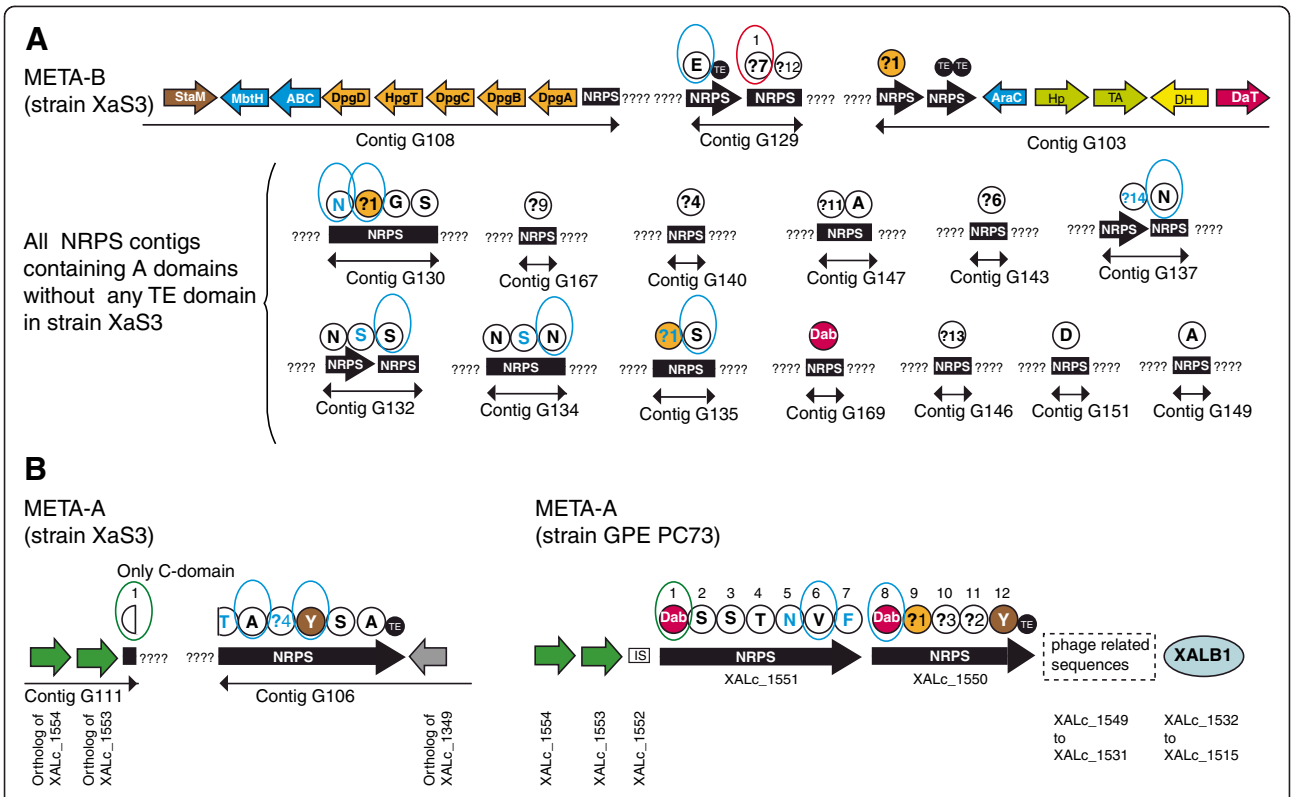


Figure 4 Representation of NRPS sequences of *Xanthomonas* spp. strain XaS3. A. Representation of contigs of *Xanthomonas* spp. strain XaS3 that contain a nucleotide sequence encoding at least one A domain and/or a nucleotide sequence encoding non-NRPS genes similar to genes found in the locus META-B of strain *X. albilineans* GPE PC73. **B.** Comparison of the locus META-A of *X. albilineans* strain GPE PC73 with the two contigs of *Xanthomonas* spp. strain XaS3 predicted to belong to a locus similar to this locus META-A. Brown arrows: *staM* and *staM'* genes predicted to be required for β-hydroxylation of two unidentified amino acids, respectively. Purple arrow: *daT* gene predicted to be required for biosynthesis of Dab (2,4-diamino butyric acid). Orange arrows: genes predicted to be required for biosynthesis of Dpg (3,5-dihydroxyphenylglycine). Dark blue arrows: other genes conserved in all META-B loci (ABC transporter, MbtH-like protein and transcriptional regulator AraC). Green arrows and yellow arrow: genes conserved in strains X11-5A and XaS3 predicted to be required for biosynthesis of an unknown non-proteinogenic amino acid (Hp: hypothetical protein gene; TA: transaminase gene; DH: lactate dehydrogenase gene). Black arrows: NRPS genes. Length of arrows is not proportional to the length of genes. All information regarding characteristics and color-code for NRPS modules are as detailed in Figure 2. IS: insertion sequence. XALB1: albicidin biosynthesis gene cluster. Incomplete circles: incomplete nucleotide sequence located at the end or at the beginning of a contig encoding only one part of a module.

C-domain of META-A was found in another contig of strain XaS3 (Figure 4). While sharing common characteristics, NRPS genes present in loci META-A, META-B or META-C in strains GPE PC73, BAI3, X11-5A or XaS3 are predicted to encode different NRPS systems, each being strain-specific. NRPS genes of each strain are described and discussed separately below.

Genomic features of NRPS genes associated with the META-B gene cluster of strain BAI3

The META-B locus of African strain BAI3 of *X. oryzae* pv. *oryzae*, which is situated next to a transcription activator-like (TAL) effector locus, contains three NRPS genes: NRPSa (14,484 bp), NRPSb (22,410 bp) and NRPSc (19,479 bp) (Figure 3). These three genes encode a NRPS system consisting of 17 modules of the characteristic domain arrangement C-A-PCP. One double TE-domain is present at the terminal module of NRPSa. Phylogenetic analysis identified the N-terminally located C-domain of the first module of NRPSb as the starter C-domain (Additional file 2). This C-starter domain shares 74% amino acid similarity with the starter C-domain of the META-B locus of strain GPE PC73 (Additional file 4). Therefore, the first module of NRPSb was assigned as the initiation module and the last module of NRPSa, which harbors the TE domain, was assigned as the termination module (Figure 3). Phylogenetic analysis of C-domains identified four modules exhibiting a dual C/E domain that are predicted to follow modules that assemble an amino acid with D-configuration in strain BAI3 (Figure 3; Additional file 1).

The META-B biosynthesis gene cluster of strain BAI3 also contains a pseudogene consisting of three degenerated NRPS modules that contain a frame-shift mutation or an insertion sequence, respectively (Figure 3). Ten out

Table 3 Similarities between NRPS-associated genes present in loci META-B of *Xanthomonas* spp. XaS3, *X. albilineans* strain GPE PC73 and *X. oryzae* pv. *oryzae* strains BAI3 and X11-5A

Proteins encoded in the locus META-B of *Xanthomonas* spp. strain XaS3	Predicted functions	Overall amino acid identities / similarities with proteins encoded by the locus META-B of *Xanthomonas* spp. XaS3		
		X. albilineans strain GPE PC73	*X. oryzae* pv. *oryzae* strain BAI3	*X. oryzae* pv. *oryzae* strain X11-5A
StaM	β-hydroxylation	98% / 99%	90% / 95%	90% / 94%
StaM'	β-hydroxylation	Not encoded	Not encoded	87% / 92%
MbtH	Assembly of amino acids by NRPS	95% / 96%	94% / 96%	94% / 96%
ABC	Secretion of the synthesized peptides	91% / 93%	78% / 83%	77% / 81%
DpgD	Biosynthesis of Dpg	94% / 96%	86% / 90%	86% / 90%
HpgT	Biosynthesis of Dpg	90% / 93%	80% / 88%	83% / 91%
DpgC	Biosynthesis of Dpg	92% / 95%	80% / 87%	81% / 88%
DpgB	Biosynthesis of Dpg	78% / 86%	69% / 79%	69% / 79%
DpgA	Biosynthesis of Dpg	92% / 94%	85% / 89%	86% / 90%
AraC	Transcriptional regulation	93% / 96%	57% / 72%	57% / 71%
DaT	Biosynthesis of Dab	95% / 97%	Not encoded	76% / 84%
Hypothetical protein	Unknown	Not encoded	Not encoded	76% / 84%
Transaminase	Unknown	Not encoded	Not encoded	74% / 82%
Lactate dehydrogenase	Unknown	Not encoded	Not encoded	79% / 89%

Dab: 2,4-diamino butyric acid; Dpg: 3,5-dihydroxyphenyl-glycine.

of 17 amino acids assembled by the locus META-B of strain BAI3 were predicted on the basis of analysis of specificity-conferring signatures of A-domains (Additional file 1; see Figure 3). Interestingly, one of the degenerated modules exhibits a signature specific to Dab, suggesting that a putative ancestor of strain BAI3 possessed the gene *daT* required for biosynthesis of Dab. Moreover, five unknown specificity-conferring signatures of A-domains were found in strain BAI3. Remarkably, the META-B locus is the only NRPS locus identified in the whole genome of strain BAI3.

Genomic features of NRPS genes associated with the META-B gene cluster of strain X11-5A

The contig 23.1 of the American strain X11-5A of *X. oryzae* pv. *oryzae* contains several NRPS-associated genes also present in the META-B gene cluster of strain GPE PC73. Two contigs (23.1 and 114.1) of strain X11-5A each contain a nucleotide sequence encoding a chain-terminating TE-domain located just upstream from a nucleotide sequence encoding a C-domain that shares more than 75% amino acid similarity with starter C-domains of the META-B loci of strains GPE PC73 and BAI3, and which segregates in the same phylogenetic clade as these starter C-domains (Additional files 2, 4). Therefore, both contigs 23.1 and 114.1 should belong to the META-B locus, suggesting that, in strain X11-5A, this locus possesses two starter C-domains and two chain-terminating TE domains, and consequently encodes two independent NRPS systems that should be involved in the biosynthesis of two different compounds. The older ancestor of META-B, considered similar to locus META-B of strain X11-5A, likely also encoded two NRPS systems. In strain BAI3, a pseudogene consisting of three degenerated NRPS modules, as well as a *syrP* pseudogene, may be remnants of lost ancestral genes that were required for a second NRPS system likely encoded by the older ancestor of META-B. Similarly, a pseudogene encoding a stand alone double TE-domain in strain GPE PC73 (XALc_1055) may be a remnant of a lost ancestral NRPS gene that was required for a second NRPS system. A total of 29 different nucleotide sequences encoding A-domains were found in all contigs of strain X11-5A (Figure 3; Additional file 1). The absence of nucleotide sequences encoding a TE-domain other than those encoded by contigs 23.1 and 114.1 suggests that all these 29 sequences encoding A-domains belong to the META-B gene cluster of strain X11-5A. Indeed, the total number of nucleotide sequences encoding A-domains belonging to this gene cluster may be even higher because of repeated DNA regions that are not included in the draft genome sequence and that may encode additional A-domains. Amino acids assembled by 14 of these 29 A-domains were predicted on the basis of specificity-conferring signatures of A-domains (Additional

file 1; see Figure 3). Seven unknown specificity-conferring signatures of A-domains were identified in strain X11-5A (Additional file 1). Phylogenetic analysis of C-domains identified six modules exhibiting dual C/E domains in this strain (see Figure 3; Additional file 1).

Genomic features of NRPS genes associated with gene clusters META-A and META-B of *Xanthomonas* spp. strain XaS3

The nucleotide sequence of contig G129 of *Xanthomonas* spp. strain XaS3 encodes a TE-domain located just upstream from a sequence encoding a starter C-domain. This C-domain shares more than 71% amino acid similarity with starter C-domains of the META-B loci of strains GPE PC73 and BAI3, which segregate in the phylogenetic clade of starter C-domains (Additional files 2 and 4). Therefore, contig G129 should belong to the META-B locus. Contigs G108 and G103 are also predicted to belong to the META-B gene cluster because they contain several NRPS-associated genes specific to this locus (Figure 4). Contig G103 contains the same gene as strain GPE PC73 encoding the stand alone double TE-domain predicted to be a pseudogene (ortholog of XALc_1055). Based on these resemblances with the META-B locus of strain GPE PC73, strain XaS3 is predicted to possess a META-B gene cluster (Figure 4) encoding one NRPS system involved in the biosynthesis of one peptide. Contig G111 of strain XaS3 contains a starter C-domain located at the same position as the starter C-domain of locus META-A of strain GPE PC73, and downstream genes orthologous to XALc_1553 and XALc_1554 (Figure 4). This starter C-domain shares 82% amino acid similarity with the starter C-domain of the locus META-A of strain GPE PC73 (Additional file 4). Contig G106 of strain XaS3 contains an incomplete NRPS gene encoding six NRPS modules, with the last module containing a chain-terminating TE domain. The 5′-part of this incomplete NRPS gene is located in a region that does not encode any NRPS in strain GPE PC73, upstream from the orthologous gene of XALc_1349 (Figure 4). However, based on phylogenetic analyses of C-domains, strain XaS3 possesses only two starter C-domains, indicating that contig G106 belongs to the same META-A gene cluster as the starter C-domain of contig G111 (Figure 4). The 5′-part of META-A in strain GPE PC73 is located upstream of both the terminus of replication and the albicidin biosynthesis gene cluster XALB1, which is not present in the genome of strain XaS3. Acquisition of XALB1 by the ancestor of *X. albilineans* or, alternatively, loss of XALB1 by the ancestor of strain XaS3 may explain why the upstream segments of META-A from strains GPE PC73 and XaS3 are located in two different regions. A total of 33 different nucleotide sequences encoding A-domains were found in all contigs of strain XaS3 (Figure 4; Additional file 1): four of these are located in contig G129, which contains nucleotide sequences encoding the chain-terminating TE-domain and the starter C-domain of META-B; and six are located in contig G106 containing the incomplete NRPS gene of META-A. The remaining 23 nucleotide sequences encoding A-domains, which are not located in contigs encoding TE-domains, may belong either to META-A or META-B. The total number of nucleotide sequences encoding A-domains in strain XaS3 may be higher than 33 because of repeated DNA regions that are not included in the draft genome sequence and that may encode additional modules. Amino acids assembled from 21 of these 33 A-domains were predicted on the basis of specificity-conferring signatures of A-domains (Figure 4; Additional file 1). Nine unknown specificity-conferring signatures of A-domains were found in strain XaS3 (Additional file 1). Phylogenetic analysis identified nine modules exhibiting a dual C/E domain (Figure 4; Additional file 1).

Genome mining reveals a NRPS gene cluster specific to *X. oryzae* pv. *oryzicola* strain BLS256

In *X. oryzae* pv. *oryzicola* strain BLS256, the region containing the XaPPTase gene also contains two large NRPS genes: XOC_2575 (9,513 bp) and XOC_2574 (11,622 bp). These two genes encode a NRPS consisting of six modules each with the three domains (C-A-PCP) followed by a chain-terminating double TE domain (Figure 5). Only the A, PCP, C and TE domains were identified and no auxiliary domains were found in this system. These genes are the only NRPS genes present in strain BLS256. Phylogenetic analysis of C-domains showed that the C-domain of the first module of XOC_2575 belongs to the same clade as starter C-domains identified by Rausch et al. [6]. Elongation modules of strain BLS256, that all exhibit a dual C/E domain, are predicted to catalyze epimerization of five amino acids (Figure 5; Additional file 1). The double TE-domain might catalyze cyclization of the peptide. Orthologs of NRPS genes of strain BLS256, which are present in *X. axonopodis* pv. *citri* strain 306, are only partially conserved and do not encode a complete NRPS system; these should be considered as pseudogenes (Figure 5).

Genome mining of *Xanthomonas sacchari* reveals genes required for biosynthesis of new non-ribosomal peptide(s)

A non-annotated draft sequence of the genome of *X. sacchari* strain NCPPB4393, isolated from an insect collected on a diseased banana plant, was published recently [45]. The species *X. sacchari*, which is phylogenetically related to *X. albilineans*, was also isolated on sugarcane and milled rice [46,47]. However, no disease caused by this species on any plant has been described to date. Interestingly, the genome of *X. sacchari* strain NCPPB4393 contains the XaPPTase gene (Table 1), which is not located at the same position as on the chromosomes

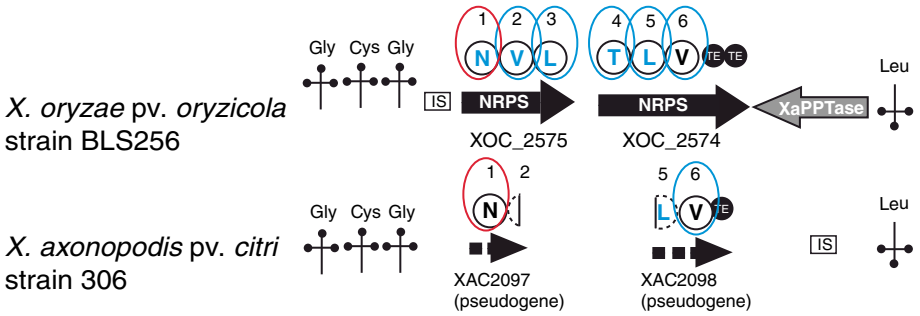


Figure 5 Representation of the NRPS locus present in the same region as XaPPTase in *X. oryzae* pv. *oryzicola* strain BLS256. Comparison with the similar genomic region of *X. axonopodis* pv. *citri* strain 306. NRPS modules are represented by circles. Large circles indicate complete NRPS modules (containing domains C, A and PCP). The amino acid predicted to be assembled by the corresponding module is indicated within each circle. Small black circles represent TE domains. Black arrows: NRPS genes. The starter module circled in red exhibits a C-domain belonging to the same phylogenetic clade as starter C-domains identified by Rausch et al. [6] (Additional file 2). Elongation modules circled in blue exhibit a C-domain belonging to the same phylogenetic clade as dual C/E domains identified by Rausch et al. [6] (Additional file 2). Amino acids predicted to be epimerized by these domains are in blue. ••: tRNA. Orientation of the tag indicates the orientation of the tRNA gene in the genomic regions. Amino acid specificity of each tRNA is indicated above or below each tag according to the orientation of the tRNA gene. IS: insertion sequence. Dotted arrows: pseudogenes. Dotted and incomplete circles: incomplete and non-functional modules.

of *X. albilineans* or *X. oryzae*. This genome also contains the genes *daT* (42% identical to XALc_1054), *staM* (38% identical to XALc_1066) and *syrP* (36% identical to XALc_1524). These three latter genes are clustered in a single locus. In addition, the genome of *X. sacchari* strain NCPPB4393 contains several contigs encoding NRPS-related sequences, among them adenylation domains specific to Dab, valine, glutamic acid, phenylalanine, asparagine, proline and an unknown substrate (unknown signature DAWLLGCTFK). However, the NCPPB4393 genome does not contain any sequences that are closely related to META-A, META-B, META-C, XALB1 or the short NRPS genes described above, indicating that non-ribosomal peptide(s) potentially produced by *X. sacchari* is/are specific to this species.

Genome mining by *in silico* analysis demonstrates a great biosynthetic potential for non-ribosomally synthesized peptides in the genus *Xanthomonas*

While sharing numerous common characteristics, the META-B gene clusters of all four investigated *Xanthomonas* spp. strains do not encode identical small molecules. This was concluded from the A-domain specificities of the NRPSs, and the hypothetical peptides are therefore considered strain-specific. Starter C-domains of all four loci META-B share at least 71% amino acid similarity (Additional file 4) suggesting that, although there is currently no experimental evidence, they might be expected to effect *N*-acylation with a structurally related β-hydroxy-carboxylic acid. Furthermore, the four META-B loci exhibit several C/E dual domains, indicating that they catalyze the assembly of residues with D-configuration. BLAST analyses identified only 188 GenBank entries that exhibit similar dual C/E domains. These 188 entries include nine genes analyzed in the present study (five NRPS genes of strain GPE PC73, two NRPS genes of strain BLS256, and two pseudogenes of strain 306 of *X. axonopodis* pv. *citri*). Draft genome sequence data of strains XaS3 and X11-5A did not indicate the number or distribution of residues in peptides assembled by their respective META-B gene cluster. However, data clearly suggest that these peptides are specific to each strain. In strain X11-5A, the META-B biosynthesis gene cluster is even predicted to control the biosynthesis of two different lipopeptides. A total of five different lipopeptides that do not resemble any compound described to date are predicted for the four *Xanthomonas* spp. strains due to the presence of the META-B gene cluster. These five lipopeptides could be secreted by the ABC transporter encoded by orthologs of XALc_1064, which are present in all META-B gene clusters.

The only common characteristic shared between the loci META-A and META-C of strain GPE PC73 and the locus META-A of strain XaS3, is their unusual starter C-domain, which segregates in a distant phylogenetic clade and shares less that 42% amino acid similarity with starter C-domains of META-B loci or with starter C-domains identified in numerous bacteria by Rausch et al. [6] (Additional file 4). These unusual starter C-domains might catalyze linkage of the first residue with a non-amino acid substrate. A total of three different compounds, not resembling any known compound, are predicted to be produced by *Xanthomonas* spp. strains GPE PC73 and XaS3 by the gene clusters META-A or META-C. Short NRPS genes of strain GPE PC73 might be involved in the biosynthesis of two different small molecules consisting of an amino acid (Gly for XALc_0364 and an unknown amino acid for XALc_1145) linked to a β-hydroxy

-carboxylic acid and a glycosyl residue. The only short NRPS gene of strain XaS3 might be involved in biosynthesis of a small molecule identical to that synthesized by XALc_0364 because of the high similarity of these two genes. According to *in silico* analyses, strain BLS256 could synthesize a cyclolipopeptide with the amino acid sequence D-Asn-D-Val-D-Leu-D-Thr-D-Leu-L-Val, which does not resemble any compound described to date (Figure 5). *X. sacchari* strain NCPPB4393 may also produce new non ribosomal peptide(s). Transporters involved in secretion of these unknown compounds remain to be identified as they are not encoded by the NRPS loci.

A total of 14 unknown specificity-conferring signatures of A-domains were found in the four strains GPE PC73, XaS3, X11-5A and BAI3. Only signatures "Unknown2" and "Unknown4" were found in a recently published database that includes 5,118 adenylation domains with unknown signatures [48]. The other 12 unknown NRPS signatures are not present in this database and could be specific to the genus *Xanthomonas*. The signature "Unknown1" is predicted to be specific for Dpg because of its presence in all four strains and because of the absence in these strains of any signature similar to those specific to the non-proteinogenic amino acid identified by [48]. Strains XaS3 and X11-5A possess a higher number of unknown signatures, suggesting that the three genes specific to both strains (transaminase, lactate dehydrogenase and a hypothetical protein) are involved in the biosynthesis of additional non-proteinogenic amino acids. However, these two strains share only one specific unknown signature ("Unknown9"), whereas some unknown signatures are shared by three strains ("Unknown2, -4, -6 and -7"). Eight other unknown signatures are present in only one strain ("Unknown3, -5, -8, -10, -11, -12, -13, and -14"). "Unknown5" was found only in the short gene XALc_1145 of strain GPE PC73. The presence of a high number of unknown signatures specific to the genus *Xanthomonas* suggests that numerous unknown non-proteinogenic amino acids are assembled by the novel NRPS genes identified in the genome of the four *Xanthomonas* spp. The biosynthetic capacities for those amino acids have yet to be identified. In conclusion, our results yield the expectation that numerous non-proteinogenic amino acids as well as residues with D-configuration should confer unusual structures to compounds synthesized by *Xanthomonas* spp. strains; identification of these compounds represents an exciting area for future study in these organisms.

PCR screening of 94 plant pathogenic bacteria for the XaPPTase gene and genes associated to NRPS in META-B

A collection of 94 strains belonging to the main genera of plant-pathogenic bacteria was screened by PCR. Primers designed to amplify DNA fragments of the XaPPTase gene, the ABC transporter gene present in META-B, the *dpgB* gene, the *dpgC* gene, the *hpgT* gene and the *daT* gene, were designed (Additional file 4). The XaPPTase gene was found in all analyzed strains of *X. albilineans* and *X. oryzae*. It was also found in strain CFBP2539 of *Xanthomonas translucens* pv. *secalis*, strain CFBP4642 of *Xanthomonas cassavae*, strain CFBP2431 of *Pseudomonas corrugata*, strain CFBP5593 of *Pseudomonas brassicacearum* and strain CFBP1192 of *Xylophilus ampelinus* (Additional file 5). The XaPPTase gene of *X. cassavae* is very similar to that of *X. oryzae*, while the XaPPTase gene of *P. corrugata* is very similar to that of *P. brassicacearum* (Additional file 5).

The ABC transporter gene was found in all four analyzed strains of *X. albilineans*, in all five analyzed African strains of *X. oryzae* pv. *oryzae*, in strains UPB497 and CFBP2286 of *X. oryzae* pv. *oryzicola* and in strain CFBP2539 of *X. translucens* pv. *secalis* (Additional file 5). Three additional genes of META-B (the *dpgB* gene, the *dpgC* gene and the *hpgT* gene) were found in all strains possessing the META-B ABC transporter, except for strain CFBP2286 of *X. oryzae* pv. *oryzicola*, which does not possess any of these genes (Additional file 5). The *daT* gene was found in all analyzed strains of *X. albilineans*, in strain UPB497 of *X. oryzae* pv. *oryzicola* and in strain CFBP2539 of *X. translucens* pv. *secalis* (Additional file 5). In summary, the META-B gene cluster seems to be present in all analyzed strains of *X. albilineans*, in all analyzed African strains of *X. oryzae* pv. *oryzae*, in strain UPB497 of *X. oryzae* pv. *oryzicola* and in strain CFBP2539 of *X. translucens* pv. *secalis*.

Conclusions

This study analyzed published genome sequences of *Xanthomonas* spp. together with unpublished draft genome sequences of *Xanthomonas* spp. XaS3 and *X. oryzae* pv. *oryzae* strain BAI3 for their genetic capacity to producing small molecules. Table 4 summarizes the *in silico* data on the non-ribosomally synthesized peptides predicted to be produced by these *Xanthomonas* strains. Unfortunately, because of the presence of unknown signatures and unusual starter C-domains, we were unable to predict possible structures of any of the products of the identified biosynthesis gene clusters. However, this study revealed that four strains of the genus *Xanthomonas* possess up to three novel homologous gene clusters, termed META-A, META-B or META-C, encoding NRPS peptides. A phosphopantetheinyl transferase (XaPPTase), which is essential for activation of NRPS enzymes, was identified in all strains investigated. Furthermore, sequence alignment of these META-clusters from *Xanthomonas* strains is indicative of the biosynthesis of lipopeptides or peptides linked to another non-amino acid substrate, and involving the biosynthesis and incorporation of non-proteinogenic

Table 4 *In silico* predicted data on non-ribosomally synthesized peptides potentially produced by *Xanthomonas* strains

Xanthomonas strains	Non-ribosomally synthesized peptides	
	Lipopeptides containing non-proteinogenic amino acids Dpg, Dab and/or at least one β-hydroxy-amino acid*	**Others**
X. albilineans strain GPE PC73	- Lipopeptide synthesized by META-B with the amino acid sequence Gly-Unk2-Thr-D-Asn-Tyr-Asn-D-Tyr-Ser-D-Asn-Asp-Asn-Dpg-Dpg-Ser-D-Thr-Unk2	- Glycine linked to a β-hydroxy-carboxylic acid and a glycosyl residue
	- Lipopeptide synthesized by META-A with the amino acid sequence Dab-Ser-Ser-Thr-D-Asn-Val-D-Phe-Dab-Dpg-Unk3-Unk2-Tyr	- Unknown amino acid linked to a β-hydroxy carboxylic acid and a glycosyl residue
	- Lipopeptide synthesized by META-C with the amino acid sequence Unk4-Phe-Dab-Asn-Ala-D-Unk3-Gln	
Xanthomonas spp. strain XaS3	- Two lipopeptides synthesized by META-B and META-A, respectively.**	- Glycine linked to a β-hydroxy-carboxylic acid and a glycosyl residue
X. oryzae pv. *oryzae* strain BAI3	-Lipopeptide synthesized by META-B with the amino acid sequence Gly-Unk7-Unk2-Thr-Asn-D-Thr-Tyr-D-Ile-Ser-Asn-D-Asn-Ser-Dpg-Dpg-Unk6-D-Dpg-Unk4	
X. oryzae pv. *oryzae* strain X11-5A	- Two lipopeptides synthesized by META-B.**	
X. oryzae pv. *oryzicola* strain BLS256		- A cyclolipopeptide with the amino acid sequence D-Asn-D-Val-D-Leu-D-Thr-D-Leu-Val
X. sacchari strain NCPPB4393		No precise data on the number and nature of the non-ribosomally synthesized peptides produced by this strain because of the lack of information (only draft genome sequence including several unassembled contigs that contain NRPS sequences)

Unk: Unknown amino acid; Dab: 2,4-diamino butyric acid; Dpg: 3,5-dihydroxyphenyl-glycine.
* The number and nature of amino acid(s) β-hydroxylated by proteins encoded by *staM*, *staM'* or *syrP* remain unidentified.
** Only partial amino acid sequences of these lipopeptides are available due to insufficient information (contigs containing gene clusters META-B or META-A are unassembled).

amino acids Dpg, Dab and amino acid(s) of unknown identity including at least one β-hydroxy-amino acid. The identity of several amino acids could not be predicted from the signature sequences of A-domain specificities. Hence, at least eight different peptides may be synthesized, the partial sequences of which have been predicted. This study revealed that each peptidic sequence is strain-specific (Table 4). This suggests that other *Xanthomonas* spp. strains, although they all possess the META-B locus, are expected to produce structurally different peptides, varying in size and amino acid composition. Small molecules synthesized by other NRPS genes analyzed in strain BLS256 of *X. oryzae* pv. *oryzicola*, strain NCPPB4393 of *X. sacchari* and strain GPE PC73 of *X. albilineans* are also likely candidates for the biosynthesis of new metabolites from *Xanthomonas* spp. strains (Table 4). If these biosynthetic pathways are functional, the resulting peptides may exert different biological functions in plant–bacteria interactions. Interestingly, these biosynthetic pathways seem to be shared only by strains of *Xanthomonas* associated with monocotyledonous plants, suggesting a putative involvement of novel non-ribosomally synthesized peptides in plant–bacteria interactions. The XaPPTase gene might be a useful probe for further genome mining of *Xanthomonas* spp. and related strains. In summary, this extensive *in silico* study shows that the genus *Xanthomonas* constitutes a promising reservoir of new non-ribosomally synthesized peptides. Experimental elucidation of this promising biosynthetic potential should contribute to the study of plant-bacteria interactions as well as to drug discovery. Interestingly, a first step of this elucidation was recently achieved with the isolation of the lipopeptide synthesized by the gene cluster META-B of *X. albilineans* strain GPE PC73. The nominal molecular mass of this lipopeptide is 2,293 Da and subsequent MS/MS-experiments revealed an amino acid sequence which excellently matches the one predicted by *in silico* analysis of A-domain specificities of META-B NRPSs (G.H. Völler, unpublished data).

Methods

Bacterial strains

X. albilineans strain GPE PC73 isolated in Guadeloupe was sequenced recently [27]. *X. albilineans* strains were grown for 48h on modified Wilbrink's medium [49] or on XAS selective growth medium [50] at 28°C. *X. oryzae* pv. *oryzae* strain BAI3 (isolated in Burkina Faso [33]) was grown for 24h on PSA medium [51] supplemented with appropriate antibiotics at 28°C.

Design of PCR primers

Primers used to screen a collection of 94 plant pathogenic bacteria for the presence of the XaPPTase gene as well as genes associated to NRPS in META-B by PCR were designed based on genome sequence information of strains *X. albilineans* GPE PC73, *X. oryzae* pv. *oryzae* BAI3, and *X. translucens* pv. *undulosa* UPB513 (Claude Bragard, unpublished data). Primer sequences are listed in Additional file 4. Primers used to determine the complete sequence of the 82-kb length region containing META-B in strain BAI3 were designed based on the draft genome sequence of *X. oryzae* pv. *oryzae* strain BAI3 (Genbank accession n° JQ348075).

PCR screening for the presence of the XaPPTase gene and genes associated to NRPS in META-B

DNA templates were prepared by suspending a freshly grown colony in 100 μl sterile nuclease-free water. PCR amplifications were performed in an automated thermal cycler (GeneAmp PCR System 9700; Life Technologies, Carlsbad, CA, USA). The 20-μl PCR reaction mix consisted of 5 μl bacterial suspension, 4 μl of 5x GoTaq buffer (Promega, Madison, WI, USA), 250 μM dNTP mix, 0.2 μM of each primer, 1 unit of GoTaq Polymerase (Promega, Madison, WI, USA), and sterile nuclease-free water to final volume. The PCR program was 94°C for 4 min, 35 cycles at 94°C for 30 seconds, Tm (melting temperature) for 30 seconds, and 72°C for 30 seconds, with a final 72°C extension for 8 min. A 8-μl aliquot of each amplified product was analyzed by electrophoresis through a 1.5% agarose gel. PCR products were sequenced with primers used for their respective amplification. Screened bacteria are listed in Additional file 6.

Genome sequencing

The draft genome sequence of strain BAI3 of *X. oryzae* pv. *oryzae* was determined by a Sanger/pyrosequencing hybrid approach. A shotgun library was constructed with 10-kb sized fragments obtained after mechanical shearing of the total genomic DNA, and cloned into the vector pCNS (pSU-derived). Sequencing with vector-based primers was carried out on an ABI 3730 Applera Sequencer. A total of 5,921 reads (~1 fold coverage) were analyzed and assembled with 518,656 (~23 fold coverage) 454 GS FLX reads (Roche Applied Science; http://www.roche.com). Sequence assembly was performed using Arachne "HybridAssemble" version (Broad Institute; http://www.broad.mit.edu/), which combines the 454 contigs with Sanger reads. As the Sanger reads contribution was not quite sufficient for the scaffolding, a mate-paired 454 library with 8-kb insert size was constructed and 179,755 (~4 fold coverage) 454 GS FLX reads were added into the assembly. To further improve quality, Illumina technology was applied (36-bp reads at ~50 fold coverage), eventually resulting in an assembly of 67 scaffolds with a mean scaffold length of 80 kb.

Draft genome sequence of *Xanthomonas* spp. strain XaS3 (also called strain GPE39), was obtained using short read technology only. After mechanical shearing of genomic DNA, fragments with an insert size of 300–500 bp were used for Illumina library construction (http://www.illumina.com). A total of 64,891,306 GAIIx reads were produced, corresponding to a paired-end sequencing with 2 × 76 nt reading lengths. Assembly was performed using a combination of the SOAPdenovo (1.05) and Velvet (1.1.04) short read assemblers [52,53]. In a first step, SOAPdenovo was run with kmer values ranging from 25 to 73 (step of 4) to generate contigs. In a second step, Velvet was run with parameters "-cov_cutoff 5 -min_contig_lgth 100 -max_divergence 0.05 -exp_cov auto -min_pair_count 10" to generate scaffolds, by using raw reads and SOAPdenovo contigs as input data. The range of kmer values used for Velvet was the same as for SOAPdenovo. The assembly with the maximum N50 value was selected, and remaining gaps were filled with SOAPGapCloser software. The XaS3 draft assembly spans 3,529,206 bp, and consists of 73 contigs of length ranging from 526 to 512,021 bp with a N50 of 168,272 bp. Annotated nucleotide sequences of the regions identified in these 73 contigs as containing A-domains and/or NRPS associated genes are provided in Additional file 7.

Determination of the complete sequence of the 82-kb length region containing META-B in *X. oryzae* pv. *oryzae* strain BAI3

On the basis of the draft genome sequence of strain BAI3, three scaffold DNA sequences were identified as belonging to META-B. The sizes of these DNA regions are 35,406 bp, 16,409 bp and 22,029 bp, respectively. Seventeen clones from a 10-kb shotgun library of strain BAI3 were identified by BLAST analyses to contain fragments of the META-B NRPS cluster. These clones are ABP0AAB3YI02, ABP0AAB3YP05, ABP0AAB8YK02, ABP0AAB8YE14, ABP0AAB5YJ09, ABP0AAB4YK08, ABP0AAB3YL17, ABP0AAB6YE03, ABP0AAB6YC14, ABP0AAB5YG12, ABP0AAB7YB12, ABP0AAB5YK06, ABP0AAB5YJ04, ABP0AAB7YD04, ABP0AAB3YG10, ABP0AAB5YE07 and ABP0AAB3YM11. Sequences of the insert ends were mapped onto the three scaffold DNA sequences. Four clones (ABP0AAB4YK08, ABP0AAB3YL17, ABP0AAB7YB12 and ABP0AAB5YK06) were identified as bridging two of the three scaffold DNA sequences. All other clones were mapped to within one of the three scaffold DNA sequences. Clones ABP0AAB4YK08, ABP0AAB3YL17, ABP0AAB7YB12 and ABP0AAB5YK06 were digested with restriction enzyme *Xho*I and partial digestion with restriction enzyme *Sal*I followed by re-ligation. The restriction site *Xho*I is

present only in the polylinker of pCNS vector and several *Sal*I restriction sites are present in the inserts. Clone ABP0AAB7YB12 was also digested with restriction enzyme *Bam*HI followed by re-ligation. The borders of the resulting clones were sequenced using universal primers (M13R for *Bam*HI borders and M13F for *Xho*I borders). One clone resulting from an internal *Sal*I-mediated deletion of clone ABP0AAB5YK06, and harboring an insert of 3.6 kb, was sequenced with primers MRK16, MRK17, MRK18, MRK19 and MRK20 (Additional file 8). A clone resulting from an internal *Sal*I-mediated deletion of clone ABP0AAB7YB12 and harboring an insert of 1.9 kb was sequenced with primers MRK19 and MRK20 (Additional file 8). Clone ABP0AAB5YK06 was sequenced with primers MRK7, MRK8, MRK18, MRK22R and MRK23R. The new sequence information allowed the three scaffold DNA sequences to be merged into a contiguous scaffold sequence of 82 kb. Contigs of strain BAI3 were mapped to this sequence and three sets of primers were designed in order to sequence (i) gaps between contigs, (ii) regions between NRPS and a *tal* gene, and (iii) regions with frameshifts, respectively (Additional file 8). These primers were used to sequence a clone resulting from an internal *Sal*I-mediated deletion of clone ABP0AAB4YK08 and harboring an insert of 5.0 kb and to sequence clones ABP0AAB3YG10, ABP0AAB3YM11, ABP0AAB5YE07, ABP0AAB5YJ04, ABP0AAB6YC14, ABP0AAB6YE03, ABP0AAB7YB12 and ABP0AAB7YD04 (Additional file 8). DNA sequencing was performed by Beckman Coulter Genomics (Takely, UK). Sequence reads were used to assemble the complete 81,740-bp META-B region of strain BAI3 (Genbank accession n° JQ348075).

In silico analyses of NRPS

Specificity of adenylation domains in NRPS and signatures were predicted using software available at http://nrps.informatik.uni-tuebingen.de/Controller?cmd=SubmitJob [54]. *In silico* analyses were performed on NRPS genes present in the finished annotated genome sequence of *X. albilineans* strain GPE PC73 (GenBank accession n°: NC_013722.1), in the non-annotated draft genome sequence of *X. oryzae* pv. *oryzae* strain X11-5A (GenBank accession n°: AFHK00000000.1; sequences analyzed in the current study are annotated in Additional file 9), in the unpublished annotated sequence of the 82-kb length region containing META-B in *X. oryzae* pv. *oryzae* strain BAI-3 (Genbank accession n° JQ348075), in the unpublished sequence of the contigs of strain *Xanthomonas* spp. XaS3 (sequences analyzed in the current study were annotated in Additional file 7), in the finished annotated genome sequence of *X. oryzae* pv. *oryzicola* strain BLS256 (nucleotide GenBank accession n°: CP003057.1), in the finished annotated genome sequence of *X. axonopodis* pv. *citri* strain 306 (nucleotide GenBank accession n°: AE008923.1), and in the non-annotated draft genome sequence of *X. sacchari* strain NCPPB4393 (nucleotide GenBank accession n°: AGDB00000000.1).

Phylogenetic analysis

The phylogenetic tree presented in Additional file 2 was reconstructed using the maximum likelihood method implemented in the PhyML program. The LG substitution model was selected, assuming an estimated proportion of invariant sites (of 0.01) and four gamma-distributed rate categories to account for rate heterogeneity across sites. The locus analyzed was the C-domain. Multiple alignments of the amino acid sequences of the C-domain and for all taxa were performed using ClustalW. The phylogenetic tree was calculated with PhyML; http://atgc.lirmm.fr/phyml/, version 2.4.4. Five hundred bootstrap replicates were performed with PhyML program. Data are available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.gh7h8.

Additional files

Additional file 1: Domains, signature sequences and predicted assembled residues of the NRPS genes described in this study. Sheet 1: Domains, signature sequences and predicted assembled residues of the NRPS genes present in loci META-A, META-B and META-C of *X. albilineans* strain GPE PC73. Sheet 2: Domains, signature sequences and predicted assembled residues of the NRPS genes present in the gene cluster META-B of *X. oryzae* pv. *oryzae* strain BAI3. Sheet 3: Domains, signature sequences and predicted assembled residues of the NRPS genes present in the NRPS contigs of *X. oryzae* pv. *oryzae* strain X11-5A. Sheet 4: Domains, signature sequences and predicted assembled residues of the NRPS genes present in the NRPS contigs of *Xanthomonas* spp. strain XaS3. Sheet 5: Domains, signature sequences and predicted assembled residues of the NRPS genes identified in the genome of *X. oryzae* pv. *oryzicola* strain BLS256 in the same region as the XaPPTase gene.

Additional file 2: Tree of the amino acid sequences of C-domains of strains GPE PC73, XaS3, X11-5A, BAI3, BTAi and BLS256 together with C-domains identified by Rausch et al. [6] as starter C-domains or as dual C/E domains. The tree was constructed using the maximum likelihood method and GTR as substitution model. Bootstrap percentages retrieved in 100 replications are shown at the main nodes. The scale bar (0.2) indicates the number of amino acid substitutions per site.

Additional file 3: Comparison of short NRPS genes and their associated overlapping glycosyltransferase genes of *X. albilineans* strain GPE PC73 with similar genes present in the genome of other bacteria. Table A: Comparison of short NRPS genes. **Table B**: Comparison of glycosyltransferase genes. Presence/absence of a gene similar to the XaPPTase gene in the genome of other bacteria.

Additional file 4: Primers used for PCR screening of a collection of 94 plant pathogenic bacteria for the presence of XaPPTase gene and several genes associated to NRPS in META-B.

Additional file 5: Summary of the results of the PCR screening of the collection of 94 plant pathogenic strains.

Additional file 6: Collection of strains screened for the presence of XaPPTase gene and genes associated with NRPS in META-B.

Additional file 7: Annotated nucleotide sequence of the regions encoding A-domains and/or NRPS associated genes in the contigs of strain XaS3.

Additional file 8: Primers used to determine the 82-kb length region containing META-B in strain BAI-3.

Additional file 9: Annotation of the contigs of the published draft genome sequence of *X. oryzae* pv. *oryzae* strain X11-5A which were analysed in the current manuscript.

Abbreviations

A-domain: Adenylation domain; ACP-domain: Acyl carrier protein domain; C-domain: Condensation domain; Cy-domain: Heterocyclization domain; Dab: 2,4-diamino butyric acid; Dpg: 3,5-dihydroxyphenyl-glycine; E-domain: Epimerization domain; C/E domain: Condensation/epimerization domain; NMR: Nuclear magnetic resonance; NRPS: Non-ribosomal peptide synthesis; NRPSs: Non-ribosomal peptide synthetases; MLSA: Multi locus sequence analyses; MS: Mass spectrometry; PCP-domain: Peptidyl carrier protein domain; P-pant: 4′-phosphopantetheinyl; PPTase: 4'-phosphopantetheinyl transferase; TAL: Transcription activator-like; TE-domain: Thioesterase domain.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MR supervised cloning experiments for determination of the complete sequence of the 82-kb length region containing META-B in strain BAI3, performed most of *in silico* analysis of NRPS loci, conceived the study and drafted part of the manuscript. RK discovered the presence of large NRPS in the genome of strain BAI3, contributed to determination and analysis of the sequence of contigs of strain BAI3 and drafted part of the manuscript. MM contributed to *in silico* analysis of NRPS loci and drafted part of the manuscript. VB provided sequencing reads used to determine the draft genome sequences of strains BAI3 and XaS3. SC (Carrere) contributed to determination of the sequence of contigs of strains BAI3 and XaS3. GPR, CG, MH, GHV and SR contributed to *in silico* analysis of NRPS loci. CB performed PCR screening of the collection of 94 plant pathogenic bacteria. JN performed cloning experiments for determination of the 82-kb length region containing META-B in strain BAI3. IP performed phylogenetic analysis and drafted part of the manuscript. VV conceived the sequencing project of strain BAI3. SP supervised PCR screening of the collection of 94 plant pathogenic bacteria. PR conceived the sequencing project of strains XaS3 and GPE PC73 and revised the manuscript. RDS and SC (Cociancich) conceived the study, contributed to *in silico* analysis of NRPS loci, supervised the preparation of the manuscript and drafted part of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Claude Bragard for access to sequence data of *X. translucens*. We thank Helen Rothnie for English editing. CG was supported by a post-doctoral fellowship awarded by the Agence Nationale de la Recherche, France (ANR-09-BLAN-0413-01). MM was supported by a PhD Fellowship awarded by the Région Languedoc-Roussillon and by the Centre de Coopération Internationale en Recherche Agronomique pour le Développement, France. GPR was supported by a PhD fellowship awarded by the Région Languedoc-Roussillon and by the Centre National de la Recherche Scientifique, France (grant 07-845). PR was supported by a Marie Curie Outgoing International Fellowship from the European Commission. The project was supported by a joint research grant of the Agence Nationale de la Recherche to MR (ANR-09-BLAN-0413-01) and the Deutsche Forschungsgemeinschaft to RDS (DFG SU239/11-1).

Author details

[1]CIRAD, UMR BGPI, Montpellier Cedex 5, F-34398, France. [2]IRD, UMR RPB, Montpellier Cedex F-34394, France. [3]CEA/DSV/IG/Genoscope, Centre National de Séquençage, Evry Cedex F-91057, France. [4]INRA, UMR IRHS, Beaucouzé F-49071, France. [5]INRA, UMR LIPM, Castanet-Tolosan Cedex F-31326, France. [6]Institut für Chemie, Technische Universität Berlin, Berlin D-10623, Germany. [7]UMR PVBMT, Université de la Réunion, Saint-Denis, La Réunion F-97715, France.

Received: 11 January 2012 Accepted: 22 September 2013
Published: 27 September 2013

References

1. Donadio S, Monciardini P, Sosio M: **Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics.** *Nat Prod Rep* 2007, **24:**1073–1109.
2. Gross H, Loper JE: **Genomics of secondary metabolite production by *Pseudomonas* spp.** *Nat Prod Rep* 2009, **26:**1408–1446.
3. Finking R, Marahiel MA: **Biosynthesis of nonribosomal peptides.** *Annu Rev Microbiol* 2004, **58:**453–488.
4. Sieber SA, Marahiel MA: **Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics.** *Chem Rev* 2005, **105:**715–738.
5. De Bruijn I, De Kock MJD, Yang M, De Waard P, Van Beek TA, Raaijmakers JM: **Genome-based discovery, structure prediction and functional analysis of cyclic lipopeptide antibiotics in *Pseudomonas* species.** *Mol Microbiol* 2007, **63:**417–428.
6. Rausch C, Hoof I, Weber T, Wohlleben W, Huson D: **Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution.** *BMC Evol Biol* 2007, **7:**78.
7. Balibar CJ, Vaillancourt FH, Walsh CT: **Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains.** *Chem Biol* 2005, **12:**1189–1200.
8. Witting K, Süssmuth RD: **Discovery of antibacterials and other bioactive compounds from microorganisms-evaluating methodologies for discovery and generation of non-ribosomal peptide antibiotics.** *Curr Drug Targets* 2011, **12:**1547–1559.
9. Challis GL, Ravel J, Townsend CA: **Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.** *Chem Biol* 2000, **7:**211–224.
10. May JJ, Kessler N, Marahiel MA, Stubbs MT: **Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases.** *Proc Natl Acad Sci U S A* 2002, **99:**12120–12125.
11. Stachelhaus T, Mootz HD, Marahiel MA: **The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.** *Chem Biol* 1999, **6:**493–505.
12. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT: **A new enzyme superfamily - the phosphopantetheinyl transferases.** *Chem Biol* 1996, **3:**923–936.
13. Gehring AM, Bradley K, Walsh CT: **Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then Acylated by EntE using ATP and 2,3-dihydroxybenzoate.** *Biochemistry* 1997, **36:**8495–8503.
14. Walsh CT, Gehring AM, Weinreb P, Quadri LEN, Flugel RS: **Post-translational modification of polyketide and nonribosomal peptide synthases.** *Curr Opin Chem Biol* 1997, **1:**309–315.
15. Quadri LEN, Sello J, Keating TA, Weinreb PH, Walsh CT: **Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin.** *Chem Biol* 1998, **5:**631–645.
16. Saddler GS, Bradbury JF: **Validation of publication of new names and new combinations previously effectively published outside the IJSEM.** *Int J Syst Evol Microbiol* 2005, **55:**2235–2238.
17. Rott P, Davis M: **Leaf scald**. In *A guide to sugarcane diseases*. Edited by Rott P, Bailey R, Comstock J, Croft B, Saumtally A. Montpellier: CIRAD-ISSCT; 2000:38–44.
18. Birch RG: ***Xanthomonas albilineans* and the antipathogenesis approach to disease control.** *Mol Plant Pathol* 2001, **2:**1–11.
19. Hashimi SM, Wall MK, Smith AB, Maxwell A, Birch RG: **The phytotoxin albicidin is a novel inhibitor of DNA gyrase.** *Antimicrob Agents Chemother* 2007, **51:**181–187.
20. Birch RG, Patil SS: **Antibiotic and process for the production thereof.** *US patent* 1985, 4,525,354.
21. Royer M, Costet L, Vivien E, Bes M, Cousin A, Damais A, Pieretti I, Savin A, Megessier S, Viard M, *et al*: **Albicidin pathotoxin produced by *Xanthomonas albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance renes.** *Mol Plant Microbe Interact* 2004, **17:**414–427.
22. Huang G, Zhang L, Birch RG: **Analysis of the genes flanking *xabB*: a methyltransferase gene is involved in albicidin biosynthesis in *Xanthomonas albilineans*.** *Gene* 2000, **255:**327–333.
23. Huang G, Zhang L, Birch RG: **A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans*.** *Microbiology* 2001, **147:**631–642.

24. Hashimi SM, Birch RG: **Functional analysis of genes for benzoate metabolism in the albicidin biosynthetic region of *Xanthomonas albilineans*.** *Appl Microbiol Biotechnol* 2010, **87:**1475–1485.
25. Huang G, Zhang L, Birch RG: **Albicidin antibiotic and phytotoxin biosynthesis in *Xanthomonas albilineans* requires a phosphopantetheinyl transferase gene.** *Gene* 2000, **258:**193–199.
26. Birch RG, Patil SS: **Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*.** *J Gen Microbiol* 1985, **131:**1069–1075.
27. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Couloux A, Darrasse A, Gouzy J, Jacques M-A, *et al*: **Genomic insights into strategies used by Xanthomonas albilineans with its reduced artillery to spread within sugarcane xylem vessels.** *BMC Genomics* 2012, **13:**658.
28. Triplett LR, Hamilton JP, Buell CR, Tisserat NA, Verdier V, Zink F, Leach JE: **Genomic analysis of *Xanthomonas oryzae* isolates from rice grown in the United States reveals substantial divergence from known *X. oryzae* pathovars.** *Appl Environ Microbiol* 2011, **77:**3930–3937.
29. Ochiai H, Inoue Y, Takeya M, Sasaki A, Kaku H: **Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity.** *Jpn Agric Res Q* 2005, **39:**275–287.
30. Lee B-M, Park Y-J, Park D-S, Kang H-W, Kim J-G, Song E-S, Park I-C, Yoon U-H, Hahn J-H, Koo B-S, *et al*: **The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice.** *Nucleic Acids Res* 2005, **33:**577–586.
31. Salzberg SL, Sommer D, Schatz M, Phillippy A, Rabinowicz P, Tsuge S, Furutani A, Ochiai H, Delcher A, Kelley D, *et al*: **Genome sequence and rapid evolution of the rice pathogen Xanthomonas oryzae pv. oryzae PXO99A.** *BMC Genomics* 2008, **9:**204.
32. Bogdanove AJ, Koebnik R, Lu H, Furutani A, Angiuoli SV, Patil PB, Van Sluys M-A, Ryan RP, Meyer DF, Han S-W, *et al*: **Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp.** *J Bacteriol* 2011, **193:**5450–5464.
33. Gonzalez C, Szurek B, Manceau C, Mathieu T, Séré Y, Verdier V: **Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa.** *Mol Plant Microbe Interact* 2007, **20:**534–546.
34. White FF, Potnis N, Jones JB, Koebnik R: **The type III effectors of *Xanthomonas*.** *Mol Plant Pathol* 2009, **10:**749–766.
35. Daugrois JH, Dumont V, Champoiseau P, Costet L, Boisne-Noc R, Rott P: **Aerial contamination of sugarcane in Guadeloupe by two strains of *Xanthomonas albilineans*.** *Eur J Plant Pathol* 2003, **109:**445–458.
36. Quadri LEN, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT: **Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases.** *Biochemistry* 1998, **37:**1585–1595.
37. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre J, Jaubert M, Simon D, Cartieaux F, Prin Y, *et al*: **Legumes symbioses: absence of Nod genes in photosynthetic *Bradyrhizobia*.** *Science* 2007, **316:**1307–1312.
38. Lautru S, Oves-Costales D, Pernodet J-L, Challis GL: **MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145.** *Microbiology* 2007, **153:**1405–1412.
39. Buchko GW, Kim C-Y, Terwilliger TC, Myler PJ: **Solution structure of Rv2377c-founding member of the MbtH-like protein family.** *Tuberculosis (Edinb)* 2010, **90:**245–251.
40. Pootoolal J, Thomas MG, Marshall CG, Neu JM, Hubbard B, Walsh CT, Wright GD: **Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009.** *Proc Natl Acad Sci USA* 2002, **99:**8962–8967.
41. Ikai H, Yamamoto S: **Identification and analysis of a gene encoding L-2,4-diaminobutyrate:2-ketoglutarate 4-aminotransferase involved in the 1,3-diaminopropane production pathway in *Acinetobacter baumannii*.** *J Bacteriol* 1997, **179:**5118–5125.
42. Pelzer S, Süssmuth R, Heckmann D, Recktenwald J, Huber P, Jung G, Wohlleben W: **Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908.** *Antimicrob Agents Chemother* 1999, **43:**1565–1573.
43. Tseng CC, Vaillancourt FH, Bruner SD, Walsh CT: **DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic pathway.** *Chem Biol* 2004, **11:**1195–1203.
44. Singh GM, Fortin PD, Koglin A, Walsh CT: **Beta-hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: characterization of two candidate hydroxylases AspH and SyrP in *Pseudomonas syringae*.** *Biochemistry* 2008, **47:**11310–11320.
45. Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J, Grant M: **Draft genome sequences of *Xanthomonas sacchari* and two banana-associated Xanthomonads reveal insights into the *Xanthomonas* group 1 clade.** *Genes* 2011, **2:**1050–1065.
46. Vauterin L, Hoste B, Kersters K, Swings J: **Reclassification of *Xanthomonas*.** *Int J Syst Bacteriol* 1995, **45:**472–489.
47. Ikeda S, Fuji S-I, Sato T, Furuya H, Naito H, Ytow N, Ezura H, Minamisawa K, Fujimura T: **Microbial diversity in milled rice as revealed by ribosomal intergenic spacer analysis.** *Microbes Environ* 2007, **22:**165–174.
48. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O: **NRPSpredictor2– a web server for predicting NRPS adenylation domain specificity.** *Nucleic Acids Res* 2011, **39:**362–367.
49. Rott P, Abel M, Soupa D, Feldman P, Letourmy P: **Population dynamics of *Xanthomonas albilineans* in sugarcane plants as determined with an antibiotic-resistant mutant.** *Plant Dis* 1994, **78:**241–247.
50. Davis MJ, Rott P, Baudin P, Dean JL: **Evaluation of selective media and immunoassays for detection of *Xanthomonas albilineans*, causal agent of sugarcane leaf scald disease.** *Plant Dis* 1994, **78:**78–82.
51. Soto-Suarez M, Bernal D, Gonzalez C, Szurek B, Guyot R, Tohme J, Verdier V: ***In planta* gene expression analysis of *Xanthomonas oryzae* pathovar *oryzae*, African strain MAI1.** *BMC Microbiol* 2010, **10:**170.
52. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, *et al*: **De novo assembly of human genomes with massively parallel short read sequencing.** *Genome Res* 2010, **20:**265–272.
53. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18:**821–829.
54. Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH: **Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs).** *Nucleic Acids Res* 2005, **33:**5799–5808.

doi:10.1186/1471-2164-14-658
Cite this article as: Royer *et al.*: Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genomics* 2013 **14**:658.

BRIEF COMMUNICATION
PUBLISHED ONLINE: 19 JANUARY 2015 | DOI: 10.1038/NCHEMBIO.1734

The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine

Stéphane Cociancich[1,3], Alexander Pesic[2,3], Daniel Petras[2,3], Stefanie Uhlmann[2], Julian Kretz[2], Vivien Schubert[2], Laura Vieweg[2], Sandrine Duplan[1], Mélanie Marguerettaz[1], Julie Noëll[1], Isabelle Pieretti[1], Manuela Hügelland[2], Sebastian Kemper[2], Andi Mainz[2], Philippe Rott[1], Monique Royer[1]* & Roderich D Süssmuth[2]*

Albicidin is a potent DNA gyrase inhibitor produced by the sugarcane pathogenic bacterium *Xanthomonas albilineans*. Here we report the elucidation of the hitherto unknown structure of albicidin, revealing a unique polyaromatic oligopeptide mainly composed of *p*-aminobenzoic acids. *In vitro* studies provide further insights into the biosynthetic machinery of albicidin. These findings will enable structural investigations on the inhibition mechanism of albicidin and its assessment as a highly effective antibacterial drug.

Bacteria of the genus *Xanthomonas* belong to a large group of Gram-negative plant pathogens that cause enormous economic losses in various crops, such as sugarcane, rice, soybean, cotton, wheat and citrus species[1]. *X. albilineans* is a xylem-invading plant pathogen that causes leaf scald disease in sugarcane[2]. Sugarcane is cultivated on 26 million ha in more than 60 locations around the world and is used for the production of sugar, bioethanol and rum[3]. Albicidin (**1**) is a phytotoxic compound of *X. albilineans*[4] that blocks DNA gyrase in sugarcane chloroplasts[5,6]. Apart from its importance for the pathogenicity of *X. albilineans*, albicidin is bactericidal at low nanomolar concentrations to a wide range of Gram-positive and Gram-negative bacteria, such as *Enterobacter aerogenes*, *Escherichia coli*, *Haemophilus influenza*, *Klebsiella pneumonia*, *Shigella sonnei* and *Staphylococcus aureus*[7]. Albicidin targets the bacterial DNA gyrase (topoisomerase II)[8], which is an essential enzyme for DNA replication, transcription and gene regulation because of its ability to modulate the extent of DNA supercoiling[9]. Albicidin is a potent inhibitor of this supercoiling activity, with a 50% inhibitory concentration of ~40 nM, which is comparable to the activity of other important DNA gyrase inhibitors such as the coumarins and quinolones[8]. Whereas the coumarins affect the ATPase activity of the GyrB subunit[10], albicidin interferes in the catalytic DNA cleavage-religation cycle of the GyrA subunit and traps a conformational state different than that targeted by quinolones[8]. These unique characteristics qualify albicidin as a potential lead compound for antibacterial drug development. However, structural characterization of albicidin has been impeded for decades because of the extremely low production rate of albicidin in its original host organism *X. albilineans* and the unknown substrates of its enigmatic hybrid polyketide synthase/nonribosomal peptide synthetase (PKS/NRPS) machinery[11]. The development of a viable heterologous expression system in *Xanthomonas axonopodis* pv. *vesicatoria* resulted in a sixfold increase in albicidin yields[11].

As the structural analysis by means of high-resolution MS (HR-MS) and NMR spectroscopy requires sufficient amounts of albicidin, we optimized the heterologous production of albicidin in *X. axonopodis* pv. *vesicatoria* and established an extended purification protocol using preparative HPLC (Online Methods). The purification procedure was guided by antibacterial tests to ensure isolation of bioactive albicidin (**Supplementary Results**, **Supplementary Table 1**). Final yields of purified albicidin were in the range of 1 mg per 100 l cell culture. Identical molecules were verified by comparison of albicidin samples produced by the original host *X. albilineans* and the heterologous host *X. axonopodis* pv. *vesicatoria* (**Supplementary Fig. 1**).

The structure of albicidin is depicted in **Figure 1a**. The phytotoxin is composed of the nonproteinogenic α-amino acid β-cyano-L-alanine (Cya-3) as well as the aromatic δ-amino acids *p*-aminobenzoic acid (pABA-2 and pABA-4) and 4-amino-2-hydroxy-3-methoxybenzoic acid (pMBA-5 and pMBA-6), respectively. The pentapeptide of these building blocks is N-terminally linked to 3-(4-hydroxyphenyl)-2-methyl acrylic acid, which is a methyl derivative of *p*-coumaric acid (MCA-1). HPLC-HR-Orbitrap ESI-MS of purified albicidin revealed for what is to our knowledge the first time the tentative molecular formula $C_{44}H_{39}O_{12}N_6$ $[M+H]^+$ (expected *m*/*z* 843.2636 $[M+H]^+$, calculated *m*/*z* 843.2621 $[M+H]^+$, Δ*m* 1.78 p.p.m.; **Supplementary Fig. 2**). To facilitate heteronuclear NMR experiments, we also prepared ^{15}N-labeled albicidin, which gave an expected *m*/*z* of 849.2478 $[M+H]^+$ (calculated *m*/*z* 849.2443 ($C_{44}H_{39}O_{12}{}^{15}N_6$) $[M+H]^+$, Δ*m* 4.12 p.p.m.; **Supplementary Fig. 2**). The observed difference of 6 Da thus confirms the presence of six nitrogen atoms in the molecule. The sequential walk by tandem MS experiments reveals the peptidic backbone of albicidin with *b*- and *y*-ion fragment series (**Fig. 1b**). Consequently, the accurate number of nitrogen atoms and molecular formulae are derived for each fragment by comparison of unlabeled and ^{15}N-labeled albicidin (**Fig. 1b**,**c**).

Various two-dimensional NMR experiments were performed to ultimately determine the constitution of albicidin (**Supplementary Table 2**). NMR chemical shifts and scalar coupling constants are listed in **Supplementary Table 3**. **Figure 1d** summarizes the observed through-bond and through-space NMR connectivities. In brief, the ^1H NMR spectrum of albicidin contains various aromatic signals (δ_H 6–8 p.p.m.) with coupling constants characteristic of *p*-substituted aromatic spin systems (**Supplementary Table 3**). The ^1H-^{15}N HMQC spectrum with ^{15}N chemical shifts in the range of 110–130 p.p.m. confirms the presence of five amide moieties (**Supplementary Fig. 3**). The sixth nitrogen atom is visible in the ^1H-^{15}N HMBC spectrum at δ_N 250 p.p.m. (**Supplementary Fig. 3**) and is part of the nitrile group of Cya-3. As shown in **Supplementary Figure 4**, chiral gas chromatography-MS analysis indicates an L configuration for Cya-3. The remaining four amide groups belong to the aromatic

[1]Cirad, UMR Biologie et Génétique des Interactions Plante-Parasite, Montpellier, France. [2]Institut für Chemie, Technische Universität Berlin, Berlin, Germany. [3]These authors contributed equally to this work. *e-mail: monique.royer@cirad.fr or suessmuth@chem.tu-berlin.de

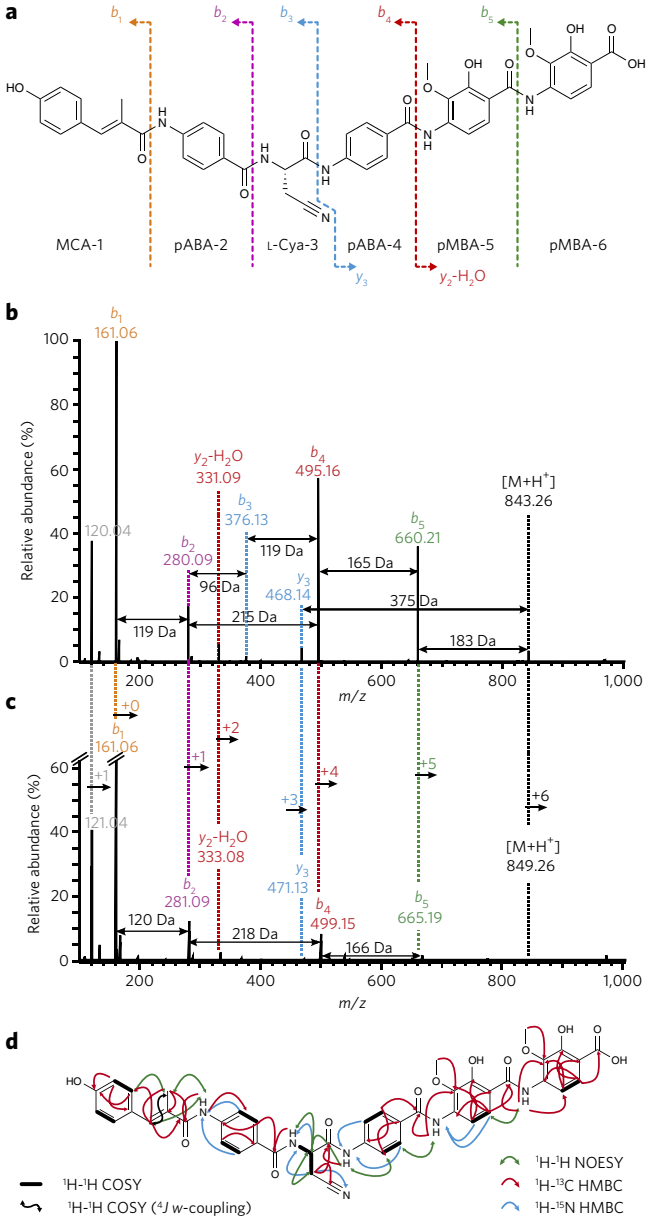


Figure 1 | Structural characterization of albicidin. (**a**) Albicidin is composed of a methyl derivative of *p*-coumaric acid (MCA-1), the nonproteinogenic α-amino acid L-Cya-3 as well as the aromatic δ-amino acids *p*-aminobenzoic acid (pABA-2 and pABA-4) and 4-amino-2-hydroxy-3-methoxybenzoic acid (pMBA-5 and pMBA-6). (**b**,**c**) The HR-ESI-(+)-Orbitrap-MS/MS fragment spectrum is shown for unlabeled albicidin (**b**) and ^{15}N-labeled albicidin (**c**). The observed *b*-ion and *y*-ion fragment series together with their mass shifts are color coded and indicated in the structure. (**d**) NMR spectroscopic through-bond (COSY, HMBC) and through-space correlations (NOESY) are illustrated in the albicidin structure.

spin systems pABA-2/pABA-4 and pMBA-5/pMBA-6, respectively (**Fig. 1d**). Finally, the sequential arrangement of the identified building blocks including the polyketide-derived MCA-1 at the N terminus of albicidin is unambiguously established by analysis of ^{1}H-^{13}C heteronuclear multiple bond coherence (HMBC) and ^{1}H-^{1}H NOESY data (**Supplementary Fig. 6**).

On the basis of the structure of albicidin and the arrangement of genes in the *alb* gene cluster (**Supplementary Table 4** and **Supplementary Fig. 7**), we propose the following pathway for the formation of albicidin (**Fig. 2a**): the N-terminal building block MCA-1 of albicidin is assembled by the concerted action of the PKS module of Alb01 (**Fig. 2b**), whereas NRPS-1 of Alb01 performs the coupling of thioester-bound MCA-1 to the subsequent building block, namely pABA-2. The delivery of pABA can be accomplished either by the cluster-intrinsic gene set of *alb17* and *alb18* or by their homologs from primary metabolism (*pabAB* and *pabC*, respectively)[12]. The substrate specificities of the involved NRPS modules were determined experimentally using a radioactivity-based ATP-PP$_i$ exchange assay (Online Methods). As illustrated in **Supplementary Figure 8**, the adenylation (A) domains of NRPS-1 (pABA-2) as well as NRPS-3 (pABA-4) indeed show the highest activity for the substrate pABA.

The biosynthesis of albicidin proceeds with the incorporation of L-Cya-3. The collinearity rule[13] assigns the A domain of NRPS-2 (Alb01) as responsible for this biosynthetic step. However, neither cyanoalanine nor asparagine were activated by NRPS-2 in the ATP-PP$_i$ exchange assay (**Supplementary Fig. 8**). Bioinformatic analysis suggests that NRPS-2 is most likely not functional in terms of a conventional A domain owing to the low sequence conservation of essential core motifs (**Supplementary Fig. 9**). Instead, we assign the incorporation of L-Cya-3 to the single-standing and thus *trans*-acting NRPS-2* (Alb04), which is essential for albicidin biosynthesis[2]. Such a *trans*-complementing mechanism has already been described, for example, for bleomycin biosynthesis[14]. In agreement with *in silico* predictions (**Supplementary Fig. 10**), NRPS-2* reveals specificity for cyano-L-alanine and has even more pronounced specificity for L-asparagine in the *in vitro* assay (**Supplementary Fig. 8**). It should be taken into account that the ATP-PP$_i$ exchange assay generally yields end-point data only and may thus imply activation of less preferred substrates. The A domain of module NRPS-2* contains an unusual 342-amino-acid insertion harboring an ATP-binding motif (SGGKD), which is strongly conserved among members of the adenosine nucleotide α-hydrolase (α-ANH-like III) superfamily (**Supplementary Fig. 9**)[15]. Hence, we postulate a two-step scenario for the activation and *in situ* processing of L-asparagine to cyano-L-alanine by NRPS-2*: (i) adenylation of the α-carboxylic acid moiety and storage as a thioester and (ii) phosphorylation of the side chain amide oxygen (**Supplementary Fig. 11**). Subsequent dephosphorylation would cause a formal elimination of water and thus the formation of cyano-L-alanine, which is then attached to the growing peptide chain.

The following biosynthetic step includes the activation and coupling of pABA-4 by NRPS-3 (Alb01), consistent with the *in vitro* substrate specificity of its A domain (**Supplementary Fig. 8**). Finally, residues pMBA-5 and pMBA-6 are incorporated by modules NRPS-4 and NRPS-5 (Alb09), respectively, with the thioesterase domain of NRPS-5 mediating the release of albicidin. Although pABA is used as a direct substrate by NRPS-1 and NRPS-3 (Alb01), our experimental data imply that NRPS-4 and NRPS-5 (Alb09) preferentially activate the substrates 2- and 3-hydroxy-pABA rather than pMBA (**Supplementary Fig. 8**). We therefore propose prior formation of 3-hydroxy-pABA (pAHBA) from pABA through hydroxylation by the benzoyl-CoA oxygenase–like enzyme Alb12 (**Fig. 2b**). Subsequently activated pAHBA, which is attached to the T domains of NRPS-4 or NRPS-5, is most likely processed to pMBA by hydroxylation and methylation through the putative β-hydroxylase Alb08 and the essential methyltransferase Alb02 (ref. 16), respectively (**Fig. 2b**). Notably, the multiple sequence alignment of several T domains reveals a nonconserved insertion of 13 amino acids for both T domains of NRPS-4 as well as NRPS-5 (**Supplementary Fig. 9**). Given published T domain structures[17,18], the sequence insertion apparently extends the loop region between helices I and II and might reflect additional protein-protein interaction modes in the course of *in situ* pAHBA processing.

In addition, multiple sequence alignment of all of the A domains of the albicidin gene cluster (**Supplementary Fig. 9**) demonstrates that NRPS-1 and NRPS-3 as well as NRPS-4 and NRPS-5 share

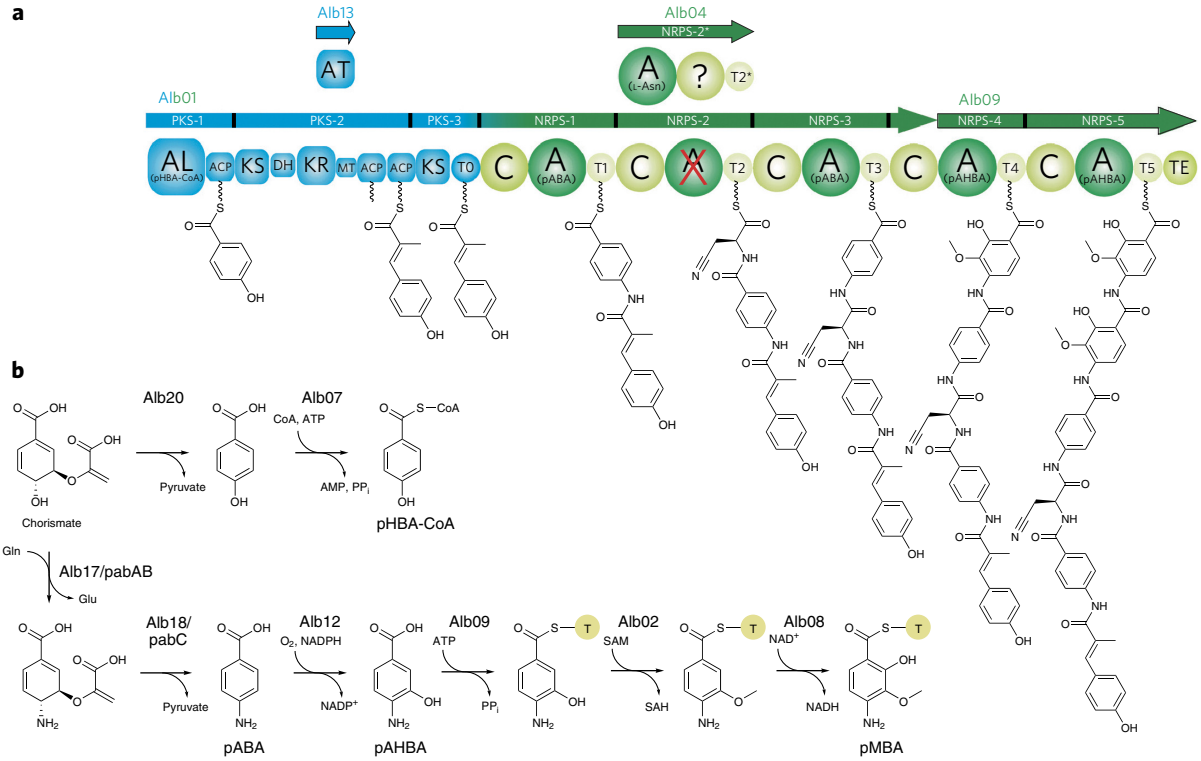


Figure 2 | Model of albicidin biosynthesis. (**a**) Proposed biosynthetic assembly line for albicidin. Substrates of the NRPS are indicated at the A domains. PKS and NRPS modules are colorcoded in blue and green, respectively. The PKS-NRPS hybrid consists of the following domains: acyl-CoA ligase (AL), acyl carrier protein (ACP), *trans*-acting acyltransferase (AT), ketoreductase (KR), ketosynthase (KS), dehydrogenase (DH), methyltransferase (MT), peptide carrier (T), condensation (C), adenylation as well as thioesterase (TE) domains. (**b**) Proposed pathways for the biosynthesis of the MCA-1 precursor pHBA-CoA and for the δ-amino acids pABA and pMBA.

virtually 100% sequence identity in functional core motifs that have been previously described to confer substrate specificity[13]. This observation is consistent with the similar chemical structures of the building blocks pABA and pMBA. Notably, the activation of aromatic δ-type amino acid substrates rather than α-amino acid substrates appears to arise from an altered localization of a key aspartic acid residue interacting with the corresponding amino moiety of the substrates (**Supplementary Fig. 10**)[19].

In the search for new potent antibiotics, albicidin represents a fascinating lead structure. This work uncovers the unique chemical structure of a new type of gyrase inhibitor and reveals *p*-aminobenzoic acids as a new class of substrate for NRPSs. The foundations are laid for further work on the biosynthesis and chemical synthesis of albicidin and analogs, which can then be followed by structure-activity relationship studies as well as investigations on the mode of action by X-ray crystallography.

Received 18 July 2014; accepted 20 November 2014;
published online 19 January 2015

Methods

Methods and any associated references are available in the online version of the paper.

References

1. Young, J.M., Park, D.C., Shearman, H.M. & Fargier, E. *Syst. Appl. Microbiol.* **31**, 366–377 (2008).
2. Royer, M. *et al. Mol. Plant Microbe Interact.* **17**, 414–427 (2004).
3. Dal-Bianco, M. *et al. Curr. Opin. Biotechnol.* **23**, 265–270 (2012).
4. Birch, R.G. & Patil, S.S. *Phytopathology* **73**, 1368–1374 (1983).
5. Birch, R.G. & Patil, S.S. *Physiol. Mol. Plant Pathol.* **30**, 207–214 (1987).
6. Birch, R.G. & Patil, S.S. *Physiol. Mol. Plant Pathol.* **30**, 199–206 (1987).
7. Birch, R.G. & Patil, S.S. Antibiotic and process for the production thereof. US patent 4,525,354 (1985).
8. Hashimi, S.M., Wall, M.K., Smith, A.B., Maxwell, A. & Birch, R.G. *Antimicrob. Agents Chemother.* **51**, 181–187 (2007).
9. Corbett, K.D. & Berger, J.M. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 95–118 (2004).
10. Maxwell, A. & Lawson, D.M. *Curr. Top. Med. Chem.* **3**, 283–303 (2003).
11. Vivien, E. *et al. Antimicrob. Agents Chemother.* **51**, 1549–1552 (2007).
12. Dosselaere, F. & Vanderleyden, J. *Crit. Rev. Microbiol.* **27**, 75–131 (2001).
13. Stachelhaus, T. & Marahiel, M.A. *FEMS Microbiol. Lett.* **125**, 3–14 (1995).
14. Du, L., Sanchez, C., Chen, M., Edwards, D.J. & Shen, B. *Chem. Biol.* **7**, 623–642 (2000).
15. Aravind, L., Anantharaman, V. & Koonin, E.V. *Proteins* **48**, 1–14 (2002).
16. Huang, G., Zhang, L. & Birch, R.G. *Gene* **255**, 327–333 (2000).
17. Koglin, A. *et al. Science* **312**, 273–276 (2006).
18. Tanovic, A., Samel, S.A., Essen, L.O. & Marahiel, M.A. *Science* **321**, 659–663 (2008).
19. Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. *Chem. Biol.* **6**, 493–505 (1999).

Acknowledgments

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG SU239/11-1; SU 18-1), by the Cluster of Excellence 'Unifying Concepts in Catalysis (UniCat)' (DFG) and by a grant from the Agence Nationale de la Recherche (ANR-09-BLAN-0413-01).

Author contributions

S.C., A.P., S.D., M.M., J.N., I.P., M.H., P.R. and M.R. performed the cultivation and isolation of albicidin. A.P., D.P., V.S., S.K. and A.M. performed the structural elucidation of albicidin. D.P., S.U., A.M. and M.R. performed the bioinformatic analysis of the biosynthesis genes. D.P., S.U. and M.H. produced the proteins and performed the *in vitro* assays. J.K. synthetized the substrates for the *in vitro* assay. S.C., A.P., D.P., A.M., M.R. and R.D.S. designed the study and analyzed the data. S.C., A.P., D.P., L.V., A.M., M.R. and R.D.S. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information and chemical compound information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to M.R. or R.D.S.

ONLINE METHODS

Fermentation of heterologous albicidin producer. Liquid cultures of a heterologous albicidin-producing strain (strain Xves-alb of *Xanthomonas axonopodis* pv. *vesicatoria*)[11] were prepared in plastic tubes (Corning Inc., Corning, New York, USA) of 500 ml to avoid adsorption of albicidin to the glass surface. Cultures were grown under agitation for 5 d at 28 °C in 72 tubes each containing 200 ml (14.4 l) of an optimized medium (XVM3B medium; **Supplementary Table 5**). Several of these batches had to be prepared for purification of unlabeled albicidin corresponding to ~1,500 tubes of 200 ml medium each (~ 300 l medium in total). For preparation of ^{15}N-labeled albicidin, the same medium was used for feeding of isotopically labeled $(^{15}NH_4)_2SO_4$ (^{15}N-XVM3B medium;**Supplementary Table 5**).

Isolation and purification of albicidin. Isolation of albicidin was adapted from a previous procedure[7,20]. Briefly, XAD-7 Amberlite was added to the fermentation broth of strain Xves-alb to adsorb albicidin from the supernatant. Albicidin was eluted from XAD-7 during a MeOH/H_2O step gradient (20–100% MeOH, Δc = 20%) at a final concentration of 100% MeOH. Bioactivity-guided fractions were concentrated in a Genevac Speedvac (Ipswich, Great Britain) and redissolved in MeOH. After centrifugation, the supernatant was purified by preparative HPLC on an Agilent 1100 system (Agilent, Waldbronn, Germany) at a detection wavelength of 310 nm on a C18 reversed phase column (GromSil 120 ODS 5 ST, 10 µm; 250 × 20 mm Grace, Rottenburg-Hailfingen, Germany) using a linear MeOH gradient starting from 35% MeOH plus 0.1% HCOOH to 80% MeOH plus 0.1% HCOOH for 40 min with a flow rate of 15 ml/min. Albicidin eluted at a retention time R_t of 33 min. The fractions were tested for antibacterial activity, freeze-dried and redissolved in aqueous tetrahydrofuran (44%) with acetic acid (1%) and purified using a semipreparative Agilent 1100 HPLC system (Agilent, Waldbronn, Germany) at a detection wavelength of 310 nm, using isocratic conditions on a polymeric reversed phase (PRP-1, 5 µm; 305 × 7 mm, Hamilton Bonaduz, Switzerland). Albicidin eluted at a retention time R_t of 68 min and was obtained as a white solid (3 mg unlabeled albicidin and <1 mg ^{15}N-labeled albicidin).

Final yields of isolated albicidin were in the range of 1 mg per 100 l cell culture. The previous description of the albicidin heterologous host claimed higher yields (~1 mg per 1 l cell culture)[11]. These yields were measured directly in the cell culture supernatant, which does not take into account unavoidable losses during the purification process. Moreover, it seems that the formula used to quantify albicidin[21] overestimated the effective amount of albicidin in the supernatant.

Antibacterial assay. Petri dishes containing 20 ml of LB agar were overlaid with a mixture of 2 ml of exponential phase bacterial culture (2×10^7 cells/ml in LB medium) with 2 ml of molten 1.5% noble agar. 20 µl of various concentrations of albicidin were deposited in 5-mm-diameter wells cut in the overlaid plates. After overnight incubation at 37 °C, the width of the crown of growth inhibition surrounding each well was recorded. Bacterial strains used in the assay are described in **Supplementary Table 1**.

Total hydrolysis and chiral GC/MS Total hydrolysis of albicidin and L-cyanoalanine was performed at 110 °C in aqueous 6 M hydrochloric acid solution (200 µl) for 6 h in glass vials. After hydrolysis, the hydrochloric acid was removed in a gentle stream of nitrogen.

For chiral GC/MS analysis of hydrolysates as well as of D-aspartic acid and L-aspartic acid, the amino acids were transformed into *N*-trifluoroacetyl/ethyl ester derivatives. In short, a total volume of 150 µl of 2 M ethanolic HCl solution was added. Then, the samples were heated for 30 min at 110 °C, and reagents were removed at 110 °C in a gentle stream of nitrogen. Acetylation was performed by adding 100 µl dichloromethane and 50 µl trifluoroacetic anhydride. The mixtures were heated again for 10 min at 110 °C. Excess reagent was removed at ambient temperature in a stream of nitrogen. Resulting residues were dissolved in 50 µl anhydrous toluene, and 1 µl was subjected to GC/MS (5975C, Agilent Technologies, Waldbronn, Germany) using an Agilent J&W CP-Chirasil L-Val GC column, Antipode D, 0.25mm, 25 m, 0.08 µm (Agilent Technologies, Waldbronn, Germany) with a split ratio of 1:50.

The temperature gradient started with an initial hold for 5 min at 70 °C followed by an increase of 2 °C/min up to 100 °C and then 3.5 °C/min up to 190 °C with a 10-min hold. The column was then heated to 200 °C within 1 min with a 1-min hold at 200 °C. Selected ion monitoring (SIM) was set to 140 *m/z* and 212 *m/z* in positive electron impact (EI) mode (5 min solvent delay, MS source: 300 °C, MS quad: 150 °C, EMV mode, gain factor 1, EM voltage 1388 V) with a flow of 1.2 ml/min with helium as carrier gas.

HPLC-HR-MS. Full-scan measurements were routinely performed on an Exactive ESI-Orbitrap-MS (Thermo Fisher Scientific GmbH, Bremen, Germany) coupled to an analytical Agilent HPLC 1200 system (Agilent, Waldbronn, Germany) equipped with a Hypersil-Gold (5 µm, 50 × 2.1 mm, Thermo Fisher Scientific GmbH, Bremen, Germany) using MeOH and H_2O + 0.1% HCOOH as mobile phase. Albicidin was analyzed during a linear gradient from 5% to 100% MeOH for 6 min followed by a washout phase of 100% MeOH for 4 min and a re-equilibration phase at 5% MeOH for 2 min. The flow rate was set to 0.25 ml/min. MS conditions were R = 100,000 (at 200 *m/z*) with a maximal C trap fill time of 500 ms. The mass range was set to 150–1,500 *m/z*.

HPLC-HR-MS/MS. MS/MS measurements were performed on an LTQ-Orbitrap XL hybrid instrument (Thermo Fisher Scientific GmbH, Bremen, Germany) coupled to an analytical HPLC 1200 system (Agilent, Waldbronn, Germany) with a Hypersil-Gold (5 µm, 50 × 2.1 mm, Thermo Fisher Scientific GmbH, Bremen, Germany) using the same gradient as described above for full scan measurements. R was set to 30,000 (at 400 *m/z*) with a maximal C trap fill time of 200 ms. HCD collision energy was set to 18% with 30 ms activation time and an isolation window of 2 *m/z* for the precursor selection (843.3 *m/z* and 849.3 *m/z*). The mass range was set to 100–1,000 *m/z*.

HPLC-QqQ-SRM. Selected reaction monitoring (SRM) of XAD extracts was carried out on an Agilent 6460 triple quadrupole instrument (Agilent Technologies, Waldbronn, Germany). A C18 Core-shell column (Supelco Ascentis Express, 100 × 2.1 mm, 5 µm, Sigma Aldrich, St. Louis, MO, USA) with a linear gradient from 20–95% ACN over 5 min was used, followed by a 2-min washout and 2 min re-equilibration, with a flow rate of 0.3 ml/min. ESI was operated in negative mode with –3,500 V capillary voltage and –1,200 V nozzle voltage. For the SRM transient, a unit precursor selection window at 841.2 *m/z* and a MS2 *m/z* value of 362.1 for a typical albicidin product ion were used. The normalized collision energy was set to 38. Dwell time was set to 200 ms. The fragmentator voltage and the cell accelerating voltage were adjusted to 135 V and 7 V, respectively.

NMR spectroscopy. NMR spectra of unlabeled and ^{15}N-labeled albicidin were acquired on an Avance III 700 MHz NMR spectrometer (Bruker, Karlsruhe, Germany) equipped with a TXI 1.7-mm room temperature probe head. Samples of 50 µl in d_8-tetrahydrofuran (THF) were filled into 1.7 mm NMR tubes. All experiments were performed at 298 K. Chemical shifts were referenced relative to tetramethylsilane (TMS; ^{1}H, ^{13}C) and liquid NH_3 (^{15}N), respectively. **Supplementary Table 2** and **Supplementary Table 3** summarize relevant NMR acquisition parameters and the NMR data for albicidin, respectively. All NMR experiments used standard parameter sets employed by the manufacturer (Bruker, Karlsruhe, Germany).

Bioinformatic analysis. DNA sequence information was obtained from a previous study[2]. Nucleotide sequences were translated using the software Clonemanager (Sci-Ed Software, Cary, NC, USA). Putative protein functions were determined using pBLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi)[22]. Multiple alignments of NRPS were performed with Clustal W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/)[23]. The final adjustment to the sequence of GrsA was performed manually. NRPS substrate prediction was drawn out using the PKS-NRPS predictor of the University of Maryland (http://nrps.igs.umaryland.edu/nrps/)[24]. Structural prediction of A-domain binding pockets was carried out using I-TASSER[25].

Construction of expression plasmids. The *alb04* gene (A/T didomain of NRPS-2*) and selected parts of genes *alb01* (A/T didomain of NRPS-1, NRPS-2 and NRPS-3) and *alb09* (A domains of NRPS-4 and NRPS-5), respectively, were amplified from the cosmid pALB571 (ref. 11) by PCR with Herculase polymerase (Thermo Scientific, Waltham, Massachusetts, USA) using the primers described in **Supplementary Table 6**. PCR products were digested with restriction enzymes *Not*I and NcoI, cloned into the vector pETtrx_1c[26] or pQE80 and subsequently transformed into *E. coli* BL21 gold by either electroporation or heat shock. Transformed clones were selected on kanamycin

doi:10.1038/nchembio.1734

or ampicillin LB-agar plates (50 µg/ml) and verified by plasmid purification following restriction analysis and DNA sequencing of the inserted genes.

Gene expression and protein purification. The *E. coli* expression strains containing the vectors pETtrx_1c-NRPS-1, -NRPS-2*, -NRPS-3, -NRPS-4, -NRPS-5 and pQE80_NRPS-2 were incubated in 500 ml TB medium containing 0.2% lactose (37 °C, 200 r.p.m. for 2 h following 10 h at 18 °C) to induce *lac*-controlled gene expression. Subsequently, cells were harvested by centrifugation at 7,000*g* for 10 min. Cell pellets were resuspended in wash buffer (50 mM Tris-HCl, pH 7.8, 300 mM NaCl) containing 5% (v/v) glycerol. Cell lysis was achieved by the addition of lysozyme and sonication (3 × 2 min; 66% intensity), and the resulting cell debris was separated by centrifugation at 50,000*g* for 20 min. The supernatant was purified by Ni-affinity chromatography using an Äkta purifier (GE Healthcare, Little Chalfont, UK). Elution was performed on 1-ml or 5-ml His-Trap columns (Qiagen, Venlo, The Netherlands or GE Healthcare, Little Chalfont, UK). After loading the sample onto the column at a flow rate of 1 ml/min or 3 ml/min, the proteins were washed (10 column volumes (CV) 30 mM imidazole) and eluted with an imidazole step gradient (10 CV 200 mM imidazole and 10 CV 500 mM imidazole). Elution fractions were combined and submitted to SDS-PAGE (**Supplementary Fig. 12**). The identity of the fusion proteins was confirmed by western blotting against the poly-His tag (**Supplementary Fig. S12**). Subsequent to Ni-affinity purification, the fractions were pooled and dialyzed with an ultrafiltration device (10 kDa molecular weight cutoff, Merck-Millipore, Billerica, MA, USA) using wash buffer. Protein concentrations were determined via intrinsic absorption at 280 nm by using a NanoPhotometer (Implen, München, Germany). After addition of 20% glycerol, aliquots were shock-frozen in liquid nitrogen and stored at –80 °C for further experiments.

ATP-PP_i exchange assay. ATP mix (100 µl containing 0.1 mM PP_i, 10 mM ATP, 10 mM $MgCl_2$ and ^{32}P-labeled PP_i set to 200,000 counts/100 µl) were mixed with 20 µl of 100 mM amino acid substrates. After the addition of 100 µl ~50 µM enzyme solution, samples were incubated for 15 min at 30 °C. The reaction was stopped by adding 1 ml stop solution (1.4% activated charcoal, 3.8% perchloric acid and 12.2 g/l Na_4PP_i). ATP bound to activated charcoal was collected by centrifugation, washed three times with 1 ml H_2O and subsequently measured using a scintillation counter (PerkinElmer, Waltham, MA, USA) following the manufacturer's instructions.

20. Birch, R.G. & Patil, S.S. *J. Gen. Microbiol.* **131**, 1069–1075 (1985).
21. Zhang, L., Xu, J. & Birch, R.G. *J. Appl. Microbiol.* **85**, 1023–1028 (1998).
22. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. *J. Mol. Biol.* **215**, 403–410 (1990).
23. McWilliam, H. *et al. Nucleic Acids Res.* **41**, W597–W600 (2013).
24. Bachmann, B.O. & Ravel, J. *Methods Enzymol.* **458**, 181–217 (2009).
25. Zhang, Y. *BMC Bioinformatics* **9**, 40 (2008).
26. Bogomolovas, J., Simon, B., Sattler, M. & Stier, G. *Protein Expr. Purif.* **64**, 16–23 (2009).

frontiers in Plant Science

MINI REVIEW
published: 24 April 2015
doi: 10.3389/fpls.2015.00289

What makes *Xanthomonas albilineans* unique amongst xanthomonads?

Isabelle Pieretti[1], *Alexander Pesic*[2], *Daniel Petras*[2], *Monique Royer*[1], *Roderich D. Süssmuth*[2] *and Stéphane Cociancich*[1]*

[1] UMR BGPI, Cirad, Montpellier, France, [2] Institut für Chemie, Technische Universität Berlin, Berlin, Germany

Xanthomonas albilineans causes leaf scald, a lethal disease of sugarcane. Compared to other species of *Xanthomonas, X. albilineans* exhibits distinctive pathogenic mechanisms, ecology and taxonomy. Its genome, which has experienced significant erosion, has unique genomic features. It lacks two loci required for pathogenicity in other plant pathogenic species of *Xanthomonas*: the xanthan gum biosynthesis and the Hrp-T3SS (hypersensitive response and pathogenicity-type three secretion system) gene clusters. Instead, *X. albilineans* harbors in its genome an SPI-1 (*Salmonella* pathogenicity island-1) T3SS gene cluster usually found in animal pathogens. *X. albilineans* produces a potent DNA gyrase inhibitor called albicidin, which blocks chloroplast differentiation, resulting in the characteristic white foliar stripe symptoms. The antibacterial activity of albicidin also confers on *X. albilineans* a competitive advantage against rival bacteria during sugarcane colonization. Recent chemical studies have uncovered the unique structure of albicidin and allowed us to partially elucidate its fascinating biosynthesis apparatus, which involves an enigmatic hybrid PKS/NRPS (polyketide synthase/non-ribosomal peptide synthetase) machinery.

Keywords: ***Xanthomonas albilineans*, leaf scald disease of sugarcane, genomic features, albicidin, NRPS and PKS genes**

OPEN ACCESS

Edited by:
Nicolas Denancé, Institut National de la Recherche Agronomique, France

Reviewed by:
Joao C. Setubal, University of Sao Paulo, Brazil
Julian J. Smith, Fera Science Ltd., UK

***Correspondence:**
Stéphane Cociancich, UMR BGPI, Cirad, TA A-54/K, Campus international de Baillarguet, F-34398 Montpellier Cedex 5, France
stephane.cociancich@cirad.fr

Specialty section:
This article was submitted to Plant-Microbe Interaction, a section of the journal Frontiers in Plant Science

Received: *24 February 2015*
Accepted: *09 April 2015*
Published: *24 April 2015*

Citation:
Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD and Cociancich S (2015) What makes Xanthomonas albilineans unique amongst xanthomonads? Front. Plant Sci. 6:289. doi: 10.3389/fpls.2015.00289

Introduction

Xanthomonas albilineans (Ashby) Dowson is known to invade the xylem of sugarcane and to cause leaf scald disease (Rott and Davis, 2000; Birch, 2001). Symptoms of this disease vary from a single, white, narrow, sharply defined stripe to complete wilting and necrosis of infected leaves, leading to plant death. Dissemination of *X. albilineans* occurs mainly mechanically through use of contaminated harvesting tools and by distribution and planting of infected cuttings. However, aerial transmission and potential for epiphytic survival have also been reported for this pathogen (Autrey et al., 1995; Daugrois et al., 2003; Champoiseau et al., 2009).

Xanthomonas albilineans is a representative of the genus *Xanthomonas*, members of which are exclusively Gram-negative plant-associated bacteria that collectively cause dramatic damage to hundreds of plant species of ornamental or agronomical interest. Indeed, both monocotyledonous (e.g., rice, sugarcane, or banana) and dicotyledonous (e.g., citrus, cauliflower, bean, pepper, cabbage, and tomato) plants are targeted worldwide by various *Xanthomonas* species. While sharing numerous phenotypic characteristics, at least 27 species and over 120 pathovars (variants of pathogeny) of the genus *Xanthomonas* are currently recognized. Each pathovar individually exhibits a very

restricted host range and/or tissue-specificity and this leads to clustering of bacterial strains causing similar symptoms on the same host.

Multilocus sequence analysis (MLSA) with four housekeeping genes resulted in the distribution of *Xanthomonas* species in two clades. The main one contains the majority of species whereas the secondary clade contains *X. albilineans*, *Xanthomonas sacchari*, *Xanthomonas theicola*, *Xanthomonas hyacinthi*, and *Xanthomonas translucens* (Young et al., 2008). Phylogenetic analyses with the *gyrB* sequence indicate that this secondary group also contains several uncharacterized species of *Xanthomonas* isolated mainly on rice, banana or sugarcane (Studholme et al., 2011, 2012). Intriguingly, two multiMLSA studies with 28 genes and 228 genes, respectively, in which *X. albilineans* is the only representative of this secondary clade, resulted in the branching of *Xylella fastidiosa* between *X. albilineans* and the main clade (Rodriguez-R et al., 2012; Naushad and Gupta, 2013). *X. fastidiosa* is a xylem-limited bacterium which is insect-vectored to a variety of diverse hosts, has a reduced genome and lacks the Hrp-T3SS (hypersensitive response and pathogenicity–type III secretion system; Simpson et al., 2000).

Analysis of the *X. albilineans* genome has revealed unusual features compared to other xanthomonads, the most prominent being the absence of the Hrp-T3SS gene cluster and the occurrence of genome erosion. Furthermore, to our knowledge, *X. albilineans* is the only xanthomonad that produces the phytotoxin albicidin. This mini-review aims to summarize the characteristics that, taken together, make *X. albilineans* so unique.

Genome Erosion

The genome of *X. albilineans* strain GPE PC73 has been fully sequenced and annotated. It consists of a 3,768,695-bp circular chromosome with a G+C content of 63%, and three plasmids of 31,555-bp, 27,212-bp and 24,837-bp, respectively (Pieretti et al., 2009). This genome size is much smaller than that of any other xanthomonad sequenced to date (commonly ∼5 Mb). Examination of the genome of strain GPE PC73 together with OrthoMCL comparative analyses performed with other sequenced xanthomonads highlights several genomic features that distinguish *X. albilineans* from its near relatives (Pieretti et al., 2009, 2012; Marguerettaz et al., 2011; Royer et al., 2013).

Orthologous analyses show that *X. albilineans* and *X. fastidiosa* have experienced a convergent genome reduction during their respective speciation, with a more extensive genome reduction for *X. fastidiosa* (Pieretti et al., 2009). Based on these analyses, *X. albilineans* has lost at least 592 genes that were present in the last common ancestor of the xanthomonads. Interestingly, most of these ancestral genes are conserved in the genome of *X. sacchari* strains NCPPB4393 and LMG 476 and *Xanthomonas* spp. strains NCPPB1131 and NCPPB1132, which are the sequenced strains phylogenetically closest to *X. albilineans* (Studholme et al., 2011, 2012; Pieretti et al., 2015). This indicates that genome erosion is specific to *X. albilineans*. Convergent genome erosion of *X. albilineans* and *X. fastidiosa* could be linked to a similar adaptation to a xylem-invading lifestyle in which interactions with living plant tissues are minimal (Pieretti et al., 2009). More recently, a study of the somewhat reduced genome of *Xanthomonas fragariae* (4.2 Mb) led to the hypothesis that the convergent genome reduction observed in some xanthomonads could be linked to their endophytic lifestyle and typically to their commitment to a single host (Vandroemme et al., 2013).

Compared to other xanthomonads, a low number of insertion sequences (IS) has been found in the genome of *X. albilineans*. Taken together with a limited recombination of the chromosome and a GC skew pattern containing a low number of distortions, it was postulated that genome erosion of *X. albilineans* was mainly not due to IS and other mechanisms were proposed for this erosion (Pieretti et al., 2009). The low number of IS could be linked to the activity of CRISPR (clustered regularly interspaced short palindromic repeats) systems. Strain GPE PC73 of *X. albilineans* possesses two CRISPR loci. The first one, CRISPR-1, is conserved in *X. oryzae* pv. *oryzae*, *X. axonopodis* pv. *citri*, *X. campestris* pv. *vasculorum*, and *X. campestris* pv. *musacearum*. The second, CRISPR-2, is present in *X. campestris* pv. *raphani* (Pieretti et al., 2012). Interestingly, many spacers of CRISPR-1 and CRISPR-2 of strain GPE PC73 are identical to IS or phage-related DNA sequences present on the chromosome of this strain (Pieretti et al., 2012).

Specific Genes Linked to a Xylem-Invading Lifestyle

Although determinants for host- or tissue-specificity of *X. albilineans* remain unclear, the presence in its genome of genes encoding cell-wall-degrading enzymes (CWDEs) with specific features is probably important for its ability to spread in xylem and for pathogenicity. Indeed, all CWDEs from *X. albilineans* harbor a cellulose-binding domain (CBD) and a long linker region both adapted to the utilization of cell-wall breakdown products as carbon source and to the ability to spread in sugarcane xylem vessels (Pieretti et al., 2012). These enzymes may also be required to disrupt pit membranes in sugarcane, thereby promoting propagation of the bacteria in the plant. Interestingly, *X. fastidiosa* also encodes two CWDEs containing a long linker and a CBD. It has been shown that one of these two CWDEs is involved in the spread of *X. fastidiosa* in the xylem by increasing the pore size of pit membranes. CWDEs are therefore considered as virulence factors (Roper et al., 2007; Chatterjee et al., 2008; Pérez-Donoso et al., 2010). TonB-dependent transporters (TBDTs) may be used by *X. albilineans* to transport cell-wall-degrading products resulting from the activity of CWDEs, and thus may facilitate spread of the organism in the nutrient-poor conditions prevailing in the xylem of sugarcane. In the genome of *X. albilineans*, 35 TBDT genes have been identified, including one specific to this species and two others that are functionally associated to pathogenicity of the bacterium (Rott et al., 2011; Pieretti et al., 2012).

Lack of Hrp-T3SS

Most phytopathogenic bacteria rely on the type III secretion system (T3SS) of the hypersensitive response and pathogenicity family (Hrp1 and Hrp2, respectively). This syringe-like apparatus allows pathogens to deliver, into their host cells, proteins (type

III effectors) that modulate plant physiology and immunity for the benefit of the pathogen. Interestingly, genes encoding the injectisome and associated effectors of the Hrp-T3SS are missing in the genome of *X. albilineans*, as is also the case in the genomes of *X. sacchari* strains NCPPB4393 and LMG 476 and *Xanthomonas* spp. strains NCPPB1131 and NCPPB1132 (Studholme et al., 2011, 2012; Pieretti et al., 2015). Yet, an Hrp system is present in other close neighbor species of *X. albilineans*, such as *X. translucens* pv. *graminis* strain 29, *X. translucens* pv. *translucens* strain DSM18974, and *X. translucens* strain DAR 61454 (Wichmann et al., 2013; Gardiner et al., 2014). Although the Hrp-T3SS is described as a crucial key component in plant–host interactions for most *Xanthomonas* spp, it seems not to be essential in *X. translucens* pv. *graminis* strain 29 for xylem colonization, even though it is involved in symptom development (Ryan et al., 2011; Wichmann et al., 2013). Similarly, despite being devoid of any Hrp T3SS, *X. albilineans* displays pathogenicity and is able to cause serious damage to sugarcane.

Acquisition of a SPI-1 T3SS

The annotated sequence of the genome of *X. albilineans* strain GPE PC73 reveals the presence of a T3SS belonging to the *Salmonella* pathogenicity island-1 (SPI-1) injectisome family. Genes encoding this system are located near the terminus of the replication site of the chromosome and were probably acquired by lateral gene transfer. This secretion system, found mainly in mammals and insects bacterial pathogens or symbionts, exhibits high similarity to that described in *Burkholderia pseudomallei*—a human pathogen causing melioidosis (Stevens et al., 2002). The SPI-1 needle-like assemblies of *X. albilineans* strain GPE PC73 and *B. pseudomallei* strain K96243 are homologous. Both species share all but two genes—*orgA* and *orgB*, encoding putative oxygen-regulated invasion proteins involved in type three secretion that are not conserved in *B. pseudomallei*. The genome composition of the SPI-1 T3SS in *X. albilineans* additionally includes genes encoding translocon components (*xipB*, *xipC*, and *xipD*), injectisome components (*xsaJ* to *xsaS* and *xsaV* to *xsaZ*) and a chaperone (*xicA*). Furthermore, the locus contains 15 additional genes referred to as *xapA*–*xapO*, encoding hypothetical proteins. These genes, which show homology neither to sequences from *B. pseudomallei* nor to sequences available from protein sequence databases, are specific to *X. albilineans* and their products represent good candidates to be considered as effectors for this SPI-1 T3SS (Marguerettaz et al., 2011). Interestingly, this SPI-1 T3SS is conserved in *Xanthomonas axonopodis* pv. *phaseoli* strains CFBP 2534, CFBP 6164 and CFBP 6982, which moreover possess a second T3SS belonging to the Hrp2 family (Alavi et al., 2008; Marguerettaz et al., 2011). Pathogenicity of *X. albilineans* strains seems not to be linked to the presence of the SPI-1 T3SS in their genome; besides, no SPI-1 T3SS locus has been identified in strain PNG130 of *X. albilineans* even though it is able to spread in sugarcane. Functional analyses showed that, *in planta*, multiplication of a SPI-1 T3SS knockout mutant of *X. albilineans* was not impaired when compared to the wild-type, indicating that the SPI-1 T3SS is not required for spread in sugarcane vessels or for development of leaf scald symptoms. The role of the SPI-1 T3SS of *X. albilineans* remains unclear, although it has been conserved during its evolution in *X. albilineans* without frame-shifting indels or nonsense mutations (Marguerettaz et al., 2011). It remains possible, in conditions other than those tested with our knockout mutant, that the SPI-1 T3SS system may be required for interaction with sugarcane, as in the case of SPI-1 of *Salmonella*, which is involved in interactions with *Arabidopsis thaliana* (Schikora et al., 2011). The SPI-1 T3SS system may also be associated with other aspects of the *X. albilineans* lifestyle, e.g., an involvement in adherence as reported for *Erwinia tasmaniensis* (Kube et al., 2008) or in formation of pellicle or biofilm-like structures (Jennings et al., 2012), which could be related to epiphytic survival on sugarcane leaves. Although no insect vector has been identified for *X. albilineans* to date, we cannot rule out that the SPI-1 T3SS could be involved in insect association or might mediate persistence of the bacterium in an insect vector as was shown for *Pantoea stewartii* (Correa et al., 2012).

Lack of T6SS and the Xanthan Gum Gene Cluster

Xanthomonas albilineans lacks two other major pathogenicity factors that are common features of most xanthomonads. First, it lacks the gum gene cluster for extracellular polysaccharide (EPS) synthesis. This gene cluster is responsible for biofilm and xanthan gum formation, and is associated with pathogenesis in xanthomonads (Katzen et al., 1998; Kim et al., 2009; Galván et al., 2012). Exceptions are *X. fragariae*, which lacks the *gumN*, *gumO* and *gumP* genes, and *X. albilineans*, which lacks the complete set of gum genes, indicating those are not essential for virulence of both these pathogens (Pieretti et al., 2012; Vandroemme et al., 2013).

Xanthomonas albilineans is also devoid of any type VI secretion system (T6SS) described in other xanthomonads, as for example in *Xanthomonas fuscans* pv. *fuscans* strain 4834-R and *Xanthomonas citri* subsp. *citri* strain 306, which each contain a single T6SS (Potnis et al., 2011; Darrasse et al., 2013) or *X. translucens* strain DAR61454, which encodes two distinct T6SS (Gardiner et al., 2014). Structurally, the T6SS looks like an inverted bacteriophage. Functionally, this system is able to interact with both eukaryotic and prokaryotic cells by delivering effectors or toxins into host cells to subvert the signaling process to its own advantage, but also into other bacteria from the same habitat to outcompete them during infection (Filloux, 2013; Russell et al., 2014). Despite its multifunctional roles during host–pathogen interactions, the lack of T6SS in *Xanthomonas campestris* pv. *campestris* strain 8004, *Xanthomonas gardneri* strain 101, and *X. albilineans* seems to have no effect on pathogenesis of these xanthomonads.

Albicidin and Other Non-Ribosomally Synthesized Peptides

A unique feature of *X. albilineans* is the production of albicidin—a phytotoxin causing the white foliar stripe symptoms

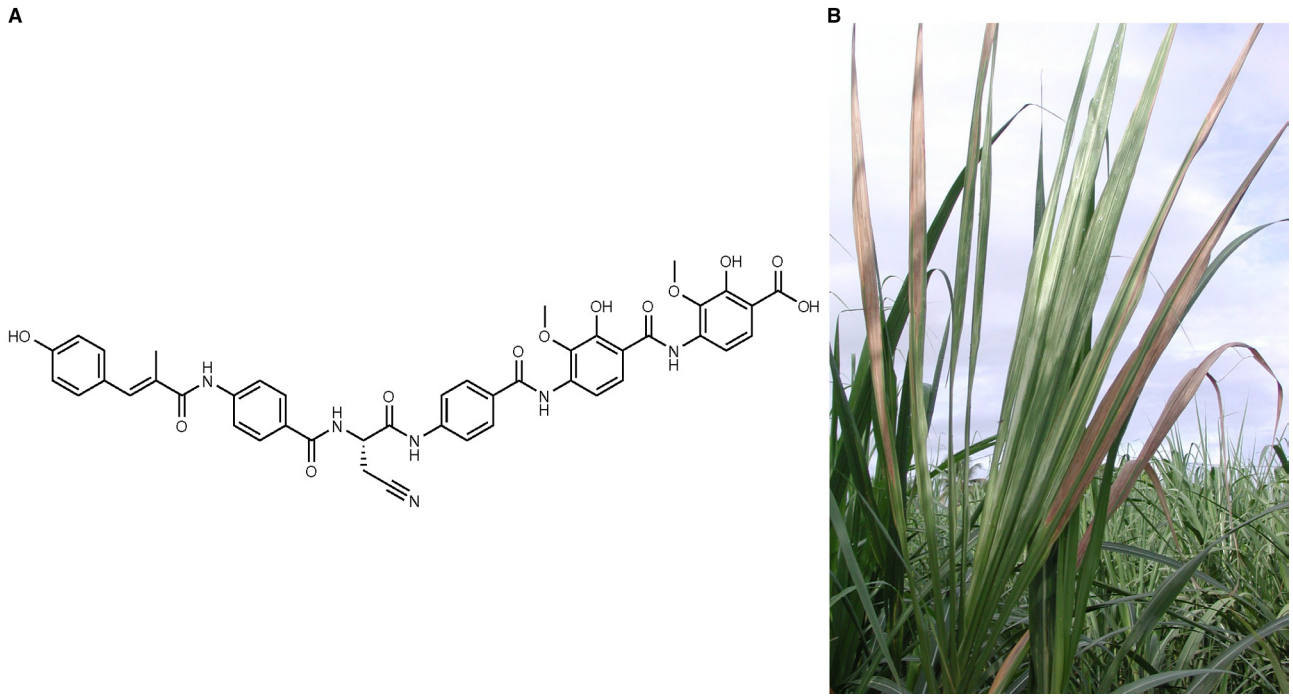


FIGURE 1 | *Xanthomonas albilineans* produces the phytotoxin albicidin—a potent gyrase inhibitor that blocks chloroplast differentiation, resulting in sugarcane leaf scald disease symptoms. (A) Structure of albicidin, a hybrid PKS/NRPS compound with unique composition including *p*-aminobenzoic acid and cyanoalanine. **(B)** Diseased sugarcane plant with characteristic leaf scald symptoms: white foliar bleaching and necrosis of infected leaves (© J. H. Daugrois/Cirad).

characteristic of leaf scald disease of sugarcane (Birch and Patil, 1985). Albicidin is a potent DNA gyrase inhibitor that blocks the differentiation of chloroplasts (**Figure 1**). It also targets bacterial gyrase by a mechanism different from that of other DNA gyrase inhibitors like coumarins and quinolones (Hashimi et al., 2007). This mode of action accounts for the potent antibacterial activity of albicidin, which inhibits the growth of Gram-positive and Gram-negative pathogenic bacteria at nanomolar concentrations (Birch and Patil, 1985). Albicidin gives a competitive advantage to *X. albilineans* against other bacteria within the xylem vessels of sugarcane (Magnani et al., 2013). Interestingly, two sugarcane-living bacteria harbor an albicidin resistance gene: *Leifsonia xyli* (Monteiro-Vitorello et al., 2004) and *Pantoea dispersa* (Zhang and Birch, 1997).

Albicidin is produced by a hybrid polyketide synthase (PKS)/non-ribosomal peptide synthetase (NRPS) enzyme complex. PKS and NRPS genes are often clustered together with a large set of regulatory, transport or modification (tailoring) genes, as well as genes involved in the biosynthesis of non-proteinogenic amino acids. In addition to a phosphopantetheinyl transferase required for activation of the PKS/NRPS system and a HtpG chaperone, the role of which remains unclear, a locus (*alb* cluster) containing 20 genes is required for albicidin biosynthesis. Among these 20 genes, 3 encode the PKS/NRPS system; 15 others act as transport, regulatory, modification or resistance genes (Royer et al., 2004).

Non-ribosomal peptide synthetases are multimodular mega-synthetases used by bacteria and fungi to produce peptides in a ribosome-independent manner (Strieker et al., 2010). Each module governs the specific incorporation of an amino acid substrate based on signature sequences in the adenylation (A) domains (Stachelhaus and Marahiel, 1995), which are loaded onto peptidyl carrier protein (PCP) domains. Elongation of the peptide is mediated by condensation (C) domains present within each module. PKSs function according to the principles of fatty acid biosynthesis (Weissman and Leadlay, 2005).

For decades, the structure elucidation of albicidin was impeded by its extremely low production yield by *X. albilineans*. A first step to overcome this bottleneck was achieved by transferring the biosynthetic genes into a heterologous host, namely *X. axonopodis* pv. *vesicatoria*, resulting in a significant increase in albicidin production (Vivien et al., 2007). Extensive HPLC purification of albicidin and thorough analysis of the purified compound by means of mass spectrometry and nuclear magnetic resonance spectroscopy then allowed us to unravel its unique structure (**Figure 1**). Albicidin proved to be a linear pentapeptide composed of cyanoalanine and *p*-amino benzoic acids N-terminally linked to a *p*-coumaric acid derivative (Cociancich et al., 2015). Although over 500 different monomers (amino acid substrates) have been identified to date as being incorporated by NRPS systems, elucidation of the structure of albicidin revealed for the first time the incorporation by NRPSs of cyanoalanine and *p*-amino benzoic acids. Moreover, the incorporation of *p*-amino benzoic acids is the first example of incorporation of a δ-aminoacid by NRPSs, since all NRPSs described to date incorporate only α or β aminoacids. The use of unusual amino acid substrates is linked to unique

features that were identified *in silico* 10 years ago within the albicidin NRPS modules sequence (Royer et al., 2004). The formation and incorporation of cyanoalanine most likely occurs *in situ* through an additional module present in the PKS-NRPS assembly line that was investigated in one of our present studies (Cociancich et al., 2015).

Chemical synthesis of albicidin is now available, allowing both production of high quantities of the compound for further study of its mode of action and activity spectrum, and the synthesis of analogs (Kretz et al., 2015). The uniqueness of its structure and the specific mode of action of this compound make albicidin a strong lead structure for antibiotic development.

Data mining of the genome of *X. albilineans* strain GPE PC73 has led to the identification, in addition to the albicidin biosynthesis locus, of five other NRPS loci (Pieretti et al., 2012; Royer et al., 2013). The first, named Meta-B, encodes megasynthases performing peptidic elongation of a 16-amino acid lipopeptide. This locus also encodes a transcription regulator belonging to the AraC family, a cyclic peptide transporter, and enzymes involved in biosynthesis of the non-proteinogenic amino acids di-amino butyric acid and dihydroxyphenylglycine. Interestingly, the NRPS locus Meta-B has been identified in the genome of strains of three other *Xanthomonas* species, namely *Xanthomonas oryzae* pv. *oryzae* strains BAI3 and X11-5A, *X. translucens* strain DAR61454 and *Xanthomonas* spp. strain XaS3 (Royer et al., 2013). Despite a similar organization of the genes within these loci, the *in silico* prediction of the sequences of the peptides produced indicates that each strain produces a different lipopeptide.

Two other NRPS gene clusters, Meta-A and Meta-C, have been identified in the genome of *X. albilineans* strain GPE PC73. They encode megasynthases that perform the biosynthesis of peptides of 12 and 7 amino acids, respectively. A partial sequence has been predicted for each of these peptides (Royer et al., 2013).

Finally, two short NRPS genes have also been identified on the chromosome of *X. albilineans*: they both encode only one NRPS module. Interestingly, there is an overlap between both these genes and a gene encoding a glycosyltransferase. It has been hypothesized that these genes encode glycosylated amino acids, to which, however, no precise function could yet be attributed (Royer et al., 2013).

Conclusion

Although most xanthomonads require pathogenicity factors such as *gum* genes, T3SS Hrp and T6SS for survival, growth and spread within host plants, *X. albilineans* lacks these pathogenicity factors, *de facto* reducing its artillery to circumvent sugarcane defense mechanisms and innate immunity. While being "disarmed" could be disadvantageous for a vascular plant pathogen, *X. albilineans* remains able to invade and spread in sugarcane, suggesting that it uses other strategies, such as stealth, i.e., being unobtrusive *in planta*, to minimize inducible host defense responses. On the other hand, the reduced genome of *X. albilineans* has specific features that may be involved in the adaptation of the bacterium to live and spread in sugarcane xylem vessels. For example, specific CWDEs and TBDTs appear to be optimized for life in the nutrient-poor sugarcane xylem environment. The uniqueness of *X. albilineans* resides also in the production of the phytotoxin and antibiotic albicidin. The recently unraveled structure and concomitant development of a chemical synthesis protocol for this compound leads to additional prospects for its use in the antibiotherapy field. According to the specificities deriving from the biological, biochemical, phylogenetic and genomic analyses described in this review, one can truly say that *X. albilineans* is quite unique amongst the genus *Xanthomonas*.

Acknowledgments

Work on albicidin was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG SU239/11-1; SU 18-1), by the Cluster of Excellence "Unifying Concepts in Catalysis (UniCat)" (DFG) and by a grant from the Agence Nationale de la Recherche (ANR-09-BLAN-0413-01). The authors are indebted to Helen Rothnie for English editing.

References

Alavi, S. M., Sanjari, S., Durand, F., Brin, C., Manceau, C., and Poussier, S. (2008). Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. phaseoli and *Xanthomonas fuscans* subsp. fuscans as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive hybridizations. *Appl. Environ. Microbiol.* 74, 3295–3301. doi: 10.1128/aem.02507-07

Autrey, L., Saumtally, S., Dookun, A., Sullivan, S., and Dhayan, S. (1995). Aerial transmission of the leaf scald pathogen, *Xanthomonas albilineans*. *Proc. Int. Soc. Sug. Cane Technol.* 21, 508–526.

Birch, R. G. (2001). *Xanthomonas albilineans* and the antipathogenesis approach to disease control. *Mol. Plant Pathol.* 2, 1–11. doi: 10.1046/j.1364-3703.2001.00046.x

Birch, R. G., and Patil, S. S. (1985). Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J. Gen. Microbiol.* 131, 1069–1075. doi: 10.1099/00221287-131-5-1069

Champoiseau, P., Rott, P., and Daugrois, J. H. (2009). Epiphytic populations of *Xanthomonas albilineans* and subsequent sugarcane stalk infection are linked to rainfall in Guadeloupe. *Plant Dis.* 93, 339–346. doi: 10.1094/pdis-93-4-0339

Chatterjee, S., Almeida, R. P. P., and Lindow, S. (2008). Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annu. Rev. Phytopathol.* 46, 243–271. doi: 10.1146/annurev.phyto.45.062806.094342

Cociancich, S., Pesic, A., Petras, D., Uhlmann, S., Kretz, J., Schubert, V., et al. (2015). The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* 11, 195–197. doi: 10.1038/nchembio.1734

Correa, V. R., Majerczak, D. R., Ammar, E.-D., Merighi, M., Pratt, R. C., Hogenhout, S. A., et al. (2012). The bacterium *Pantoea stewartii* uses two different type III secretion systems to colonize its plant host and insect vector. *Appl. Environ. Microbiol.* 78, 6327–6336. doi: 10.1128/aem.00892-12

Darrasse, A., Carrere, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genom.* 14:761. doi: 10.1186/1471-2164-14-761

Daugrois, J. H., Dumont, V., Champoiseau, P., Costet, L., Boisne-Noc, R., and Rott, P. (2003). Aerial contamination of sugarcane in Guadeloupe by two strains of *Xanthomonas albilineans*. *Eur. J. Plant Pathol.* 109, 445–458. doi: 10.1023/a:1024259606468

Filloux, A. (2013). The rise of the type VI secretion system. *F1000Prime Rep*. 5:52. doi: 10.12703/p5-52

Galván, E. M., Ielmini, M. V., Patel, Y. N., Bianco, M. I., Franceschini, E. A., Schneider, J. C., et al. (2012). Xanthan chain length is modulated by increasing the availability of the polysaccharide copolymerase protein GumC and the outer membrane polysaccharide export protein GumB. *Glycobiology* 23, 259–272. doi: 10.1093/glycob/cws146

Gardiner, D. M., Upadhyaya, N. M., Stiller, J., Ellis, J. G., Dodds, P. N., Kazan, K., et al. (2014). Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS ONE* 9:e84995. doi: 10.1371/journal.pone.0084995

Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A., and Birch, R. G. (2007). The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* 51, 181–187. doi: 10.1128/aac.00918-06

Jennings, M. E., Quick, L. N., Ubol, N., Shrom, S., Dollahon, N., and Wilson, J. W. (2012). Characterization of *Salmonella* type III secretion hyper-activity which results in biofilm-like cell aggregation. *PLoS ONE* 7:e33080. doi: 10.1371/journal.pone.0033080

Katzen, F., Ferreiro, D. U., Oddo, C. G., Ielmini, M. V., Becker, A., Pühler, A., et al. (1998). *Xanthomonas campestris* pv. *campestris gum* mutants: effects on xanthan biosynthesis and plant virulence. *J. Bacteriol.* 180, 1607–1617.

Kim, J.-G., Li, X., Roden, J. A., Taylor, K. W., Aakre, C. D., Su, B., et al. (2009). *Xanthomonas* T3S effector XopN suppresses PAMP-triggered immunity and interacts with a tomato atypical receptor-like kinase and TFT1. *Plant Cell* 21, 1305–1323. doi: 10.1105/tpc.108.063123

Kretz, J., Kerwat, D., Schubert, V., Grätz, S., Pesic, A., Semsary, S., et al. (2015). Total synthesis of albicidin: a lead structure from *Xanthomonas albilineans* for potent antibacterial gyrase inhibitors. *Angew. Chem. Int. Ed.* 54, 1969–1973. doi: 10.1002/anie.201409584

Kube, M., Migdoll, A. M., Müller, I., Kuhl, H., Beck, A., Reinhardt, R., et al. (2008). The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. *Environ. Microbiol.* 10, 2211–2222. doi: 10.1111/j.1462-2920.2008.01639.x

Magnani, G. S., Cruz, L. M., Weber, H., Bespalhok, J. C., Daros, E., Baura, V., et al. (2013). Culture-independent analysis of endophytic bacterial communities associated with Brazilian sugarcane. *Genet. Mol. Res.* 12, 4549–4558. doi: 10.4238/2013.October.15.3

Marguerettaz, M., Pieretti, I., Gayral, P., Puig, J., Brin, C., Cociancich, S., et al. (2011). Genomic and evolutionary features of the SPI-1 type III secretion system that is present in *Xanthomonas albilineans* but is not essential for xylem colonization and symptom development of sugarcane leaf scald. *Mol. Plant Microbe Interact.* 24, 246–259. doi: 10.1094/mpmi-08-10-0188

Monteiro-Vitorello, C. B., Camargo, L. A., Van Sluys, M. A., Kitajima, J. P., Truffi, D., do Amaral, A. M., et al. (2004). The genome sequence of the gram-positive sugarcane pathogen *Leifsonia xyli* subsp. xyli. *Mol. Plant Microbe Interact.* 17, 827–836. doi: 10.1094/mpmi.2004.17.8.827

Naushad, H. S., and Gupta, R. S. (2013). Phylogenomics and molecular signatures for species from the plant pathogen-containing order Xanthomonadales. *PLoS ONE* 8:e55216. doi: 10.1371/journal.pone.0055216

Pérez-Donoso, A. G., Sun, Q., Roper, M. C., Greve, L. C., Kirkpatrick, B., and Labavitch, J. M. (2010). Cell wall-degrading enzymes enlarge the pore size of intervessel pit membranes in healthy and *Xylella fastidiosa*-infected grapevines. *Plant Physiol.* 152, 1748–1759. doi: 10.1104/pp.109.148791

Pieretti, I., Bolot, S., Carrère, S., Barbe, V., Cociancich, S., Rott, P., et al. (2015). Draft genome sequence of *Xanthomonas sacchari* LMG 476. *Genome Announc.* 3, e00146-15. doi: 10.1128/genomeA.00146-15

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., et al. (2009). The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genom.* 10:616. doi: 10.1186/1471-2164-10-616

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Couloux, A., et al. (2012). Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genom.* 13:658. doi: 10.1186/1471-2164-13-658

Potnis, N., Krasileva, K., Chow, V., Almeida, N., Patil, P., Ryan, R., et al. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genom.* 12:146. doi: 10.1186/1471-2164-12-146

Rodriguez-R, L., Grajales, A., Arrieta-Ortiz, M., Salazar, C., Restrepo, S., and Bernal, A. (2012). Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol.* 12:43. doi: 10.1186/1471-2180-12-43

Roper, M. C., Greve, L. C., Warren, J. G., Labavitch, J. M., and Kirkpatrick, B. C. (2007). *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. *Mol. Plant Microbe Interact.* 20, 411–419. doi: 10.1094/mpmi-20-4-0411

Rott, P., and Davis, M. (2000). "Leaf scald," in *A Guide to Sugarcane Diseases*, eds P. Rott, R. Bailey, J. Comstock, B. Croft, and A. Saumtally (Montpellier: CIRAD-ISSCT press), 339.

Rott, P., Fleites, L., Marlow, G., Royer, M., and Gabriel, D. W. (2011). Identification of new candidate pathogenicity factors in the xylem-invading pathogen *Xanthomonas albilineans* by transposon mutagenesis. *Mol. Plant Microbe Interact.* 24, 594–605. doi: 10.1094/mpmi-07-10-0156

Royer, M., Costet, L., Vivien, E., Bes, M., Cousin, A., Damais, A., et al. (2004). Albicidin pathotoxin produced by *Xanthomonas albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. *Mol. Plant Microbe Interact.* 17, 414–427. doi: 10.1094/mpmi.2004.17.4.414

Royer, M., Koebnik, R., Marguerettaz, M., Barbe, V., Robin, G., Brin, C., et al. (2013). Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genom.* 14:658. doi: 10.1186/1471-2164-14-658

Russell, A. B., Peterson, S. B., and Mougous, J. D. (2014). Type VI secretion effectors: poisons with a purpose. *Nat. Rev. Microbiol.* 12, 137–148. doi: 10.1038/nrmicro3185

Ryan, R. P., Vorhölter, F.-J., Potnis, N., Jones, J. B., Van Sluys, M.-A., Bogdanove, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* 9, 344–355. doi: 10.1038/nrmicro2558

Schikora, A., Virlogeux-Payant, I., Bueso, E., Garcia, A. V., Nilau, T., Charrier, A., et al. (2011). Conservation of *Salmonella* infection mechanisms in plants and animals. *PLoS ONE* 6:e24112. doi: 10.1371/journal.pone.0024112

Simpson, A., Reinach, F., and Arruda, P. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406, 151–157. doi: 10.1038/35018003

Stachelhaus, T., and Marahiel, M. A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* 125, 3–14. doi: 10.1111/j.1574-6968.1995.tb07328.x

Stevens, M. P., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., et al. (2002). An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Mol. Microbiol.* 46, 649–659. doi: 10.1046/j.1365-2958.2002.03190.x

Strieker, M., Tanovic, A., and Marahiel, M. A. (2010). Non-ribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* 20, 234–240. doi: 10.1016/j.sbi.2010.01.009

Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., et al. (2011). Draft genome sequences of *Xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* Group 1 Clade. *Genes* 2, 1050–1065. doi: 10.3390/genes2041050

Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., et al. (2012). Correction: Studholme et al., Draft genome sequences of *Xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* Group 1 Clade. Genes 2011, 2, 1050–1065. *Genes* 3, 88–89. doi: 10.3390/genes3010088

Vandroemme, J., Cottyn, B., Baeyen, S., De Vos, P., and Maes, M. (2013). Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genom.* 14:829. doi: 10.1186/1471-2164-14-829

Vivien, E., Pitorre, D., Cociancich, S., Pieretti, I., Gabriel, D. W., Rott, P. C., et al. (2007). Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* 51, 1549–1552. doi: 10.1128/aac.01450-06

Weissman, K. J., and Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Micro.* 3, 925–936. doi: 10.1038/nrmicro1287

Wichmann, F., Vorhölter, F.-J., Hersemann, L., Widmer, F., Blom, J., Niehaus, K., et al. (2013). The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol. Plant Pathol.* 14, 576–588. doi: 10.1111/mpp.12030

Young, J. M., Park, D. C., Shearman, H. M., and Fargier, E. (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004

Zhang, L., and Birch, R. (1997). The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9984–9989. doi: 10.1073/pnas.94.18.9984

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACS chemical biology

Letters

pubs.acs.org/acschemicalbiology

The *O*-Carbamoyl-Transferase Alb15 Is Responsible for the Modification of Albicidin

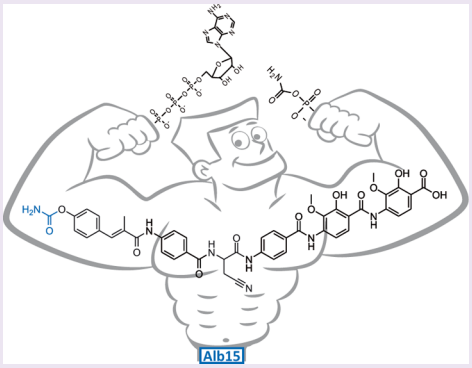
Daniel Petras,[†] Dennis Kerwat,[†] Alexander Pesic,[†] Benjamin-F Hempel,[†] Leonard von Eckardstein,[†] Siamak Semsary,[†] Julie Arasté,[‡] Mélanie Marguerettaz,[‡] Monique Royer,[‡] Stéphane Cociancich,[‡] and Roderich D. Süssmuth*[,†]

[†]Institut für Chemie, Technische Universität Berlin, Straße des 17. Juni 124, 10623 Berlin, Germany

[‡]Cirad, UMR BGPI, F-34398 Montpellier Cedex 5, France

Supporting Information

ABSTRACT: Albicidin is a potent antibiotic and phytotoxin produced by *Xanthomonas albilineans* which targets the plant and bacterial DNA gyrase. We now report on a new albicidin derivative which is carbamoylated at the *N*-terminal coumaric acid by the action of the ATP-dependent *O*-carbamoyl-transferase Alb15, present in the albicidin (*alb*) gene cluster. Carbamoyl-albicidin was characterized by tandem mass spectrometry from cultures of a *Xanthomonas* overproducer strain and the gene function confirmed by gene inactivation of *alb15* in *X. albilineans*. Expression of *alb15* in *Escherichia coli* and *in vitro* reconstitution of the carbamoyltransferase activity confirmed albicidin as the substrate. The chemical synthesis of carbamoyl-albicidin finally enabled us to assess its bioactivity by means of *in vitro* gyrase inhibition and antibacterial assays. Compared to albicidin, carbamoyl-albicidin showed a significantly higher inhibitory efficiency against bacterial gyrase (~8 vs 49 nM), which identifies the carbamoyl group as an important structural feature of albicidin maturation.



Sugarcane leaf scald, a lethal disease of sugarcane, is caused by the xylem-invading bacterium *Xanthomonas albilineans*,[1] which belongs to the class of *γ-Proteobacteria*. Transmission of the pathogen usually occurs through human practices during the reproduction of the sugarcane plants by preparing cuttings of already infected plants or by using contaminated cutting tools.[2] However, potential aerial transmission has been reported as well.[3,4] The plant colonization occurs through the xylem vessels of the plants where *X. albilineans* induces chlorosis, which becomes apparent by characteristic white pencil-like lines along the leaves. As it has been shown recently, in more advanced stages of sugarcane leaf scald, *X. albilineans* also spreads into the surrounding tissue of the xylem tubes.[5] A characteristic feature of *X. albilineans* is its ability to synthesize the phytotoxin albicidin, which is encoded in the *alb* gene cluster, a polyketide synthase-nonribosomal-peptide synthetase hybrid (PKS-NRPS). Albicidin, which was first reported decades ago, is a small molecule showing a high inhibitory activity against plant and bacterial DNA-gyrase[6−8] and is one of the pathogenic factors of sugarcane leaf scald.[9] Because albicidin is only produced in minute amounts (~10 μg/L) in cultures of *X. albilineans*, its structure remained elusive for more than 30 years. The high antibacterial activity against a wide range of Gram-positive and Gram-negative strains[10] combined with a new inhibition mechanism of gyrase makes albicidin a potential candidate for the development of new anti-infectives.[11] Unlike fluoroquinolones and coumarins, which target the ATP-binding pocket,[12] albicidin stabilizes the covalent DNA-gyrase complex.[11]

As new antibacterial drugs are urgently needed for the treatment of arising antibiotic resistant strains, we put a significant effort into the increase of albicidin production by using a heterologous *Xanthomonas* strain,[13] which finally enabled us to solve the structure of albicidin by a combination of extensive MS/MS and NMR experiments.[7] Having already established the structure of albicidin, we performed its total synthesis, aiming at structure activity relation (SAR) studies to assess albicidin derivatives for drug development purposes.[14] On the other hand, the structure set the basis for our biochemical investigations on the biosynthesis of albicidin in which we introduced para-amino benzoic acids (pABA) as a new class of substrates activated by nonribosomal peptide synthetases (NRPS).[7] Next to pABA as a characteristic structural feature, albicidin (Figure 1A) has at the *N*-terminus a coumaric acid derivative and L-cyanoalanine as the central amino acid, which is most likely produced at an insertional module of a *trans*-acting NRPS domain based on L-asparagine as a precursor.[7] By the combination of *in silico* analysis of the proteins present in the albicidin cluster and *in vitro* testing of the adenylation domain (A-domain) substrate specificity, we

Received: December 5, 2015
Accepted: February 17, 2016
Published: February 17, 2016

DOI: 10.1021/acschembio.5b01001
ACS Chem. Biol. 2016, 11, 1198−1204

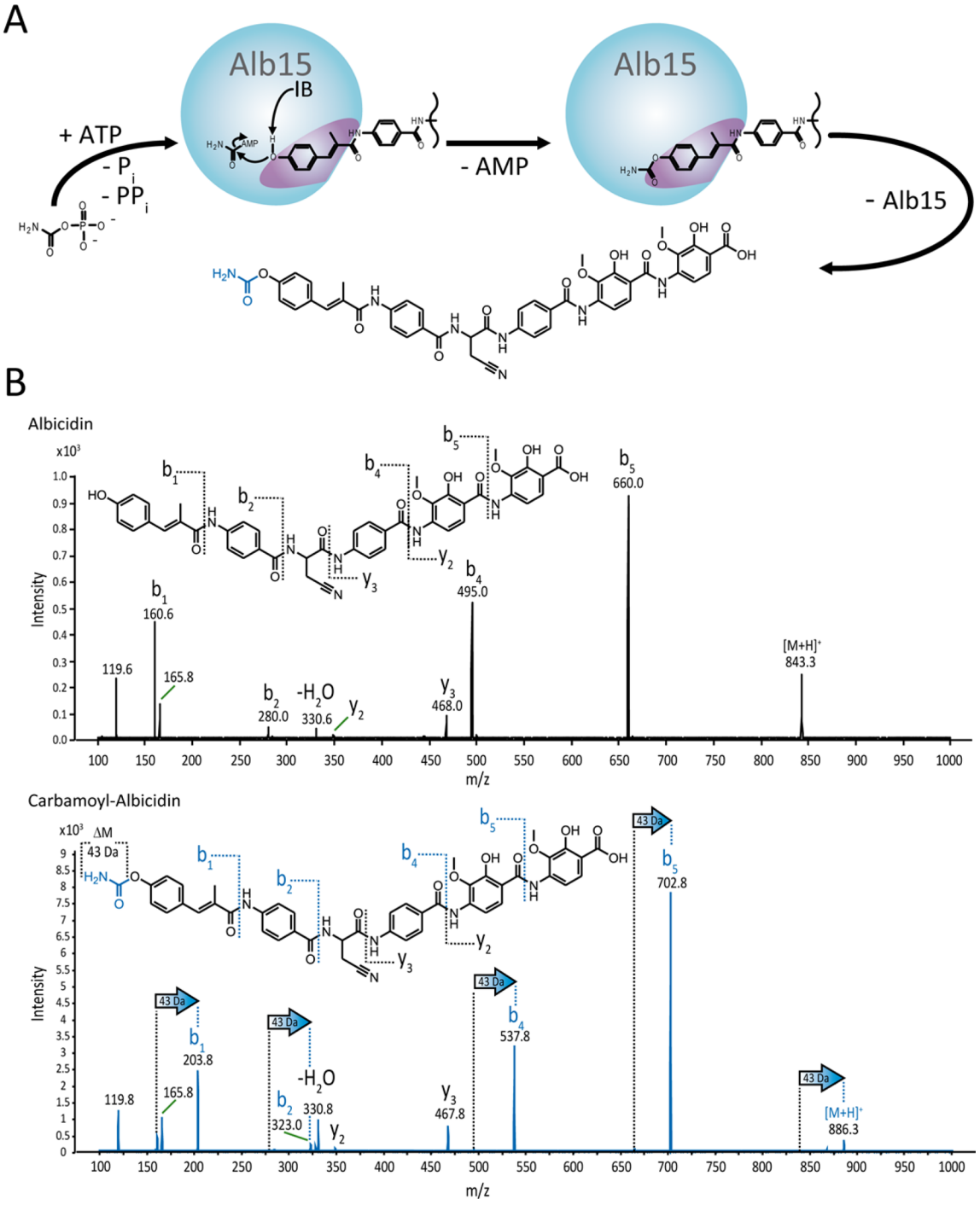


Figure 1. (A) Reaction scheme of post-NRPS carbamoylation of albicidin. As an ATP-dependent *O*-carbamoyltransferase, Alb15 catalyzes first the reaction from carbamoyl-phosphate and ATP to carbamoyl-AMP followed by the carbamoylation of albicidin. (B) MS/MS structural elucidation and comparison of albicidin and carbamoyl-albicidin. The observed b-ion and y-ion fragments and the mass shift through carbamoylation are indicated in the spectra and structure.

proposed a comprehensive biosynthesis model for albicidin.[7] Nevertheless, some gene functions present in the cluster remained unclear in terms of PKS-NRPS assembly as well as the tailoring reactions involved in post-NRPS processing of albicidin.

A putative *O*-carbamoyl transferase gene is part of the albicidin biosynthesis gene cluster (*alb* cluster),[6,7] which finds its precedence in other natural products, e.g., cephalomycin,[15] novobiocin,[16] and tobramycin.[17] Here, we show the structural elucidation and biosynthesis investigation of the putative carbamoylated albicidin derivative through a combination of tandem mass spectrometry, *in vitro*, and gene knockout experiments. Finally, we confirmed the structure by the total synthesis of carbamoyl-albicidin, which furthermore enabled us to study the impact on bioactivity and pharmacological properties, which are part of our ongoing synthetic optimization of the albicidin structure as a new antibacterial drug.

During the purification of natural albicidin by preparative HPLC and LC-MS analysis of the chromatographic fractions,

DOI: 10.1021/acschembio.5b01001
ACS Chem. Biol. 2016, 11, 1198−1204

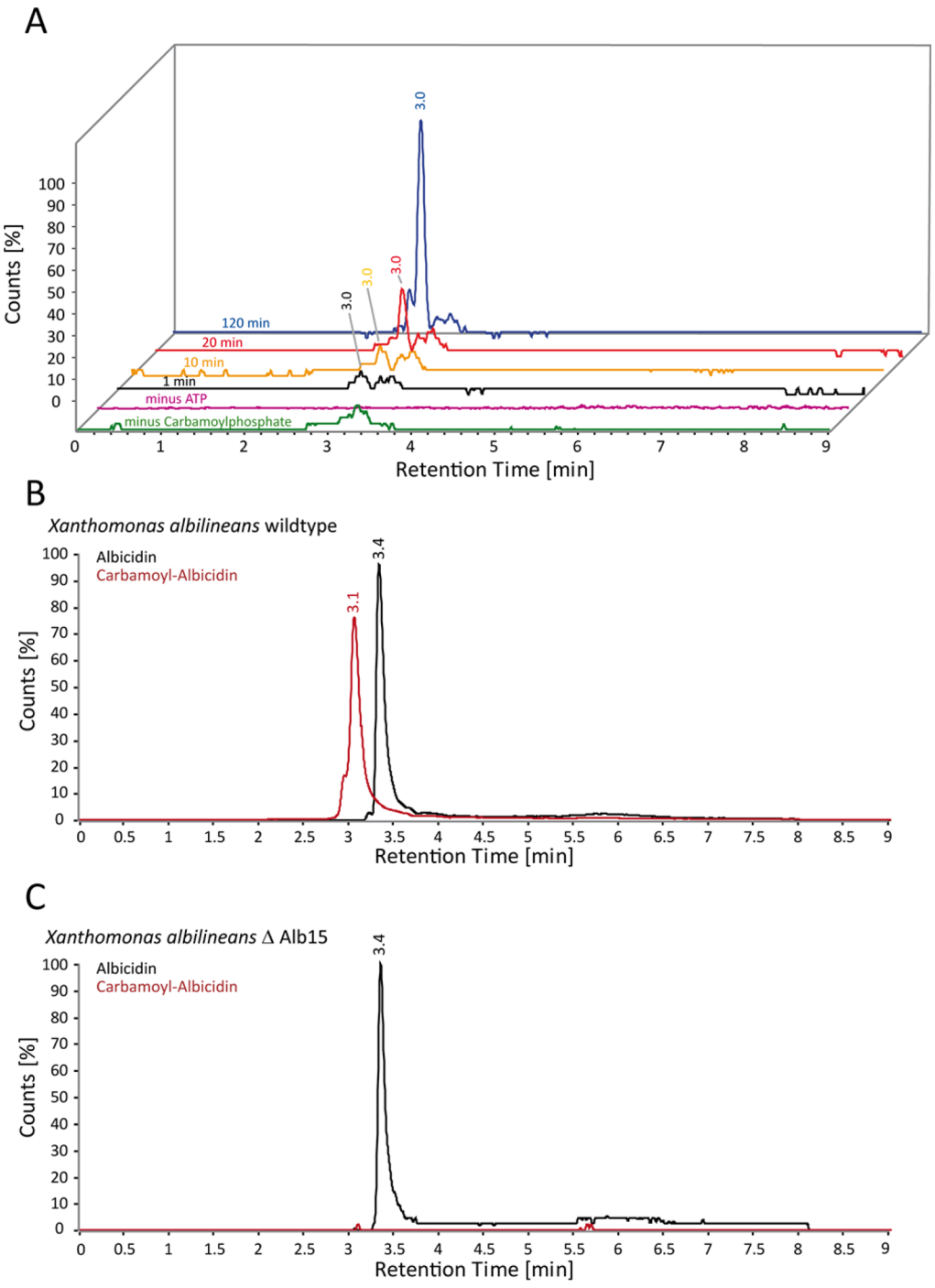


Figure 2. (A) *In vitro* carbamoylation of albicidin. Multiple reaction monitoring (MRM) chromatograms, of carbamoyl-albicidin (886 $m/z \rightarrow$ 703 m/z), at different time points after incubation with Alb15, ATP, and carbamoyl-phosphate and negative controls of the *in vitro* conversion assay are shown. (B) MRM chromatogram of XAD extract of *X. albilineans* wildtype. (C) MRM chromatogram of XAD extract of *X. albilineans* Δ*alb15* gene inactivation mutant. MRM transitions were 843 $m/z \rightarrow$ 660 m/z for albicidin (black) and 886 $m/z \rightarrow$ 703 m/z for carbamoyl-albicidin (red). The presence of carbamoyl-albicidin (R_t = 3.1 min) can only be seen in the wildtype, whereas albicidin (R_t = 3.4 min) is present in both extracts.

we identified another minor abundant compound with a delta mass of +43.0030 Da compared to albicidin ($[M + H]^+$ = 843.2636 Da). The exact mass of the pseudomolecular ion of this new compound $[M + H]^+$ = 886.2666 Da corresponds to a molecular formula of $C_{45}H_{40}O_{13}N_7$ (mass error Δm = −2.0 ppm, calculated exact mass $[M + H]^+$ = 886.2684 Da). This differs from the molecular formula of albicidin by a fragment corresponding to CONH, which could be interpreted as a carbamoyl group. This assumption is in line with the presence of a putative carbamoyltransferase gene (*alb15*) in the albicidin gene cluster. Alb15 shows high similarity (∼30% identity) to TobZ, an ATP-dependent *O*-carbamoyltransferase involved in the carbamoylation of tobramycin of which structural data are available.[17] Interestingly, all amino acid residues in TobZ, which are involved in complexation of a metal ion and ATP binding, are highly conserved in Alb15 as well (alignment is shown in the Supporting Information, Figure 1). This indicates that Alb15 is most likely an ATP-dependent carbamoyltransferase.

DOI: 10.1021/acschembio.5b01001
ACS Chem. Biol. 2016, 11, 1198−1204

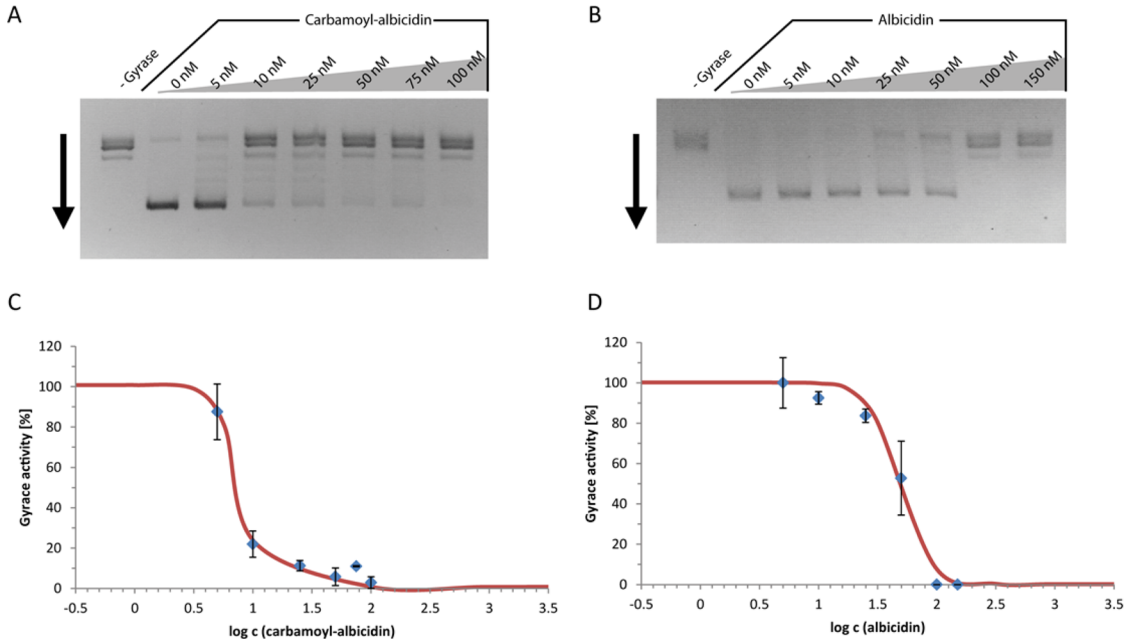


Figure 3. (A, B) *In vitro* determination of the half maximal inhibitory concentration (IC_{50}) of carbamoyl-albicidin and of albicidin against bacterial gyrase. (C, D) Densitometric analysis of the gyrase assay shown in A and B. The intensity of the lower supercoiled lane is plotted against the concentration of carbamoyl-albicidin or albicidin. Error bars represent the standard deviation of triplicates. For the determination of the IC_{50}, the density of relaxed DNA at 0 nM carbamoyl-albicidin or albicidin was set as 100% gyrase activity, e.g., no inhibition. A regression curve was fitted, based on the following function: $f(x) = 100/(1 + 10\text{^}((x\text{-}LOG(IC_{50}))*slope))$. IC_{50} values were determined to be ~49 nM for albicidin and ~8 nm for carbamoyl-albicidin.

In order to unambiguously identify the carbamoylation site and due to only minute amounts of carbamoyl-albicidin from fermentations, which were insufficient for NMR experiments, we performed mass spectrometric product ion scans of both albicidin and carbamoyl-albicidin (Figure 1), respectively. Albicidin ($[M + H]^+$ = 843.3 Da) shows a characteristic b-ion series at 160.6 Da (b_1), 280.0 Da (b_2), 495.0 Da (b_4), and 660.0 Da (b_5). Additionally, we observed two characteristic y-ions at 330.6 Da (y_2-H_2O) and 468.0 Da (y_3). For the b-ion series of the putative carbamoyl-albicidin, we observed a characteristic pattern of ions at 203.8 Da (b_1), 323.0 Da (b_2), 537.8 Da (b_4), and 702.8 Da (b_5), which is shifted throughout the spectrum by 43.0 Da. The y-ions observed for albicidin could be seen for carbamoyl-albicidin at 330.8 Da (y_2-H_2O) and 467.8 Da (y_3) as well. This indicates that the carbamoyl group must be located at the *N*-terminus, attached to the *para*-hydroxy group of the methyl-coumaric acid. The ultimate proof for such a structural arrangement came from the total synthesis of carbamoyl-albicidin. The main synthesis strategy is based on an orthogonal protecting group at the *para*-hydroxy function that allows site-specific carbamoylation before the global deprotection step (shown in Supporting Information Scheme 1), according to the recently published total synthesis of albicidin by our group.[14] Due to the fact that acidic, basic, and hydrogenolytic conditions could not be applied for the total synthesis, we decided to use a *tert*-butylsilyl (TBS) protecting group, which is mildly cleavable with a fluorine source. The carbamoyl moiety was introduced by using chlorosulfonyl isocyanate (CSI) followed by global deprotection according to Dauvergne *et al.*[18] To this end, synthetic carbamoyl-albicidin had an identical retention time in analytical HPLC runs as the product isolated from the host organism. The exact mass of synthetic carbamoyl-albicidin was determined as $[M-H]^-$ = 884.2554 Da (mass error Δm = −3 ppm, calculated mass 884.2527 Da). The pattern of the MS/MS spectrum of synthetic carbamoyl-albicidin exactly matches with the pattern of the natural product, thus ultimately giving proof of the proposed structure (Supporting Information Figure 2).

In order to characterize the gene function of *alb15*, a gene inactivation mutant was generated. Mass spectrometric characterization by multiple reaction monitoring (MRM) of *X. albilineans*−Δ*alb15* compared to *X. albilineans* wild type showed the absence of carbamoyl-albicidin in the mutant strain, while albicidin still could be detected in both cultures (Figure 2). Subsequently, we cloned the *alb15* gene into the expression vector pETtrx_1c and heterologously expressed the gene as a thioredoxin-Alb15-His_6 fusion protein in *E. coli BL 21 gold*. After protein purification through Nickel-affinity chromatography and gel filtration, yields obtained for Alb15 were ~5 mg/$L_{culture}$ (Supporting Information Figure 3). The purity and identity of the fusion protein was verified by SDS-PAGE (Supporting Information Figure 3), in-gel trypsin digestion and LC-MS/MS analysis of the tryptic peptides (Supporting Information Figure 4). We then reconstituted the carbamoylation by incubating purified Alb15 with ATP, albicidin, and carbamoyl-phosphate. The detection of carbamoylated albicidin was performed by ESI-MRM mass spectrometry. The MRM chromatograms in Figure 2A show four points of time (1, 10, 20, and 120 min), including negative controls (no carbamoyl-phosphate and no ATP at 120 min). We observed peaks at a characteristic retention time for carbamoyl-albicidin of R_t = 3.0 min from approximately 20 min of incubation onward. These results indicate that the formation of carbamoyl-albicidin occurs with albicidin as a substrate, most likely post-NRPS, which is in line with other known carbamoylation reactions, as for example the post-PKS carbamoylation of ansamitocin.[19,20]

To assess the effect of carbamoylation on the physicochemical properties and antibacterial activity of albicidin which may

DOI: 10.1021/acschembio.5b01001
ACS Chem. Biol. 2016, 11, 1198−1204

be of significance for further medicinal chemistry studies, we performed a simple agar halo assay against *E. coli* DH5α (Supporting Information Figure 5). We observed inhibition zones that were approximately 15% larger for carbamoyl-albicidin in comparison to those of albicidin (at 1, 2, and 10 ng/spot). To quantify the antibacterial activity in a more detailed way, we determined the minimal inhibitory concentration (MIC) against two Gram-negative (*E. coli* K-12 and *Salmonella typhimurium*) and two Gram-positive (*Bacillus subtilis* and *Mycobacterium phlei*) strains (Supporting Information Table 2). We observed MICs of ∼0.2 ng/μL for albicidin and carbamoyl-albicidin against both Gram-positive strains. The MIC of carbamoyl-albicidin against *Salmonella typhimurium* was 3.1 ng/μL compared to 6.3 ng/μL for albicidin, which corresponds to an approximately 50% stronger inhibition. For *E. coli* K-12, we observed MIC values of ∼0.1 ng/μL for carbamoyl-albicidin and an approximately 50% better value for albicidin (0.06 ng/μL). In comparison to the apramycin, these MICs were more than 1 order of magnitude lower. To investigate the influence of the carbamoyl-group on the inhibitory potential of gyrase, we performed *in vitro* gyrase DNA supercoiling assays. The assay is based on different migration of DNA-topoisomers generated in an ATP-dependent manner by gyrase in agarose gels. Hence, we incubated relaxed DNA with bacterial gyrase at different concentrations of carbamoyl-albicidin and albicidin, respectively. As shown in the representative gels in Figure 3A and B and in the densitometric plots in Figure 3C and D, no supercoiling occurs in the reaction without gyrase (w/o gyrase) and in a dose-dependent inhibition (0−100 nM carbamoyl-albicidin) which reaches from inhibition of supercoiling activity (100 nM) to almost complete conversion (0 nM). Through a regression curve and numeric fitting, we calculated the IC_{50} of carbamoyl-albicidin to be ∼8 nM, which is nearly a 6 times higher inhibition compared to the albicidin control of ∼49 nM, the latter being in a similar range as the results from our previous study.[14]

In summary, it was shown that Alb15 is a carbamoyltransferase that transfers a carbamoyl moiety to the hydroxy-group of the coumaric acid residue at the *N*-terminus of albicidin, most likely as a post-NRPS reaction. Compared to albicidin, carbamoyl-albicidin exhibits higher inhibition of bacterial gyrase, while the antibacterial effects are strain-dependent. Besides the biological relevance, these findings are important for the deepened understanding of gyrase inhibition by albicidin. Hence, the higher *in vitro* activity strongly suggests that the *N*-terminal region of albicidin is important for gyrase interaction. Based on these findings, we will further investigate the influence of *N*-terminal modifications in our synthetic studies on albicidin as a potentially new anti-infective drug.

EXPERIMENTAL SECTION

Fermentation, Isolation, and Purification of Carbamoyl-albicidin. Liquid cultures of a heterologous albicidin-producing strain (strain Xves-alb of *Xanthomonas axonopodis* pv. *vesicatoria*)[13] were prepared in plastic tubes (Corning Inc.). Cultures were grown under agitation (100 rpm) for 5 days at 28 °C in 72 tubes each containing 200 mL (total = 14.4 L) of the optimized XVM3B medium (K_2HPO_4 0.24 mM; KH_2PO_4 0.12 mM; $(NH_4)_2SO_4$ 10 mM; $MgSO_4{\cdot}7H_2O$ 5 mM; casamino acids 0.015%; $FeSO_4$ 0.01 mM; $CaCl_2$ 1 mM; NaCl 20 mM; glycerol 6 g/L). The isolation of carbamoylated albicidin was adapted from our previous procedure.[7] Briefly, XAD-7 Amberlite was added to the fermentation broth of strain Xves-alb to adsorb albicidin from the supernatant. After eluting the XAD material with 100% MeOH, the fraction was evaporated in a rotary evaporator and redissolved in MeOH. After centrifugation, the supernatant was purified by preparative HPLC on an Agilent 1100 system (Agilent) at a detection wavelength of 308 nm on a C_{18} reversed-phase column (GromSil 120 ODS 5 ST, 10 μm; 250 × 20 mm, Grace) using a linear MeOH gradient starting from 35% MeOH plus 0.1% HFO to 80% MeOH plus 0.1% HFO for 40 min at a flow rate of 15 mL/min. The collected fraction was freeze-dried and redissolved in aqueous tetrahydrofuran (44%) with acetic acid (1%) and subsequently purified using an analytical Agilent 1200 HPLC system (Agilent) at a detection wavelength of 308 nm, using isocratic conditions (THF 44%/HFO 1%; 1 mL/min) on a polymeric reversed-phase (PRP-1, 5 μm; 305 × 7 mm, Hamilton). An additional step of purification was performed with reversed-phase HPLC on an analytical Agilent 1200 HPLC system (Agilent) at a detection wavelength of 308 nm using an Agilent Zorbax RX-C_{18} column (250 × 4.6 mm; 5 μm; Agilent) operated at a flow rate of 1 mL/min. Elution was performed *via* a gradient of H_2O/ACN/TFA 0.05% from 10 to 80% ACN in 60 min. All fractions were checked by bioactivity test (*E. coli* halo assay) and LC-MS analysis.

Construction of the *alb15* Deletion Mutant of *X. albilineans*. The preparation of the *alb15* deletion mutant in *X. albilineans* was performed according to Rott *et al.*[21] All primers used are listed in the Supporting Information (supplemental Table 1). This method is based on the principle of double recombination. Fragments flanking the *alb15* deletion were amplified from the plasmid pYOAAB23CA09 (Genomic library of strain GPE PC73 of *X. albilineans*[22]) with primers AalbXV/BalbXV and CalbXV/DalbXV, respectively. Resulting PCR fragments were joined at the level of the complemented 24 bp-sequence of primers BalbXV and CalbXV, respectively. The resulting fragment was cloned into Strataclone vector (Stratagene, La Jolla, USA), yielding plasmid pStrata-alb15Xa. Transformants were screened with primers CriblA/CriblB. A *Bcl*I insert of pStrata-alb15Xa was then cloned into the pUFR080 (*sacB*) suicide vector,[23] digested by *Bam*HI, yielding pUFR080-alb15. Plasmid pUFR080-alb15 was subsequently introduced into the *X. albilineans* strain GPE PC73 RS by electroporation. Transformants were plated on Wilbrink agar plates without sucrose and supplemented by 1% glucose and 20 μg/mL kanamycin. Plates were incubated at 28 °C for 5 to 7 days until isolated colonies appeared, which correspond to mutants in which a first recombination occurred. At this point, to ensure the recombination occurred in the target gene (*alb15*), transformants were screened with primers CriblC/CriblB. Selected transformants were then transferred on classic Wilbrink medium to allow the second recombination to occur. To ensure that the deletion occurred, resulting colonies were screened with primers CriblC/CriblB, CriblA/CriblB, and CriblD/CriblB. The PCR product obtained with primers CriblC/CriblB was sequenced using primers CriblC and CriblB.

MS Experiments. Full-scan measurements were routinely performed on an Exactive ESI-Orbitrap-MS (Thermo Fisher Scientific GmbH) coupled to an analytical HPLC 1200 system (Agilent) using a Thermo Hypersil-Gold (5 μm, 100 × 2.1 mm) column with a linear gradient at 0.3 mL/min from 5% B to 100% B (A = water + 0.1% formic acid (HFO), B = acetonitrile + 0.1% HFO) over 6 min followed by a 4 min washout phase at 100% B and a 3 min re-equilibration phase at 5% B.

MS/MS experiments were performed on a triple-quadrupole mass spectrometer coupled to an analytical UHPLC 1290 system (Agilent) using a Grace C_{18} column (3 μm, 50 × 2.1 mm) with a linear gradient from 5% B to 100% B (A = water + 0.1% HFO, B = acetonitrile + 0.1% HFO) over 6 min followed by a 4 min washout phase at 100% B and a 3 min re-equilibration phase at 5% B. For product ion scans, m/z 843 and m/z 886 were selected as precursors with a unit mass selection window. Normalized collision energy was set to 8%.

For the detection of albicidin and carbamoyl-albicidin, the same triple-quadrupole mass spectrometer with identical chromatographic conditions was used. Detection was performed through multiple reaction monitoring (MRM) mass spectrometry using the transitions m/z 843 → m/z 660 and m/z 886 → m/z 703 as typical product ions for albicidin and carbamoyl-albicidin, respectively. Normalized collision energy was set as well to 8%.

DOI: 10.1021/acschembio.5b01001
ACS Chem. Biol. 2016, 11, 1198−1204

Cloning, Expression, and Purification of Alb15. For a detailed procedure, see the Supporting Information. In short, the *alb15* gene was amplified by PCR using the cosmid pALB540[6] as a template and cloned into pETtrx_1c and subsequently transformed into *E. coli BL 21 gold*. Protein expression was carried out in Terrific Broth medium at 37 °C and 200 rpm for 2 h followed by 16 h at 18 °C and 180 rpm. After lysis, the protein was purified by Ni-affinity chromatography and size exclusion chromatography using an Äkta purification system (GE Healthcare). The identity of the Alb15 fusion protein was verified by SDS-PAGE and in-gel trypsin digestion (supplemental Figures 3 and 4). Finally, the protein solution was concentrated to a final concentration of 3.3 mg mL^{-1} in 20% glycerin/phosphate buffer, shock frozen in liquid N_2, and stored at −20 to −80 °C.

***In Vitro* Carbamoylation Assay.** A 25-μL reaction mix containing 4 mM DTT, 400 μM $MnCl_2$, 4 mM $MgCl_2$, 40 mM carbamoyl-phosphate, and 20 mM ATP was mixed with 135 μL of HEPES buffer (100 mM; pH 6.7) and 5 μL of albicidin (2.0 mg mL^{-1}). After the addition of 30 μL of Alb15 (3.3 mg mL^{-1}), samples were incubated at 30 °C, and the reaction was stopped at different time points (1, 10, 20, and 120 min) by adding 20 μL of formic acid (100%). Carbamoylated albicidin was extracted with 300 μL of ethyl acetate. The organic supernatant was removed and evaporated in a vacuum centrifugation (Thermo Scientific), redissolved in aqueous ACN (20% + 1%HFO), and measured via MRM.

Total Synthesis of Carbamoyl-albicidin. The synthesis of carbamoyl-albicidin was performed according to Kretz *et al.*[14] In short, we modified (a detailed description can be found in the Supporting Information) the synthesis scheme by making use of the *tert*-butyldimethylsilyl (TBS) protecting group for the *para* hydroxy group of the cinnamic acid. After coupling of the pentapeptide to the cinnamic acid moiety according to Kretz *et al.*[14] and cleavage of the TBS ether with tetra-*n*-butylammonium fluoride (TBAF), the carbamoyl moiety was introduced by using a chlorosulfonyl isocyanate (CSI) reagent, subsequently followed by the final deprotection of the allyl protection groups.

***In Vivo* Bioactivity.** Halo assays were performed on 1.5% LB-agar plates (15 mL) with 0.75% LB top-agar (4 mL) containing 20 μL of an overnight culture of *E. coli* DH5α. Albicidin and carbamoyl-albicidin were applied, dissolved in DMSO to yield total amounts of 0.1, 1, 2, and 10 ng/spot. Inhibition zones were measured after overnight incubation. The assay was performed in duplicate.

The minimal inhibitory concentration was determined according to the Clinical and Laboratory Standards Institute, M31-A2:[24] In short, MHB Medium was inoculated with overnight cultures of *Bacillus subtilis*, *Mycobacterium phlei*, *E. coli* K-12 *BW25113*, and *Salmonella typhimurium* and aliquoted with 200 μL/well into sterile flat bottom microtiter plates. After the addition of the albicidin, carbamoyl-albicidin or apramycin dilution series, the plates were incubated overnight at 37 °C without shaking. Finally, the optical density (OD) at 625 nm of each well was analyzed using an Infinite 200 plate reader (Tecan).

Gyrase Activity. Gyrase supercoiling experiments were performed in a total volume of 20 μL of gyrase buffer (protocol by NEB, Frankfurt, Germany). The incubations contained 60 ng of relaxed pUC19 plasmid DNA (NEB, Frankfurt, Germany), 1 unit of DNA-gyrase (2.2 nM; NEB), and various concentrations of carbamoyl-albicidin (5−100 nM). Samples were incubated at 37 °C for 45 min and subsequently heated at 65 °C for 15 min in order to inactivate the gyrase. Electrophoretic analysis was performed on a 1% agarose gel. Staining of bands was performed with ethidium bromide. For the determination of the IC_{50}, the gels were photographed and densitometrically analyzed with ImageJ (National Institutes of Health). The peak area of the densitometric analysis of the control without inhibitor was set as 100% enzyme activity. IC_{50} values were calculated through numeric regression using solver (Microsoft) based on the logarithmic equation: $f(x) = 100/(1 + 10\text{^}((x\text{-}LOG(IC_{50}))*slope))$.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.5b01001.

Additional experimental information; supplemental Scheme 1, Tables 1 and 2, and Figures 1−5 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: suessmuth@chem.tu-berlin.de.

Author Contributions

J.A., S.C., and A.P. performed the cultivation and isolation of carbamoyl-albicidin. D.P. and A.P. performed the MS/MS experiments. D.P. performed the bioinformatic analysis. D.P. and B-F.H. performed the heterologous production of Alb15 and performed the *in vitro* carbamoylation. M.M. and M.R. produced the gene inactivation mutant. D.K. and L.v.E. synthesized carbamoyl-albicidin. D.P. and S.S. performed the antibacterial and gyrase *in vitro* assays. D.P., A.P., and R.D.S. designed the study and analyzed the data. D.P. and R.D.S. wrote the manuscript. All authors read, discussed, and approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank M. Kunert and M. Seidel for assistance in the MIC assays and A. Mainz for helpful discussions on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG SU239/11-1; SU 18-1), the Cluster of Excellence "Unifying Concepts in Catalysis" (UniCat, DFG) and by the Agence Nationale de la Recherche (ANR-09-BLAN-0413-01).

■ REFERENCES

(1) Birch, R. G. (2001) *Xanthomonas albilineans* and the antipathogenesis approach to disease control. *Mol. Plant Pathol. 2*, 1−11.

(2) Rott, P. (2000) *A Guide to Sugarcane Diseases*, Editions Quae, Cirad, Paris.

(3) Champoiseau, P., Rott, P., and Daugrois, J. H. (2009) Epiphytic Populations of *Xanthomonas albilineans* and Subsequent Sugarcane Stalk Infection Are Linked to Rainfall in Guadeloupe. *Plant Dis. 93*, 339−346.

(4) Daugrois, J. H., Dumont, V., Champoiseau, P., Costet, L., Boisne-Noc, R., and Rott, P. (2003) Aerial Contamination of Sugarcane in Guadeloupe by Two Strains of *Xanthomonas albilineans*. *Eur. J. Plant Pathol. 109*, 445−458.

(5) Mensi, I., Vernerey, M. S., Gargani, D., Nicole, M., and Rott, P. (2014) Breaking dogmas: the plant vascular pathogen *Xanthomonas albilineans* is able to invade non-vascular tissues despite its reduced genome. *Open Biol. 4*, 130116.

(6) Royer, M., Costet, L., Vivien, E., Bes, M., Cousin, A., Damais, A., Pieretti, I., Savin, A., Megessier, S., Viard, M., Frutos, R., Gabriel, D. W., and Rott, P. C. (2004) Albicidin pathotoxin produced by *Xanthomonas albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. *Mol. Plant-Microbe Interact. 17*, 414−427.

(7) Cociancich, S., Pesic, A., Petras, D., Uhlmann, S., Kretz, J., Schubert, V., Vieweg, L., Duplan, S., Marguerettaz, M., Noell, J., Pieretti, I., Hugelland, M., Kemper, S., Mainz, A., Rott, P., Royer, M., and Sussmuth, R. D. (2015) The gyrase inhibitor albicidin consists of *p*-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol. 11*, 195−197.

DOI: 10.1021/acschembio.5b01001
ACS Chem. Biol. 2016, 11, 1198−1204

(8) Huang, G., Zhang, L., and Birch, R. G. (2001) A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans*. *Microbiology 147*, 631−642.
(9) Birch, R. G., and Patil, S. S. (1985) Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in Escherichia coli. *Microbiology 131*, 1069−1075.
(10) Birch, R. G., Patil, S. S. (1985) Antibiotic and process for the production thereof, U.S. Patent No. 4,525,354.
(11) Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A., and Birch, R. G. (2007) The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob. Agents Chemother. 51*, 181−187.
(12) Maxwell, A., and Lawson, D. M. (2003) The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. *Curr. Top. Med. Chem. 3*, 283−303.
(13) Vivien, E., Pitorre, D., Cociancich, S., Pieretti, I., Gabriel, D. W., Rott, P. C., and Royer, M. (2007) Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent inhibitor of DNA gyrase. *Antimicrob. Agents Chemother. 51*, 1549−1552.
(14) Kretz, J., Kerwat, D., Schubert, V., Gratz, S., Pesic, A., Semsary, S., Cociancich, S., Royer, M., and Sussmuth, R. D. (2015) Total synthesis of albicidin: a lead structure from *Xanthomonas albilineans* for potent antibacterial gyrase inhibitors. *Angew. Chem., Int. Ed. 54*, 1969−1973.
(15) Brewer, S. J., Taylor, P. M., and Turner, M. K. (1980) An adenosine triphosphate-dependent carbamoylphosphate–3-hydroxymethylcephem *O*-carbamoyltransferase from *Streptomyces clavuligerus*. *Biochem. J. 185*, 555−564.
(16) Freel Meyers, C. L., Oberthur, M., Xu, H., Heide, L., Kahne, D., and Walsh, C. T. (2004) Characterization of NovP and NovN: completion of novobiocin biosynthesis by sequential tailoring of the noviosyl ring. *Angew. Chem., Int. Ed. 43*, 67−70.
(17) Parthier, C., Gorlich, S., Jaenecke, F., Breithaupt, C., Brauer, U., Fandrich, U., Clausnitzer, D., Wehmeier, U. F., Bottcher, C., Scheel, D., and Stubbs, M. T. (2012) The *O*-carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction. *Angew. Chem., Int. Ed. 51*, 4046−4052.
(18) Dauvergne, J., Wellington, K., and Chibale, K. (2004) Unprecedented observation of sulfonamides in the transesterification of *N*-unsubstituted carbamates with sulfonyl chlorides. *Tetrahedron Lett. 45*, 43−47.
(19) Yu, T. W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K. Q., Xu, J., Moss, S. J., Leistner, E., and Floss, H. G. (2002) The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. *Proc. Natl. Acad. Sci. U. S. A. 99*, 7968−7973.
(20) Li, Y., Zhao, P., Kang, Q., Ma, J., Bai, L., and Deng, Z. (2011) Dual carbamoylations on the polyketide and glycosyl moiety by asm21 result in extended ansamitocin biosynthesis. *Chem. Biol. 18*, 1571−1580.
(21) Rott, P., Fleites, L., Marlow, G., Royer, M., and Gabriel, D. W. (2011) Identification of new candidate pathogenicity factors in the xylem-invading pathogen *Xanthomonas albilineans* by transposon mutagenesis. *Mol. Plant-Microbe Interact. 24*, 594−605.
(22) Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., Couloux, A., Darrasse, A., Gouzy, J., Jacques, M. A., Lauber, E., Manceau, C., Mangenot, S., Poussier, S., Segurens, B., Szurek, B., Verdier, V., Arlat, M., and Rott, P. (2009) The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics 10*, 616.
(23) Castaneda, A., Reddy, J. D., El-Yacoubi, B., and Gabriel, D. W. (2005) Mutagenesis of all eight avr genes in *Xanthomonas campestris pv. campestris* had no detected effect on pathogenicity, but one avr gene affected race specificity. *Mol. Plant-Microbe Interact. 18*, 1306−1317.
(24) Shryock, T. R. (2002) *Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals: approved standard*, NCCLC, Wayne, PA.

DOI: 10.1021/acschembio.5b01001
ACS Chem. Biol. 2016, 11, 1198−1204