Régulation in labo et in situ de l'expression de génomes de souches bactériennes méthylotrophes dégradant le dichlorométhane

Bruno Maucourt

▶ **To cite this version:**

Bruno Maucourt. Régulation in labo et in situ de l'expression de génomes de souches bactériennes méthylotrophes dégradant le dichlorométhane. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Strasbourg, 2019. Français. NNT : 2019STRAJ003 . tel-02454963

HAL Id: tel-02454963
https://theses.hal.science/tel-02454963

Submitted on 25 Jan 2020

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

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

UNIVERSITÉ DE STRASBOURG

École doctorale des Sciences de la Vie et de la Santé

UMR 7156 Génétique Moléculaire, Génomique, Microbiologie

Thèse présentée par : Bruno MAUCOURT

soutenue le : 24 janvier 2019

pour obtenir le grade de : Docteur de l'Université de Strasbourg

Discipline : Sciences du vivant

Spécialité : Biologie des organismes : développement et physiologie

Régulation *in labo* et *in situ* de l'expression de génomes de souches bactériennes méthylotrophes dégradant le dichlorométhane

THÈSE dirigée par :

Mme. BRINGEL Françoise	Directrice de recherche CNRS, Université de Strasbourg, France

RAPPORTEURS :

M. NASSER William	Directeur de recherche CNRS/INSA, Université de Lyon, France
M. VAN DER MEER Jan Roelof	Professeur, Université de Lausane, Suisse

AUTRE MEMBRE DU JURY :

Mme. ARSÈNE-PLOETZE Florence	Enseignante chercheuse, Université de Strasbourg, France

Remerciements

Ce travail de thèse a été réalisé à l'Université de Strasbourg au sein de l'unité mixte de recherche 7156 (UMR 7156 – CNRS) « Génétique Moléculaire, Génomique, Microbiologie ». Je tiens à remercier son directeur, le Dr. Ivan Tarassov, ainsi que tout le personnel pour l'accueil qui m'a été réservé.

Tout d'abord, je tiens à remercier tous les membres du jury d'avoir accepté d'évaluer mon travail. Merci William Nasser, Jan Roleof van der Meer et Florence Ploetze.

Un grand merci à Françoise, ma directrice de thèse, pour la grande confiance et liberté qu'elle m'a accordées lors de ma thèse. Tu m'as permis de pouvoir découvrir de nombreux aspects du monde de la recherche tel que l'enseignement. Merci pour l'expérience que tu m'as permis d'acquérir.

Merci également à Stéphane pour m'avoir accueilli dans ton équipe et soutenu. Merci d'avoir pris du temps que ce soit pour parler de l'archéologie des souches bactériennes dégradant le dichlorométhane mais aussi de la thèse autour d'un verre.

Je remercie l'Université de Strasbourg qui a financé cette thèse par un contrat doctoral du ministère de l'enseignement supérieur, de la recherche et de l'innovation.

Je tiens également à remercier toutes les personnes qui ont participé à ce projet : Christine Carapito et Gauthier Husson pour l'aide sur les expériences de protéomique que l'on a réalisées ensemble ; Benoit Masquida pour l'aide apportée lors de la purification de protéine et la cristallographie.

Merci à tous les membres de l'équipe AIME. Merci à Christelle pour l'aide et le soutien apportés dès mon stage de master. Emilie M., merci également pour ton aide et pour avoir veillé à ma glycémie via toutes les barres chocolatées que tu m'as offertes lors de ma thèse. Thierry, ton aide et les discussions sur les billes de sucre et la fabuleuse bactérie oscillocoque ont été très enrichissantes … Yousra, merci aussi pour ton aide pour mes manips, mais surtout pour nos nombreuses

discussions. Elles m'ont été d’un grand secours pour retrouver le moral notamment quand mes manipulations fonctionnaient difficilement.

Je tiens également à remercier tous les doctorants et post-doc de l’équipe. Un grand merci à Pauline pour tous ces « débats qui se terminent toujours en monologue ». Tu as toujours été là pour m’écouter et me conseiller. Tous tes conseils m’ont été d’une grande utilité. Sabrina je ne vais pas faire « le tour du saint cordon » ni raconter de « *carabistouilles* ». Il ne me reste donc plus qu’à te dire merci pour tout ce que tu m’as appris du patois valenciennois jusqu’à la production et purification de protéines. Merci à Louis, mon compatriote lorrain, pour toutes ces conversations sur le chlorobenzoate ou sur la prononciation du mot « *narreux* ». Merci également d'avoir mis à disposition ta voiture pour ma recherche (les eaux polluées transportées dans le coffre de ta voiture). Jing, je n’ai qu’un mot à dire « 谢谢 ». Ta bonne humeur a été précieuse lors de la fin de ma thèse. Emilie G. merci pour toutes les discussions que l’on a partagées sur les bulles ~~d’air~~ d’eau et la microfluidique ainsi que sur les dernières sorties cinéma. Merci encore à Karine et à ceux qui sont passés trop rapidement au labo, Charline, Esshan, Ludo.

Un très grand merci à Alicia et Emmanuelle qui m’ont épaulé par les manipulations dans le cadre de leurs stages. Cela a été un grand plaisir pour moi de travailler avec vous deux.

Durant ces années au labo, de nombreux stagiaires sont passés au labo et ont participé à la bonne ambiance, merci à tous. Plus particulièrement merci à Gisèle, Farah, Clément, Stéphanie, Meriem, Elova (Sisi Paris), Emmanuel (maître du « *paquito* »). Merci à Caroline pour les discussions sur le gaz moutarde mais bon « pas de détails ».

Je souhaite remercier toutes les personnes rencontrées durant ces quatre années au sein de l’UMR7156 - GMGM que ce soit dans un couloir, lors d’un *journal club* ou même lors d’un match de *football*. Notamment merci à Jackson, Jean-Seb, Julien et Téo. Merci à l’équipe d’enseignants de l’UE « Histologie et cytologie » pour vos conseils.

« *Obrigado* » Mario Diniz de m’avoir accueilli dans ton laboratoire et d'avoir été le premier à me faire découvrir le monde de la recherche et de la spectrométrie de masse.

Merci à toute la promotion « Patrick Forterre » du master BMO. Ces deux années de master ont été mes plus belles années d’études. Un grand merci à Matthieu pour avoir toujours été là pour m’aider et m’écouter. Ton amitié est précieuse. Prépare ton nœud pap’ pour nos prochaines randonnées. Merci à Isa et Nico respectivement mes binômes de cinéma et de jeux de rôle. Je suis heureux d’avoir pu faire tout ce chemin avec vous deux. Pour finir, merci à toute la bande de Blauvac pour ces supers vacances.

Merci également à Déborah, Jérémy, Paul, Sarah, Carole, Phil et Fanny pour les soirées jeux de sociétés. Cela a été un plaisir de jouer avec vous tous.

Pour finir, un grand merci à ma famille. Merci à Perrine pour tous les repas que l’on a partagés les dimanches soirs. Merci Maman et Papa pour votre confiance et votre soutien. Sans vous je ne serais pas là aujourd’hui.

Sommaire

Chapitre 1
1. État de l'art sur la dégradation du DCM....1
1.1 Le dichlorométhane un polluant majeur....1
1.2 Métabolisme bactérien de dégradation du dichlorométhane....4
1.3 Mécanismes d'adaptation à la croissance avec le DCM....8
1.3.1 Adduits à l'ADN....10
1.3.2 Libération intracellulaire de chlorure....12
1.3.3 Effet solvant et stress au niveau de la membrane....13
1.3.4 Utilisation du glutathion....14
1.4 Régulation de la transcription de l'îlot *dcm* via DcmR....14
1.5 Hypothèses de travail et objectifs du projet de thèse....18

Chapitre 2
2. État de l'art sur la régulation des gènes bactériens de déshalogénases....22
Revue Maucourt *et al.*, à soumettre
Transcriptional regulation of organohalogen pollutant utilisation in bacteria

Chapitre 3
3. Une vision globale de l'adaptation au DCM par des approches « -omiques »....63
3.1 Introduction générale....63
3.2 Étude de la transcription par RNA-seq chez *M. extorquens* DM4 durant la croissance avec le DCM....64
3.2.1 Article de Chaignaud *et al*....64
3.2.2 Comparaison de l'adaptation aux méthanes chlorés entre *M. extorquens* CM4 et DM4 79
3.3 Approche protéomique pour étudier les protéines chez *M. extorquens* DM4 en présence de DCM....81
3.3.1 Article de Bibi-Triki *et al*. paru en 2018....81
3.3.2 Détection des extrémités NH_2 terminale par l'approche dN-TOP....91
3.4 Détection des peptides signaux par protéomique....92
3.5 Comparaison des résultats de transcriptomique et des protéomique....97
3.5.1 Régulation de l'expression génique....99
3.5.2 Réparation de l'ADN suite à la déshalogénation....101
3.5.3 Métabolisme des aldéhydes et plus particulièrement du formaldéhyde....105
3.5.4 Modification du métabolisme des glucides, en réponse à la croissance avec le DCM. 106
3.5.5 Métabolisme du glutathion, le cofacteur de l'enzyme DCM déshalogénase....109
3.5.6 Métabolisme énergétique en condition de déshalogénation....110
3.6 Détection des sites d'initiation de la transcription par TSS-seq et recherche de motifs de régulation....113
3.6.1 Introduction....113
3.6.2 Matériel et méthode....114
3.6.2.A Construction des librairies pour le TSS-seq....114
3.6.2.B Analyse informatique des données de TSS-seq....116

3.6.3 Résultats et discussions....120
3.6.3.A Détection globale des TSS chez *M. extorquens* DM4....120
3.6.3.B Analyse des TSS des gènes du transposon catabolique *dcm*....122
3.6.3.C Analyse détaillée des TSS des gènes plus exprimés en présence de DCM ou méthanol....124
3.6.4 Conclusions de la partie TSS-seq....132
3.7 Conclusions générales sur les approches « -omiques »....133

Chapitre 4
Article Maucourt *et al.*, à publier
Revisiting dcmR*, a key regulator of bacterial dichloromethane dehalogenation*

4. Régulation transcriptionnelle de l'îlot *dcm*....134
4.1 Présentation du chapitre....134
4.2 Matériel et méthode (précisions par rapport à l'article)....134
4.2.1 Quantification des protéines par SRM....134
4.3 Annexe du chapitre 4 : données non présentées dans l'article....176
4.3.1 Matériel et méthodes....176
4.3.1.A Production et purification de la protéine DcmR de *M. extorquens* DM4 avec une étiquette pour la purification....176
4.3.1.B Retard sur gel avec la protéine DcmR....179
4.3.1.C Cristallisation de la protéine DcmR....180
4.3.1.D Étude simultanée des promoteurs de *dcmA* et *dcmR* avec des plasmides bi-rapporteurs....180
4.3.1.E Détection de la protéine DcmA par *western blot*....184
4.3.2 Résultats et discussions....186
4.3.2.A Résultats de la production de DcmR et des approches associées....186
4.3.2.B Conjugaison des plasmides bi-rapporteur chez *M. extorquens* DM4....189
4.3.2.C Détection de DcmA par *western blot*....190
4.4 Perspectives de l'article *dcmR*....191

Chapitre 5
5. Détection du gène *dcmR in situ* et diversité....193
5.1 Introduction....193
5.2 Matériels et méthodes....195
5.2.1 Comparaison des séquences protéiques....195
5.2.2 Description de l'origine des échantillons....195
5.2.2.A Échantillons de l'aquifère du site industriel de Themeroil et extraction d'ADN. 195
5.2.2.B Echantillons prélevés de microcosmes exposés ou non au DCM....196
5.2.3 Quantification des gènes *dcmA* et *dcmR* par qPCR dans l'environnement....199
5.3 Résultats et discussions....202
5.3.1 Séquence conservée du gène *dcmR* au sein du transposon *dcm*....202
5.3.2 Diversité du contexte génétique des gènes des protéines homologues à DcmR....205
5.3.3 Quantification du gène *dcmR* dans l'environnement....211
5.4 Conclusions et perspectives....214
5.5 Annexe : article de Bringel *et al*. paru en 2017....215

Chapitre 6

6. Conclusions et perspectives.....220
6.1 Conclusions.....220
6.2 Nouveau modèle d'étude de la régulation du transposon catabolique *dcm* chez *M. extorquens* DM4.....221
6.2.1 De nouvelles stratégies pour purifier la protéine DcmR ?.....222
6.2.2 Impact des métabolites du DCM dans la régulation transcriptionnelle ?.....223
6.3 Nouvelles pistes de recherches de fonctions adaptatives à la croissance sur le DCM.....225
6.3.1 METDI3043 un nouveau gène impliqué dans l'adaptation en présence de DCM ?.....225
6.3.2 Les métabolites cellulaires ont-ils un rôle dans l'adaptation à la croissance avec le DCM ?.....228
6.4 Intérêts de l'utilisation de DcmR dans le suivi de décontamination de sites pollués.....229
6.5 Conclusions à plus long terme.....230

Chapitre7

7. Annexes.....233

Liste des figures

Table des figures

Figure 1.1 : Estimation des principales sources d'émissions du dichlorométhane....2
Figure 1.2 : Dégradation biologique et abiotique du dichlorométhane....5
Figure 1.3 : Conservation de l'organisation des gènes *dcmRABC* chez les bactéries possédant le transposon catabolique *dcm*....7
Figure 1.4 : Substrats carbonés métabolisés par *M. extorquens* DM4 et voies métaboliques associées....9
Figure 1.5 : Alkylation de l'ADN et réparation via l'ADN polymérase I....11
Figure 1.6 : Présence du domaine MEDS dans de nombreuses protéines de signalisation cellulaire. 16
Figure 1.7 : Transposon catabolique *dcm* chez *M. extorquens* DM4 et promoteurs de *dcmR* et *dcmA* 17
Figure 1.8 : Plasmide rapporteur détectant le DCM....18
Figure 3.1 : Comparaison des résultats de RNA-seq pour les souches *M. extorquens* DM4 et CM4. 80
Figure 3.2 : Comparaison des résultats de RNA-seq et RT-qPCR....81
Figure 3.3 : Principe de la technique dN-TOP pour déterminer l'extrémité NH2 terminale d'une protéine....92
Figure 3.4 : Description des peptides signaux Sec et TAT....93
Figure 3.5 : Bilan de l'approche dN-TOP....94
Figure 3.6 : Détection de peptides signaux via SignalP et correction du codon d'initiation de la traduction....95
Figure 3.7 : Catégories de protéines pour lesquelles un peptide signal a été déterminé expérimentalement....96
Figure 3.8 : Comparaison de l'abondance relative des ARNm et protéines chez *M. extorquens* DM4 lors de culture en présence de DCM ou méthanol....98
Figure 3.9 : Facteur de transcription dont l'expression est régulée....100
Figure 3.10 : Mécanismes régulant la transcription chez *M. extorquens* DM4....101
Figure 3.11 : Mécanismes de réparation de l'ADN chez *M. extorquens* DM4....104
Figure 3.12 : Expression différentielle des protéines de la famille des aldéhydes oxygénases chez *M. extorquens* DM4....106
Figure 3.13 : Prédiction des voies métaboliques liées aux glucides et plus exprimées en présence de DCM....107
Figure 3.14 : Métabolisme du glutathion chez *M. extorquens* DM4....110
Figure 3.15 : Métabolisme énergétique chez *M. extorquens* DM4....112
Figure 3.16 : Constructions et séquençages des banques « + » et « - »....117
Figure 3.17 : Impact des paramètres du logiciel TSSAR sur le nombre de TSS détecté pour des données obtenues pour un réplicat en condition DCM....118
Figure 3.18 : Exemple de détection de multiples TSS pour un même gène....119
Figure 3.19 : TSS détectés pour l'ensemble du génome de *M. extorquens* DM4....121
Figure 3.20 : Sites d'initiation de la transcription et prédiction de la structure secondaire des séquences 5'UTR des transcrits de *dcmA* et *dcmR*....123
Figure 3.21 : Sites d'initiation de la transcription, identifiés pour les gènes régulés au niveau de la transcription chez *M. extorquens* DM4....125

Figure 3.22 : Motifs d'ADN proposés par le logiciel MEME en amont des TSS.....127
Figure 3.23 : Structures prédites dans les 5'UTR des gènes plus exprimés en présence de DCM...129
Figure 3.24 : Exemples de TSS détectés.....131
Figure 3.25 : « Portrait robot » d'un gène chez *M. extorquens* DM4.....133
Figure 4.1 : Principe de la quantification de protéine par SRM.....135
Figure 4.2 : Cartes génétiques des plasmides construits en vue de la production de DcmR avec une étiquette histidine.....177
Figure 4.3 : Carte génétique du plasmide pME8312.....182
Figure 4.4 : Analyse de la protéine 10His-DcmR produite chez *E. coli*.....186
Figure 4.5 : Profil de chromatographie d'exclusion stérique de la protéine 10His-DcmR.....188
Figure 4.6 : Complémentation de fonction avec 10His-DcmR ou DcmR-6His chez le mutant STOP*dcmR*.....189
Figure 4.7 : Détection de la protéine DcmA dans différents contextes génétiques.....191
Figure 5.1 :Recherche de *dcmR* par bio-informatique et dans l'environnement.....194
Figure 5.2 : Description du site de Themeroil pollué par des solvants chlorés dont le dichlorométhane (Figure adaptée de Louis Hermon 2017 avec sa permission).....196
Figure 5.3 : Description des microcosmes du projet TOP-DCM.....198
Figure 5.4 : Conservation de la séquence de l'îlot *dcm* chez différentes souches bactériennes dégradant le DCM.....204
Figure 5.5: Diversité de séquences et de contexte génétique de protéines homologues au régulateur DcmR.....206
Figure 5.6 : Alignement des séquences protéiques de *DcmR-like*.....208
Figure 5.7 : Diversité de la protéine DcmA.....209
Figure 5.8 : Environnement génétique des gènes *dcmA* et *dcmR-like* et voie métabolique potentielle chez *Rhodospirillaceae bacterium* HHTR118.....210
Figure 5.9 : Diversité des promoteurs des gènes *dcmR* et *dcmA*.....211
Figure 5.10 : Quantification des gènes *dcmR* et *dcmA* à partir d'échantillons environnementaux..212
Figure 6.1 : Mécanismes potentiels de régulation du transposon catabolique *dcm*.....222
Figure 6.2 : Ligands possibles impliqués dans la régulation de l'îlot *dcm*.....224
Figure 6.3 : Alkylation de l'ADN par des agents chimiques.....225
Figure 6.4 : Perspectives pour l'étude de METDI3043.....227
Figure 6.5 : Approche de métabolomique pour mieux comprendre l'adaptation au DCM.....228
Figure 6.6 : Elimination d'un polluant gazeux avec un biofiltre.....230

Liste des tableaux

Tableau 3.1 Nombre de lectures alignées sur le génome.....120
Tableau 3.2 Identification des sites d'initiation de la transcription des gènes dont l'abondance des transcrits varie en présence de DCM ou méthanol.....126
Tableau 3.3 Motifs trouvés au niveau des promoteurs des gènes dont l'abondance des ARNm est régulée en présence de DCM ou méthanol.....128
Tableau 3.4 Structures secondaires prédites dans les ARNm plus exprimés en présence de DCM. 130
Tableau 4.1 Préparation d'un gel SDS-PAGE.....178
Tableau 4.2 Liste des caractéristiques et des variants du plasmide pME8312 bi-rapporteurs.....183
Tableau 5.1 Liste des amorces et plasmides utilisés pour la qPCR.....201
Tableau 5.2 Souches isolées dont le transposon catabolique *dcm* a été séquencé.....203

Liste des abréviations

ABC	ATP-*Binding Cassette*
ADN	Acide désoxyribonucléique
ADNc	Acide désoxyribonucléique complémentaire
AMP/ADP/ATP	Adénosine mono/di/triphosphate
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messager
ARNr	Acide ribonucléique ribosomique
°C	Degré Celsius
C_1	Composé carboné à un seul atome de carbone
Cl	Chlorure
CBS	Cystathione β-synthase
Ct	*Cycle threshold* (cycle seuil)
DCM	Dichlorométhane (CH_2Cl_2)
DNase	Désoxyribonucléase
dNTP	Désoxynucléotide tri-phosphate (dATP, dGTP, dCTP, dTTP)
dN-TOP	*Doublet N-terminal oriented proteomics*
DO	Densité optique
EDTA	Éthylène diamine tétra-acétique
g, mg, µg, ng	Gramme, milligramme, microgramme, nanogramme
g	Gravité
GSH	Glutathion (réduit)
GST	Glutathion-S transférase
H^+	Proton
HCl	Acide chlorhydrique
H_4MPT	Tétrahydrométhanoptérine
IS	*Insertion sequence* (séquence d'insertion)
kb, Mb	Kilo bases, méga bases
kDa	Kilo Dalton
L, mL, µL	Litre, millilitre, microlitre
LB	*Lysogeny broth* (bouillon de lysogénie)
Log_2 fc	*Logarithm* 2 *of fold change*
MEDS	*Methanogen / methylotroph, DcmR sensory*
m, cm, mm, µm, nm	Mètre, centimètre, millimètre, micromètre, nanomètre
M3	Milieu minimum pour *Methylobacterium*
mA	MilliAmpère
M, mM, µM	Molaire (mole par litre), millimolaire, micromolaire
pb	Paire de bases
PBS	*Phosphate buffered saline* (tampon phosphate salin)
PCR	*Polymerase chain reaction* (réaction de polymérisation en chaîne)
pH	Potentiel hydrogène
Pz	Piézomètre
qPCR	PCR quantitative
RBS	*Ribosome binding site* (site de fixation du ribosome)
RNase	Ribonucléase
RNA-seq	*RNA sequencing* (séquençage à haut débit d'ADNc)

ROX	Carboxy-X-rhodamine
RT-qPCR	PCR quantitative après rétro-transcription
SDS	*Sodium dodecyl sulfate* (dodécylsulfate de sodium)
SDS-PAGE	*Sodium dodecyl sulfate - polyacrylamide gel electrophoresis*
sp.	*specie*
spp.	*species*
SRM	*Selected reaction monitoring*
TEX	*Terminal exonuclease* (exonucléase terminale)
Tris	Tris(hydroxyméthyl)aminométhane
Tm	*Melting temperature* (température de fusion)
TMPP	trimethoxyphenyl phosphonium
TPP	thiamine pyrophosphate
TSS	*Transcription start site* (site d'initiation de la transcription)
U	Unité
UTR	*Untranslated region* (région non traduite)
UV	Ultraviolet
V	Volt
YFP	*Yellow fluorescent protein*

Introduction générale

Mon sujet de thèse porte sur l'adaptation bactérienne à la dégradation du dichlorométhane (DCM ; CH_2Cl_2) qui est un polluant chloré toxique émis dans l'environnement principalement par l'industrie (Muller *et al*., 2011a). Le dichlorométhane est le solvant chloré le plus produit dans le monde. Certaines bactéries sont capables de métaboliser ce polluant via un métabolisme méthylotrophe qui consiste à utiliser des composés organiques réduits sans liaison carbone-carbone. *Methylorubrum extorquens* DM4 est la souche méthylotrophe la plus étudiée pour la dégradation aérobie du DCM. Les gènes *dcmRABC* localisés sur un transposon catabolique niché sur l'îlot génomique *dcm* de 126 Kb lui confèrent la capacité d'utiliser le DCM comme seule source de carbone et d'énergie (Vuilleumier *et al*., 2009). Les gènes *dcmABC* et le gène *dcmR* sont en orientation divergente. Seul le gène *dcmA* est essentiel à l'utilisation du DCM chez *M. extorquens* DM4. Il code la protéine DcmA, une déshalogénase de la famille des glutathion-S transférases caractérisée il y a presque une trentaine d'années (La Roche & Leisinger, 1990). Son activité est modulée par le gène *dcmR* lui aussi présent dans le transposon catabolique *dcm*.

La protéine DcmR est un facteur de transcription qui régule l'expression de *dcmA* ainsi que sa propre expression (La Roche & Leisinger, 1991). Ce facteur de transcription porte un domaine hélice-tour-hélice en N terminal qui lui permettrait de se fixer à l'ADN, en reconnaissant peut-être une séquence de 12 pb conservée en amont des gènes *dcmA* et *dcmR* (La Roche & Leisinger, 1991). La protéine DcmR empêcherait ainsi la reconnaissance du promoteur par l'ARN polymérase. Du côté C terminal de DcmR, un domaine a été prédit par bioinformatique mais n'est pas encore vérifié expérimentalement (Anantharaman & Aravind, 2005). Ce second domaine appelé *methanogen / methylotroph, DcmR sensory* (MEDS) est trouvé chez des protéines de plusieurs bactéries méthylotrophes ou méthanogènes. Il s'agirait d'un domaine de fixation de composés hydrocarbonés dont le DCM fait partie et pourraient ainsi servir de ligand du domaine MEDS et conduire à une modification allostérique de DcmR. Les gènes *dcmB* et *dcmC* sont quant à eux de fonction inconnue.

Les objectifs de ma thèse étaient de répondre à plusieurs questions : i) quels sont les transcrits et les protéines dont l'abondance est modifiée lors de cultures avec DCM ou du méthanol chez *M. extorquens* DM4 ? ii) déterminer comment le facteur de transcription DcmR intervient dans la régulation de l'îlot *dcm* ? iii) le gène *dcmR* est-il détectable et conservé dans l'environnement et chez les souches dégradant le DCM isolées de sites pollués ? Pour répondre à ces questions, le manuscrit est basé sur six chapitres.

Le **chapitre 1** correspond à une synthèse bibliographique sur le métabolisme bactérien du DCM et plus particulièrement par l'enzyme DcmA. Ce chapitre aborde aussi les premières recherches sur le facteur de transcription DcmR et son impact sur la régulation du transposon catabolique *dcm*.

Le **chapitre 2** est présenté sous forme de *review* en vue d'une soumission dans *FEMS Microbiology Reviews* ou *Environmental microbiology* pour une publication. Il s'agit d'une étude bibliographique sur la régulation transcriptionnelle lors du métabolisme bactérien de composés organohalogénés. Il s'agit de la première *review* traitant de ce sujet de manière globale. Les mécanismes moléculaires de la déshalogénation ont été beaucoup étudiés ces dernières années mais la régulation de ce métabolisme beaucoup moins. Or connaître les mécanismes de régulation impliqués dans ce métabolisme permettrait de mieux orienter des stratégies de bioréhabilitation de sites pollués. Les différents mécanismes connus sont décrits et comparés avec ce qui est connu pour le métabolisme du DCM.

Le **chapitre 3** aborde l'adaptation au DCM par une approche globale d'expression du génome en condition de croissance en DCM, comparée au méthanol, qui a été initiée tant au niveau transcriptionnel par RNA-seq (Chaignaud *et al*., 2017) qu'au niveau protéique par protéomique (Bibi-Triki *et al*., 2018). Ce chapitre est basé sur deux articles scientifiques qui ont été publiés respectivement dans *Frontiers in Microbiology* et *Journal of Proteomics* : (Chaignaud *et al.,* 2017 ; Bibi-Triki *et al.,* 2018). Ces résultats ont permis de confirmer la régulation des gènes du transposon *dcm* en présence de DCM. En parallèle, une étude globale de type RNA-seq a permis de cibler l'identification de l'ensemble des sites d'initiation de la transcription en condition méthylotrophe avec le DCM ou le méthanol chez *M. extorquens* DM4. Cela a permis de confirmer la présence de deux sites d'initiation pour *dcmR* et ceux d'autres gènes régulés en présence de DCM. L'ensemble des résultats de transcriptomique et de protéomique différentielles permet de répertorier au niveau global des fonctions qui n'avaient pas été identifiées précédemment par approche de mutagenèse aléatoire et de gel 2D (Muller *et al*., 2011b). De plus, cela permet d'identifier de nouveaux éléments de régulation par le DCM, dépendant ou non du régulateur DcmR.

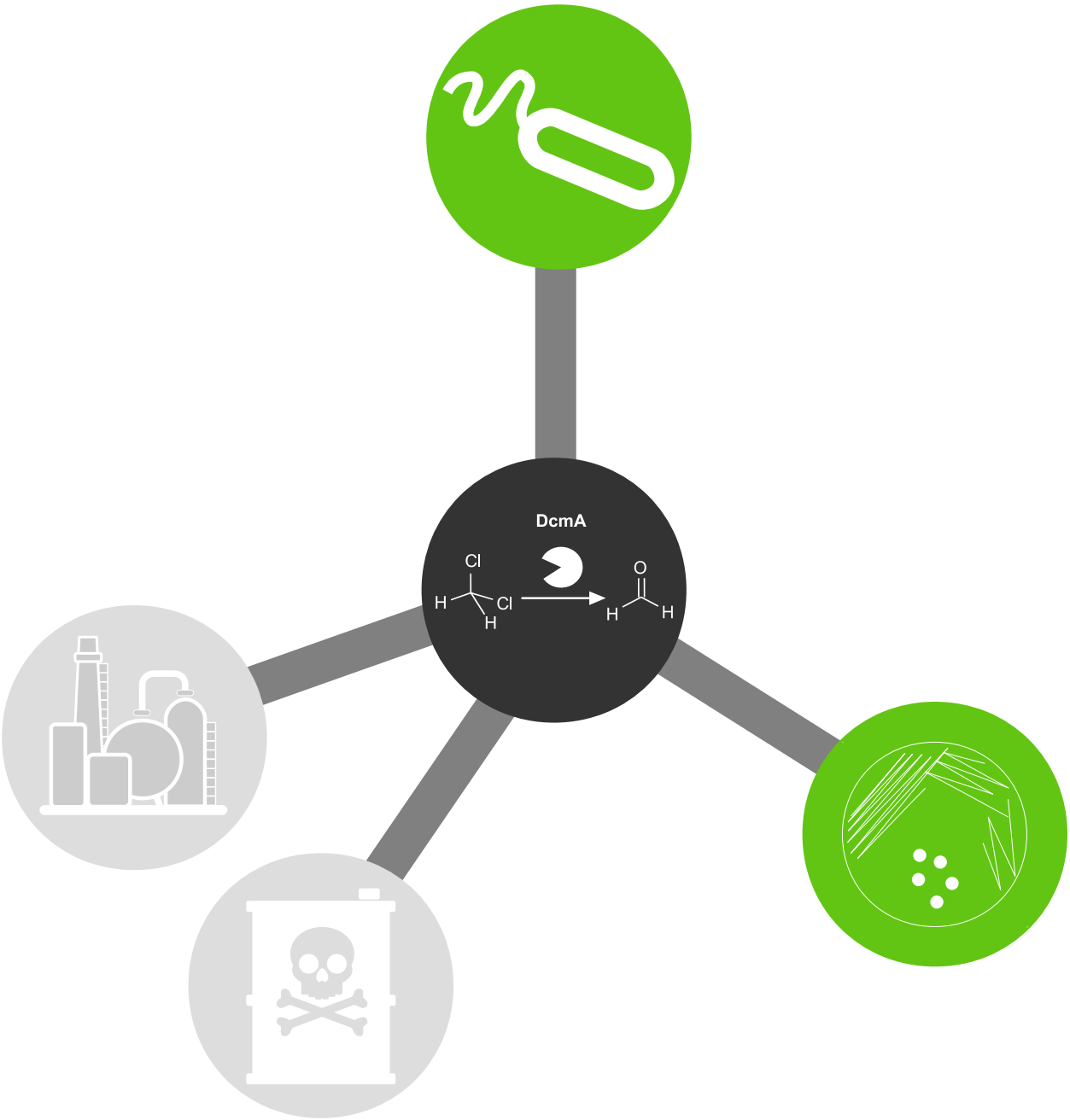
Le **chapitre 4** concerne le gène *dcmR* et la régulation du transposon catabolique *dcm*. Il a été écrit sous forme d'article dans le but d'être publié lorsque les derniers résultats complémentaires auront été obtenus. Cette étude se base sur l'analyse comparée de souches portant un gène *dcmR* muté, seul ou d'autres mutations des gènes *dcmABC* du transposon chez *M. extorquens* DM4. À partir de cultures bactériennes de la souche sauvage, et des mutants, cultivés avec du DCM comme seule source de carbone et d'électron ou avec d'autres substrats (méthanol ou succinate), le phénotype de croissance et l'expression des gènes *dcm* ont été étudiés. Pour étudier les mécanismes de régulation génétique impliquant DcmR, l'expression des gènes *dcm* a été évaluée par quantification de l'abondance des ARNm et des protéines correspondantes, et de l'activité promotrice de la transcription des régions en amont des gènes *dcmR et dcmA* clonés en aval de gènes rapporteurs (fusion transcriptionnelle). Les quatre protéines du transposon catabolique *dcm* ont été quantifiées par spectrométrie de masse avec une approche de SRM en collaboration avec Christine Carapito et Gauthier Husson (IPHC à Cronenbourg). Ces travaux ont permis de confirmer que DcmR inhibe la transcription de son propre gène ainsi que celle de *dcmA* en absence de DCM. Dans le cas de *dcmR*, cette répression a lieu au niveau d'un seul de ces promoteurs, l'autre étant exprimé de façon constitutive. Cette répression a également été observée en contexte hétérologue chez *E. coli* de manière indépendante des séquences potentielles de fixation. De plus, l'activation de la transcription lors de la dégradation du DCM se fait par un mécanisme indépendant de DcmR qui requiert la présence de séquences en amont du promoteur du gène *dcmA*. Cela laisse supposer qu'un autre facteur encore inconnu serait responsable d'une activation de la transcription de *dcmA* (cet effet n'est pas visible pour *dcmR*). Un gène (METDI3043) en dehors du transposon catabolique *dcm* a été détecté comme régulé, de façon directe ou indirecte, par les gènes *dcm*. Des résultats complémentaires qui ne sont pas décrits dans l'article sont détaillés en fin du chapitre, sous forme d'annexe.

Le **chapitre 5** décrit la diversité du gène *dcmR* et sa détection dans l'environnement. Le gène *dcmA* avait déjà été détecté dans des environnements pollués mais *dcmR* lui n'a jamais été recherché. Une analyse bioinformatique indique que le gène *dcmR* est trouvé dans d'autres contextes génétiques que celui du transposon catabolique *dcm*. Cela laisse supposer la présence d'autres voies métaboliques. Des qPCR ont permis de détecter *dcmR* et *dcmA* à la fois dans une eau d'aquifère et dans un sol de microcosme tous deux contaminés au DCM. Il s'agit de la première détection de *dcmR* dans des environnements pollués.

Le **chapitre 6** conclut sur les résultats majeurs obtenus au cours de cette thèse et les nouvelles perspectives de recherche qui en découlent. La régulation des gènes *dcmRABC* a pu être en partie élucidée pour certains aspects au cours de cette thèse. D'autres mécanismes encore non élucidés ont également été mis en évidence dans cette régulation. Ainsi, la régulation semble plus complexe que pensée précédemment et de nouvelles questions ont pu être posées à partir des résultats obtenus. En addition des gènes et protéines du transposon *dcm*, un inventaire des gènes et protéines régulés en présence de DCM a également pu être réalisé. Cet inventaire permet notamment de mieux comprendre l'adaptation de *M. extorquens* DM4 lors de la croissance avec du DCM. L'étude de diversité de DcmR ouvre de nouvelles pistes de recherches sur le domaine MEDS et les ligands pouvant être détectés, ainsi que sur une potentielle nouvelle voie métabolique. L'ensemble de ces résultats ouvre de nouvelles voies de recherches concernant la régulation des gènes *dcm*.

Le **chapitre 7** contient la bibliographie, un résumé des activités et présentations effectués durant cette thèse.

1. État de l’art sur la dégradation du DCM



Chapitre 1 : État de l’art sur la dégradation du DCM

1.1 Le dichlorométhane un polluant majeur

Le dichlorométhane (DCM) aussi appelé chlorure de méthylène est un composé halogéné de formule : CH_2Cl_2. Le DCM est émis de façon naturelle par plusieurs sources (Fig. 1) telles que les volcans (Schwandner *et al.*, 2013), la combustion de biomasse (Lobert *et al.*, 1999), les mangroves (Kolusu *et al.*, 2018), les micro- ou macro-algues (Laturnus *et al.*, 2004 ; Colomb *et al.*, 2008). De nos jours, l'origine du DCM est principalement anthropique. Il a été synthétisé pour la première fois en 1839 (Regnault, 1839). Il est depuis utilisé dans l'industrie pour ses caractéristiques de solvant : incolore, très volatil (s’évapore à 39,75 °C à une pression de 760 mm Hg (The Merck index, 2013)) et miscible dans l'eau (13,2 $g.L^{-1}$ à 25 °C (Yalkowsky *et al.*, 2010)) et dans de nombreux solvants organiques. Il dissout les graisses, les huiles et les résines. Il a de nombreuses utilisations dans l'industrie comme composant de colles, de décapants de peinture et solvant d'extraction. Il est par exemple utilisé pour extraire la caféine des grains de café (décaféination) lors de la préparation de café décaféiné. Dans l’industrie pharmaceutique, il est utilisé pour la mise en forme de médicaments (galénique). Le DCM est le solvant chloré le plus utilisé en Europe avec 100 000 tonnes commercialisées en 2013 (Eurochlor, 2014). Cette utilisation conduit à des rejets importants de DCM dans l'environnement alors même que des alternatives ont été trouvées pour remplacer ce composé qui est nocif pour la santé humaine et dangereux pour l’environnement (MacMillan *et al.*, 2012). La production mondiale continue à augmenter et continuera encore, d’après les estimations, si aucune mesure n’est prise (Feng *et al.*, 2018).

Les premières mesures du DCM atmosphérique dans les années 2000 ont montré que l’utilisation industrielle a conduit à une augmentation des émissions de ce polluant dans l’atmosphère (Elvidge *et al.*, 2015). L’analyse de l’air piégé dans des carottes de glace a permis de montrer que cette augmentation a commencé dans les années 1960 (Trudinger *et al.*, 2004). Dans l’atmosphère, le DCM a une durée de vie d’environ cinq mois en raison de sa décomposition par les rayons UV ce qui libère des groupements chlore réactifs ($Cl^{\cdot}$) qui sont néfastes pour la couche d’ozone stratosphérique (Fig. 2 ; réaction n°2). Ce polluant participe faiblement à la destruction de la couche d’ozone et du fait de sa courte durée de vie dans l’atmosphère, son usage n’a pas été banni par le protocole de Montréal interdisant l’usage de substances qui appauvrissent la couche d'ozone (Oram

et al., 2017). Ainsi, les émissions en DCM dans l'atmosphère continuent d'augmenter parallèlement à la progression de sa production (Hossaini *et al.*, 2017).

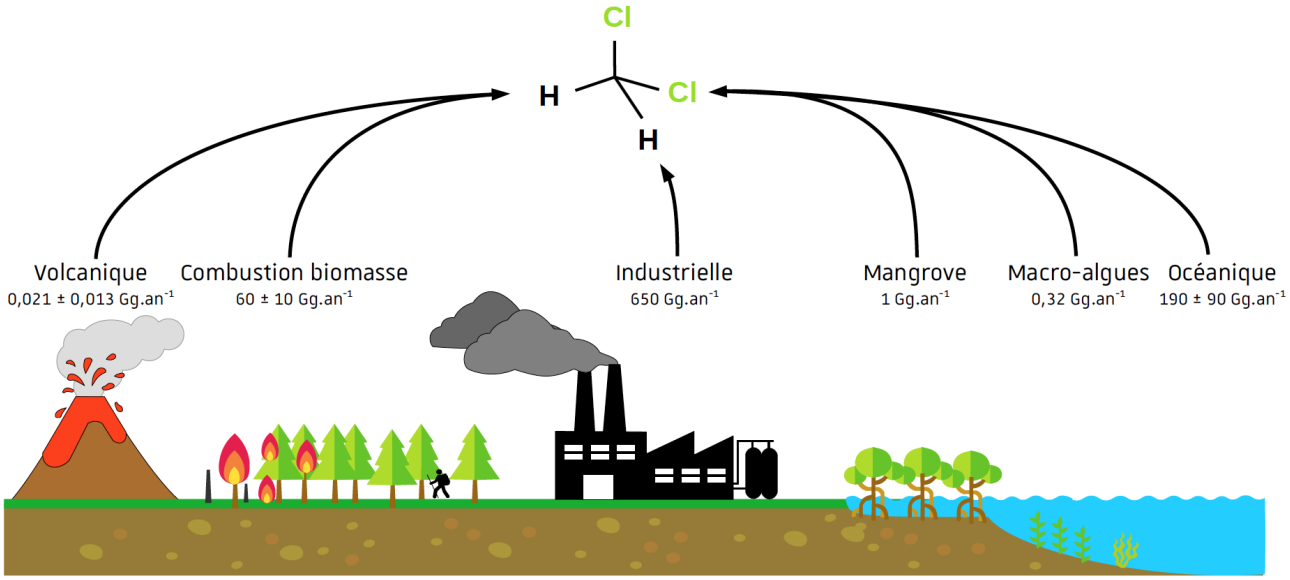


Figure 1.1: Estimation des principales sources d'émissions du dichlorométhane.
Les données numériques viennent de (Gribble, 2010) sauf pour les mangroves qui viennent de (Kolusu *et al.*, 2018). Les écarts-types de ces estimations ne sont pas disponibles dans certains cas.

Les premières données montrant des effets du DCM sur la santé humaine datent du début des années 1920-30. Il était alors utilisé comme anesthésique et solvant industriel déclenchant les premières intoxications recensées chez les personnes exposées à ce polluant dans un cadre professionnel (Bourne & Stehle, 1923 ; Collier, 1936). La toxicité du DCM a depuis été confirmée chez des animaux modèles mais aussi chez l'humain via des suivis épidémiologiques (Schlosser *et al.*, 2015) ou des tests de cultures cellulaire *in vitro* avec des cellules humaines (McDermott & Heffron, 2013). Le DCM est classé comme probablement carcinogène chez l'humain (groupe 2A) par le centre international de recherche sur le cancer (Benbrahim-Tallaa *et al.*, 2014). La toxicité du DCM à court terme (toxicité aiguë) peut s'expliquer par un effet dit « solvant » qui correspond à la capacité de ce composé à perturber les membranes lipidiques. Cependant la toxicité du DCM chez l'humain vient principalement de la formation d'intermédiaires toxiques lors de son métabolisme par les voies de détoxification impliquant soit le cytochrome P450 soit l'enzyme glutathion S-transférase (GST). L'oxydation du DCM par des cytochromes P450 conduit à la formation de monoxyde de carbone pouvant se fixer à la place de l'oxygène sur l'hémoglobine ce qui provoque des asphyxies. L'enzyme humaine de la classe *theta* des GST catalyse une réaction de conjugaison entre du DCM et du glutathion avec production de formaldéhyde. Alors que le formaldéhyde peut former des liaisons entre les protéines et l'ADN (Hu *et al.*, 2006), la réaction catalysée par la GST

génère un intermédiaire très réactif (S-chlorométhylglutathion), encore plus toxique que le DCM et le formaldéhyde qui va alkyler l'ADN (Marsch *et al.*, 2001). Ces dommages à l'ADN peuvent conduire sur le long terme (toxicité chronique) à l'émergence de cancers (Benbrahim-Tallaa *et al.*, 2014; Schlosser *et al.*, 2015).

Cette toxicité ne se limite pas aux humains et a été montrée également chez des organismes modèles comme les nématodes (Samoiloff *et al.*, 1980), la souris (Andersen *et al.*, 2017), les lentilles d'eau (Kim *et al.*, 2013) et les micro-algues (Miazek *et al.*, 2017). Chez les micro-organismes, la toxicité du DCM est plus contrastée et dépend de la capacité des micro-organismes à le dégrader. Ce polluant a peu d'effets toxiques chez *Escherichia coli* (Quintero *et al.*, 2012), *Pseudomonas putida* et *Vibrio fischeri* (Tobajas *et al.*, 2016) lorsqu'il n'est pas métabolisé. Par contre, lorsqu'il est métabolisé, il a un effet génotoxique chez *E. coli* à partir de 0,6 mM (Evans *et al.*, 2000), chez *Salmonella enterica* serovar Typhumurium (DeMarini *et al.*, 1997) et chez des bactéries méthanotrophes (Byers & Sly, 1993). Un effet solvant peut néanmoins être observé lorsqu'il est présent en forte concentration (supérieur à plusieurs dizaines de millimolaires ; (Tobajas *et al.*, 2016)). Un effet a aussi été observé dans l'environnement par la mise en évidence d'une diminution de l'activité enzymatique des micro-organismes d'un sol contaminé au DCM (1 mg de DCM pour 100 g de sol ; (Kanazawa & Filip, 1986)).

Au vue des problèmes sanitaires et environnementaux causés par ce polluant, plusieurs approches ont été développées pour dépolluer des sites industriels contaminés par du DCM (Shestakova & Sillanpää, 2013). Il s'agit notamment d'approches physico-chimiques basées sur l'adsorption sur du charbon activé, sa réduction en présence de fer zéro valent (Fig. 2 ; réaction n°1), la photo-dégradation par exposition à un rayonnement UV (Fig. 2 ; réaction n°2) ou un traitement à l'ozone (Fig. 2 ; réaction n°3) (Shestakova & Sillanpää, 2013). Ces approches sont coûteuses en énergie et peu écologiques. La découverte de bactéries se développant en utilisant le DCM comme substrat a permis le développement de nouvelles stratégies de dépollution microbiologique (Brunner *et al.*, 1980). Ce type d'approche est moins coûteux et moins consommateur en énergie ainsi que mieux accepté par le grand public. Les approches abiotiques et biologiques ne sont pas incompatibles. Au contraire leur combinaison peut permettre d'obtenir de meilleurs résultats qu'employées seules (Jianming *et al.*, 2014).

1.2 Métabolisme bactérien de dégradation du dichlorométhane

Le DCM peut être métabolisé par des bactéries par différentes voies en aérobie ou en anaérobie. Cette biodégradation n’est pas forcément liée à la croissance bactérienne. Elle peut correspondre à du cométabolisme, c’est-à-dire la transformation d’un composé ne permettant pas la croissance sauf en présence d’un autre composé servant de source de carbone et d’énergie (Frascari *et al.*, 2015). Le gène *dhlA* code une déshalogénase capable d’oxyder le DCM (Fig.2 ; réaction n°6) mais ne permet pas son utilisation comme source de carbone (Muller *et al.*, 2011a). La souche *Xanthobacter autotrophicus* GJ10 est capable d’oxyder le DCM via le gène *dhlA*. Celui-ci semble être régulé, en présence de DCM, par un facteur de transcription de la famille TetR se trouvant à proximité de *dhlA*, par un mécanisme encore inconnu (Hosoda *et al.*, 2017). Le co-métabolisme du DCM impliquant les ammonium ou méthane mono-oxygénases est dépendant de l’oxygène et ressemble au métabolisme via le cytochrome P450 (Fig.2 ; réaction n°4) (Muller *et al.*, 2011a). Ces réactions ne permettent pas la croissance des bactéries qui les catalysent, mais néanmoins elles peuvent être utilisées pour la bioréhabilitation de sites pollués si un autre substrat est présent (Frascari *et al.*, 2015).

Le DCM peut servir de source d’énergie et souvent de carbone à des micro-organismes méthylotrophes, c’est-à-dire capables d'utiliser des molécules réduites avec un seul atome de carbone ou sans liaison carbone-carbone (De Marco, 2004 ; Chistoserdova & Kalyuzhnaya, 2018). Le métabolisme du DCM par fermentation en acétate permet la croissance bactérienne mais n’est trouvé qu’en anaérobie (Fig. 2 ; réaction n°7 et 8 ; (Mägli *et al.*, 1998)). Les voies biochimiques de la fermentation sont encore peu connues néanmoins, il a été déterminé que le DCM est utilisé comme source d’énergie et qu’une partie du DCM est assimilée comme source de carbone (en complément du CO_2). Aucun gène intervenant dans ce métabolisme n’a été identifié malgré le séquençage du génome de plusieurs souches capables de fermentation (Kleindienst *et al.*, 2016 ; Chen *et al.*, 2017). De plus des données de chimie isotopique indiquent que la fermentation du DCM pourrait correspondre à des voies métaboliques différentes entre les souches *Dehalobacterium formicoaceticum* et *Candidatus* Dichloromethanomonas elyunquensis (Chen *et al.*, 2018). Au contraire, l'utilisation du DCM en aérobie comme seule source de carbone et d'électrons est bien connue chez les bactéries possédant le gène *dcmA* (Fig. 2 ; réaction n°5) (Muller *et al.*, 2011a). L'enzyme essentielle à l'utilisation du DCM en aérobie est la protéine DcmA qui fait partie de la famille des glutathion S-transférases (GST). Elle catalyse la réaction de déchloration

d'une mole de DCM en deux moles de protons (H^+) et d'ion chlorure (Cl^-), et en une mole de formaldéhyde qui est un intermédiaire central du métabolisme méthylotrophe via le cycle de la sérine (Ochsner *et al.*, 2015).

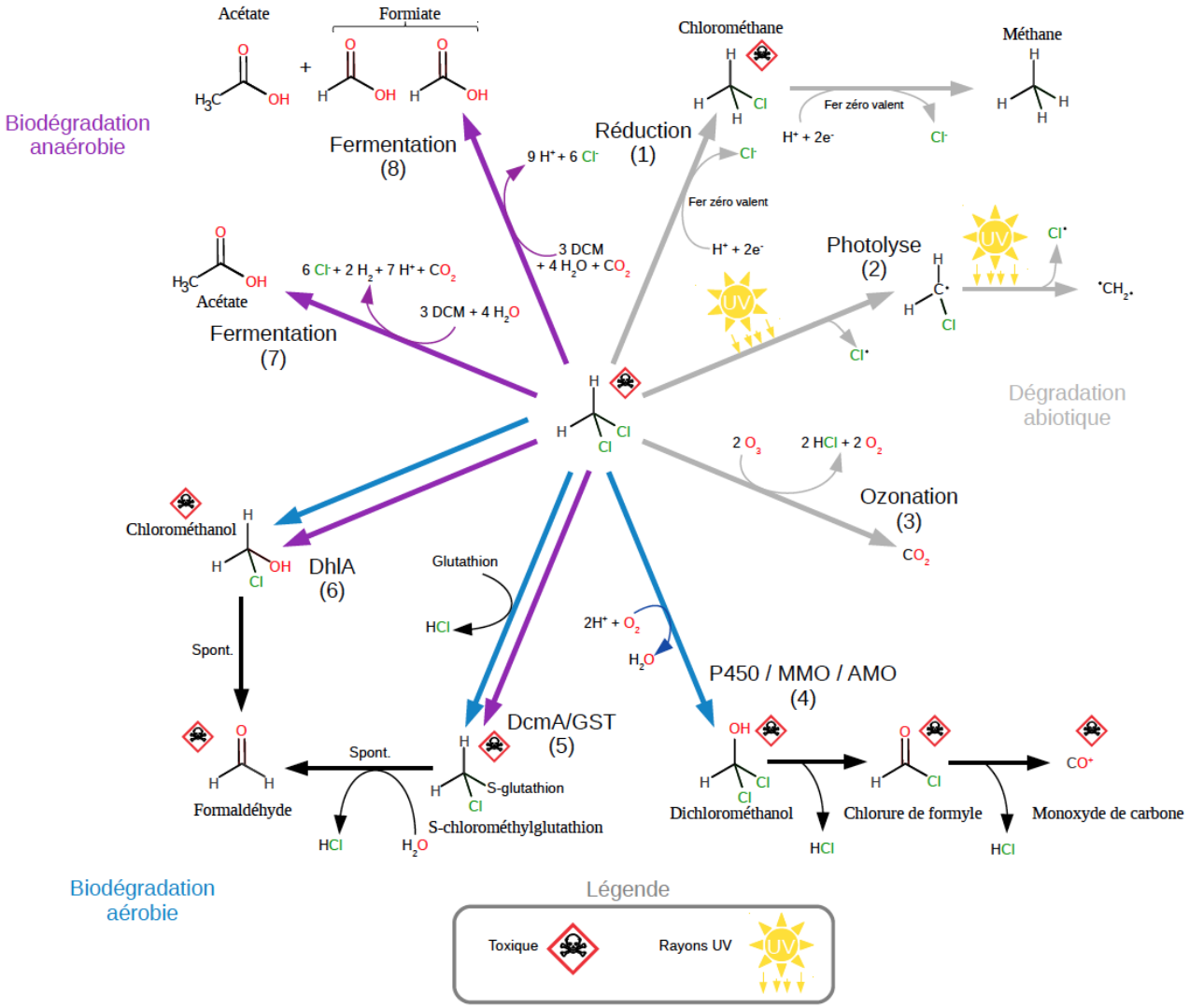


Figure 1.2: Dégradation biologique et abiotique du dichlorométhane.
Plusieurs méthodes permettent de dégrader abiotiquement le DCM (flèches grises) à des échelles industrielles (Shestakova & Sillanpää, 2013). La dégradation de façon biologique implique plusieurs voies métaboliques telles que la dégradation par les enzymes de type haloalkane déshalogenase comme DcmA/GST et DhlA soit en condition aérobie (flèche violette) ou anaérobie (flèche bleue). D'autres enzymes peuvent catalyser l'oxydation du DCM comme la méthane / ammonium mono-oxygenase (MMO/AMO) uniquement en présence d'oxygène tandis que la fermentation du DCM en acétate n'a lieu qu'en anaérobie

Chez *Methylorubrum extorquens* DM4, la souche modèle du métabolisme du DCM en aérobie, les gènes d'utilisation du DCM sont portés par un transposon catabolique possédant quatre gènes *dcm* encadrés par des séquences d'insertion (Fig. 3). Ces gènes codent une déshalogénase (DcmA), un facteur de transcription (DcmR) et deux protéines de fonction inconnue (DcmB et DcmC). Ce transposon est lui-même localisé sur un îlot génomique d'une taille de 126 kb (Vuilleumier *et al.*, 2009). Cet îlot présente un pourcentage en GC de 60 % qui diffère des 68 % du reste du génome de *M. extorquens* DM4. Ce transposon catabolique a été trouvé dans le génome de plusieurs bactéries méthylotrophes (Fig. 3), laissant supposer qu’un transfert horizontal est impliqué dans sa distribution dans l’environnement (Muller *et al.*, 2011a).

Deux groupes de gènes ont été nommés récemment *dcmAE* et *dcmB1B2CD* chez la souche *Rhodococcus* sp. JT-3 mais sans être reliés au métabolisme du DCM (Zhang *et al.*, 2018). En effet , ces gènes interviennent dans le métabolisme d’un herbicide chloré, le diclofop-méthyle, qui possède les mêmes initiales que le dichlorométhane. Chez cette souche le gène *dcmA* code également une déshalogénase de la famille des GST mais celle-ci n’est pas reliée à celle de *M. extorquens* DM4 mais plutôt à *linD* qui intervient dans le métabolisme du lindane via un mécanisme réactionnel différent (classe *eta* des GST ; (Li *et al.*, 2018)).

Le gène *dcmA* code la protéine dichlorométhane déshalogénase DcmA responsable de la déshalogénation du DCM (Roche & Leisinger, 1990). Cette enzyme de la famille des glutathion S-transférase est localisée dans le cytoplasme (Schmid-Appert, 1996). Elle est spécifique des dihalométhanes, elle ne réagit pas avec le 1-chloro-2,4-dinitrobenzène, un substrat typique des GST et n’est pas retenue sur une colonne avec du glutathion contrairement aux autres GST (Vuilleumier *et al.*, 2001 ; Mashiyama *et al.*, 2014). La protéine DcmA représente environ 25 % de la quantité des extraits cellulaires de culture obtenues en présence de DCM (Muller *et al.*, 2011b). Chez un mutant ne possédant plus le transposon *dcm*, l'ajout de *dcmA* exprimé à partir d’un plasmide permet de restaurer le phénotype sauvage de dégradation du DCM (La Roche & Leisinger, 1991) mais l'expression de *dcmA* est constitutive. Par contre, l'ajout du gène *dcmR* en plus du gène *dcmA* restaure la régulation du gène *dcmA* à savoir une expression du gène *dcmA* seulement en présence de DCM. Au vu de ces résultats, DcmR a été proposée comme inhibiteur de l'expression de *dcmA* par La Roche et Leisinger en 1991 lorsque le DCM n’est pas présents. Les gènes *dcmB* et *dcmC* en aval de *dcmA* ont des fonctions inconnues. Des analyses bioinformatiques de la séquence nucléotidique ou protéique n’ont pas permis de trouver de domaines connus chez ces gènes ni de

prédire leurs fonctions. Un mutant du gène *dcmB* par un mini-transposon est toujours capable de se développer en présence de DCM bien que moins rapidement que la souche sauvage, et présente une modification de la morphologie cellulaire après culture sur le DCM (Muller *et al.*, 2011).

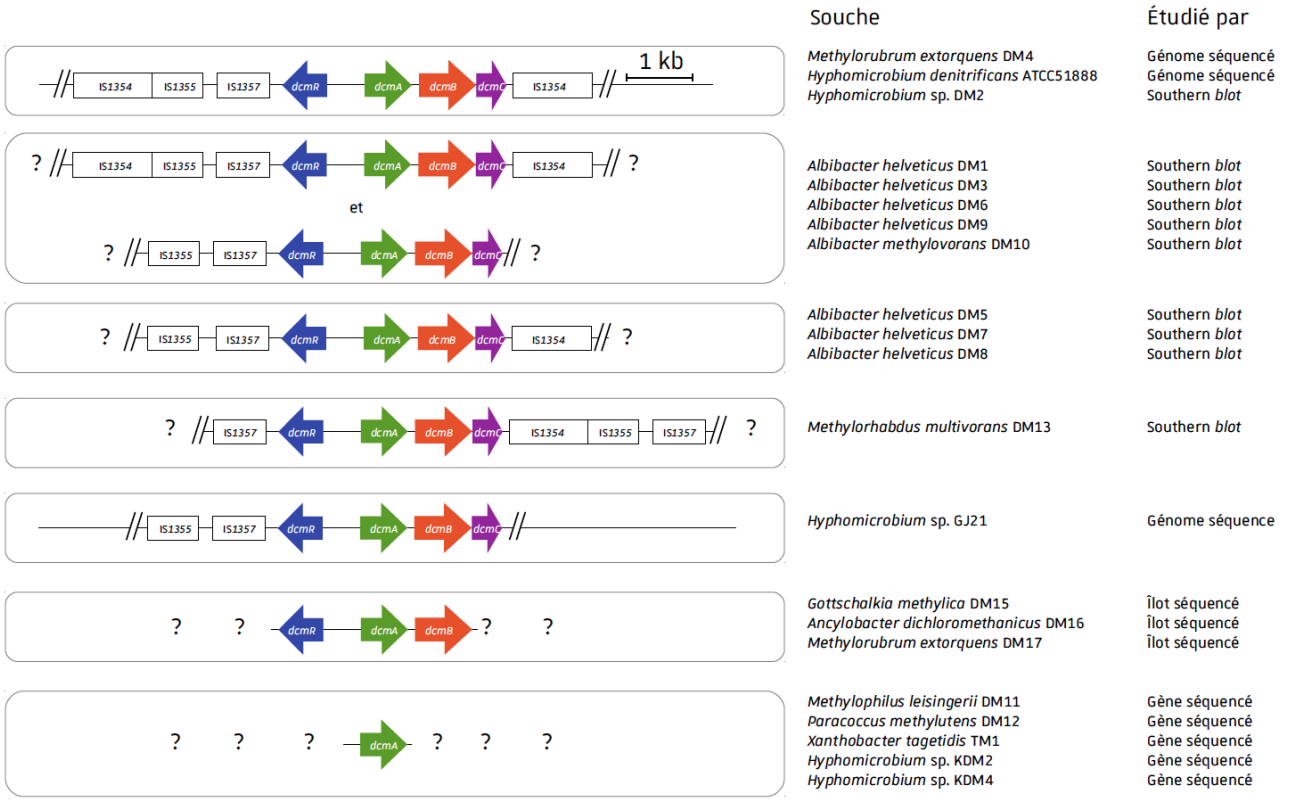


Figure 1.3: Conservation de l'organisation des gènes *dcmRABC* chez les bactéries possédant le transposon catabolique *dcm*.
Le transposon catabolique *dcm* est présent en un ou deux exemplaire(s) chez plusieurs souches bactériennes (Muller *et al.*, 2011a). Dans certains cas la séquence est connue (Firsova *et al.*, 2010 ; Brown *et al.*, 2011 ; Bringel *et al.*, 2017) dans d'autres il s'agit de données de Southern *blot* indiquant juste la présence du transposon *dcm* (Schmid-Appert *et al.*, 1997). La séquence et la conservation des gènes *dcm* sont très conservées mais les séquences d'insertion (IS) qui les entourent sont variables.

La souche modèle, *Methylorubrum extorquens* DM4 a été isolée d'un sol pollué par des alcanes chlorés sur un site industriel et initialement nommée *Pseudomonas* sp. DM4 (Gälli & Leisinger, 1985). Elle a ensuite été renommée *Methylobacterium* sp. DM4 (Gälli & Leisinger, 1988), *Methylobacterium dichloromethanicum* DM4 (Doronina *et al.*, 2000) puis *Methylobacterium extorquens* DM4 (Kato *et al.*, 2005) et récemment renommée en *Methylorubrum extorquens* DM4 (Green & Ardley, 2018). Cette bactérie de la classe des *Alphaproteobacteria* est un bacille à coloration de Gram négative, pigmenté en rose par un caroténoïde. Le génome de *M. extorquens*

DM4 a été séquencé (Vuilleumier *et al.*, 2009). Il est composé d'un chromosome de 5,94 Mb et de deux plasmides (38 et 141 kb) de fonction inconnue. *M. extorquens* DM4 se développe, en aérobie stricte, en utilisant des molécules organiques comme source de carbone, d'énergie et d'électrons (chimio-organohétérotrophe). *M. extorquens* DM4 est dit méthylotrophe facultatif, car il peut aussi utiliser des molécules à plusieurs atomes de carbone comme l'acétate ($C_2H_4O_2$), le tartrate ($C_4H_6O_6$) et le succinate ($C_4H_4O_4$) en plus des composés en C_1 tels que le DCM (CH_2Cl_2), le méthanol (CH_3OH) et la méthylamine (CH_3NH_2) (Gruffaz *et al.*, 2014) (Fig. 4). *M. extorquens* DM4 est également capable de métaboliser le L-tartrate (Doronina *et al.*, 2000). Les propriétés optiques (polarisation) de cette molécule ont été étudiées par Louis Pasteur lorsqu'il travailla à l'université de Strasbourg (Pasteur, 1848).

1.3 Mécanismes d'adaptation à la croissance avec le DCM

Le métabolisme du DCM par *M. extorquens* DM4 conduit à différents stress qui ont été identifiés *in vitro* et *in vivo* par différentes approches dont une synthèse est donnée ci-après. Pour certaines bactéries posséder le gène *dcmA* n'est pas suffisant pour permettre une bonne croissance à partir du DCM comme seule source de carbone et d'énergie. Par exemple, la croissance d'une souche de *E. coli* exprimant *dcmA* est rapidement inhibée par la production de S-chlorométhylglutathion (Evans *et al.*, 2000). Le gène *dcmA* a également été exprimé à partir d'un plasmide *en trans* chez plusieurs souches de *Methylorubrum* spp. via un plasmide (Kayser *et al.*, 2002 ; Michener *et al.*, 2014b) pour tester si des souches phylogénétiquement proches de DM4 pouvaient dégrader ce polluant. Certaines de ces souches comme *M. extorquens* CM4 sont capables de se développer avec le DCM en présence du plasmide portant *dcmA* tandis que d'autres non, comme *M. extorquens* AM1. Cela indique que la dégradation du DCM ne repose pas seulement sur le gène *dcmA* mais aussi sur la présence d'autres gènes nécessaires pour l'adaptation à ce métabolisme toxique, en impliquant différentes fonctions cellulaires de réponse face à des stress liés à la croissance sur le DCM.

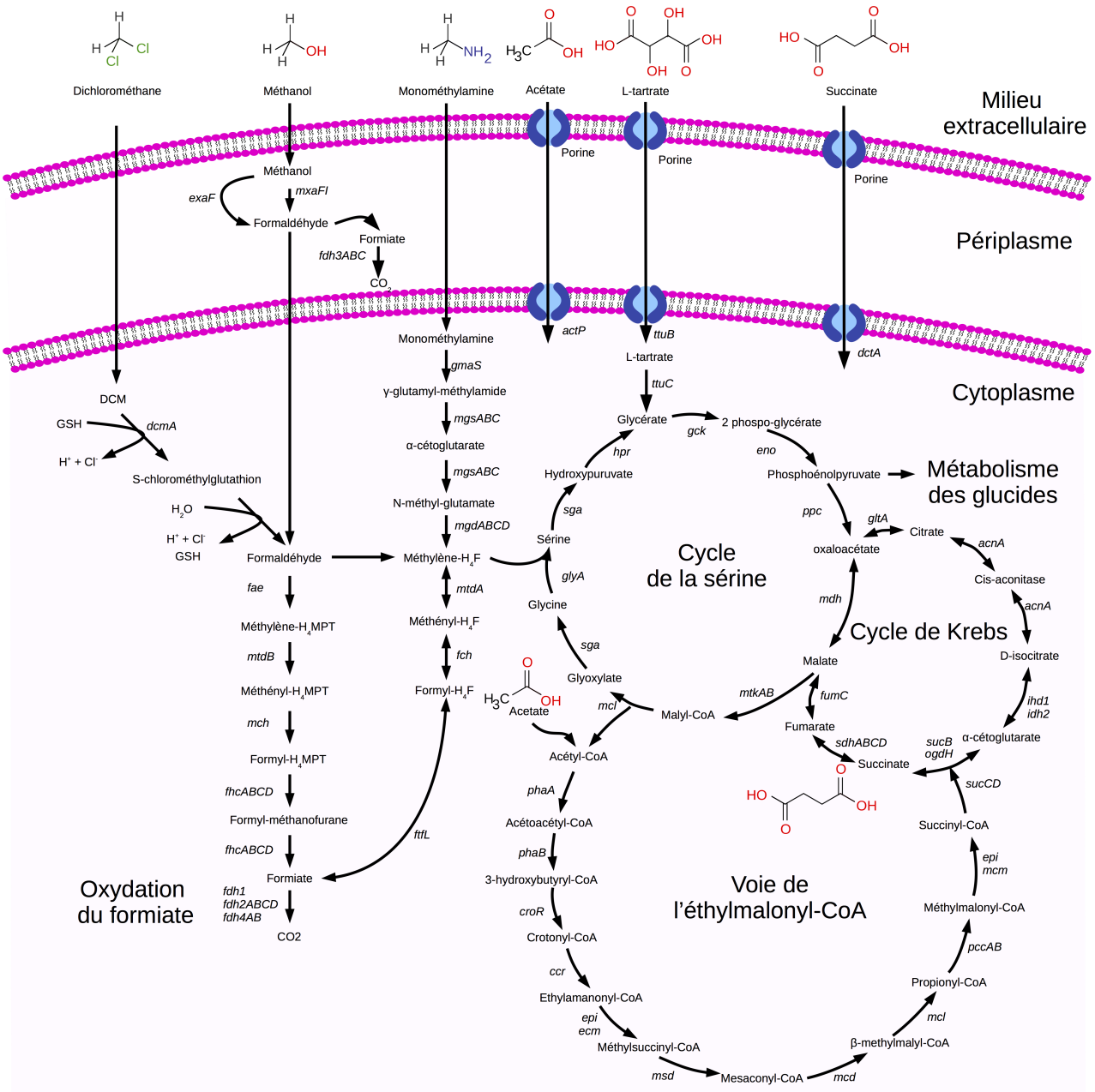


Figure 1.4: Substrats carbonés métabolisés par *M. extorquens* DM4 et voies métaboliques associées. *M. extorquens* DM4 est capable de métaboliser à la fois des substrats organiques réduits sans liaisons carbone-carbone dits « methylotrophe » et des composés avec des liaisons carbone-carbone (Doronina *et al.*, 2000). Le nom des gènes est indiqué à côté des réactions catalysées par les enzymes qu'ils codent. Les abréviations H_4MPT et H_4F correspondent respectivement aux cofacteurs tétrahydrométhanoptérine et tétrahydrofolate. L'assimilation du carbone du DCM, méthanol et monométhylamine converge au niveau du methylène-H_4F pour l'entrée dans le cycle méthylotrophe de la sérine (Ochsner *et al.*, 2015). Certains substrats tels que le succinate peuvent directement être assimilés dans le cycle de Krebs. Certains composés tel que le DCM peuvent rentrer dans le cytoplasme par diffusion tandis que d'autres passent par des transporteurs membranaires (représentés en bleu sur la figure) tel que *dctA* pour le succinate (Van Dien, 2003).

1.3.1 Adduits à l'ADN

Au cours de la première étape de dégradation du DCM, la déshalogénation du DCM catalysée par DcmA produit deux métabolites génotoxiques : le formaldéhyde et le S-chlorométhylglutathion. Ils peuvent former des adduits à l'ADN par alkylation des atomes d'azote présents dans les nucléotides (Chen *et al.*, 2016). Des expériences ont été réalisées *in vitro* avec l'enzyme DcmA de *Methylophilus leisingerii* DM11 et des extraits cellulaires additionnés de DCM (^{14}C) ou glutathion (^{35}S) marqués avec des isotopes radioactifs pour visualiser respectivement les adduits du formaldéhyde ou du S-chlorométhylglutathion (Kayser & Vuilleumier, 2001). Les adduits à l'ADN sont principalement trouvés au niveau de la base guanine et sont marqués par du ^{35}S. Cela signifie que ces adduits sont dus au S-chlorométhylglutathion et non pas au formaldéhyde. L'ajout d'une GST humaine (*theta*-1) chez *Salmonella enterica* serovar Typhimurium conduit également à des adduits au niveau des guanines lors du métabolisme du DCM (DeMarini *et al.*, 1997). En fonction de leur taille et de leur localisation, les alkylations peuvent provoquer des mutations par mésappariement de bases ou bloquer l'ADN et l'ARN polymérase ce qui conduit à des ruptures des brins d'ADN (Fig. 5A) (Shrivastav *et al.*, 2010).

Une approche de mutagenèse aléatoire par mini-transposon Tn*5*, chez *M. extorquens* DM4, a permis d'identifier deux gènes de réparation de l'ADN comme ayant un impact sur la croissance en présence de DCM (Kayser, 2001). Le gène *uvrA* impliqué dans la reconnaissance de dommages à l'ADN et la réparation par excision de nucléotide n'est pas essentiel à la croissance avec le DCM (Kayser, 2001). Par contre, le mutant *uvrA*::miniTn*5* a une croissance ralentie dans cette condition de culture. Tandis que, lorsque le gène *polA* qui code l'ADN polymérase I est muté par insertion d'un mini-transposon Tn*5* à l'intérieur du gène dans le domaine exonucléase 3'→5' (Fig. 5B), *M. extorquens* DM4 n'est plus capable de se développer avec le DCM (Kayser *et al.*, 2000). Dans cette mutation, la partie du gène *polA* en amont de l'insertion serait exprimée et coderait seulement la partie de la protéine possédant l'activité 5'→3' exonucléase impliquée dans l'élimination des amorces d'ARN utilisées lors de la réplication. L'enzyme mutée ne possède plus son activité de polymérase (Kayser *et al.*, 2000). Après croissance avec le DCM, l'ADN extrait du mutant *polA*::miniTn*5* est dégradé comme suggéré par la présence d'un « *smear* » après électrophorèse sur gel d'agarose contrairement à l'ADN extrait d'une culture en méthanol. Cette dégradation pourrait

correspondre à des cassures double-brin de l'ADN (Kayser & Vuilleumier, 2001). Pour quantifier ces dommages à l'ADN extrait de *M. extorquens* DM4 a servi de substrat *in vitro* à l'ADN polymérase I (fragment de Klenow ; Fig. 5B), en présence des dNTPs et de l'ATP radioactif (^{32}P). L'incorporation d'ATP radioactif dans l'ADN extrait de cultures avec du DCM est plus importante que pour de l'ADN extrait de cultures en méthanol ce qui suggère la présence accrue de lésions à l'ADN après croissance avec le DCM (Firsova *et al.*, 2004). Les mécanismes de réparation des dommages à l'ADN suite à la déshalogénation du DCM n'ont pas encore été élucidés.

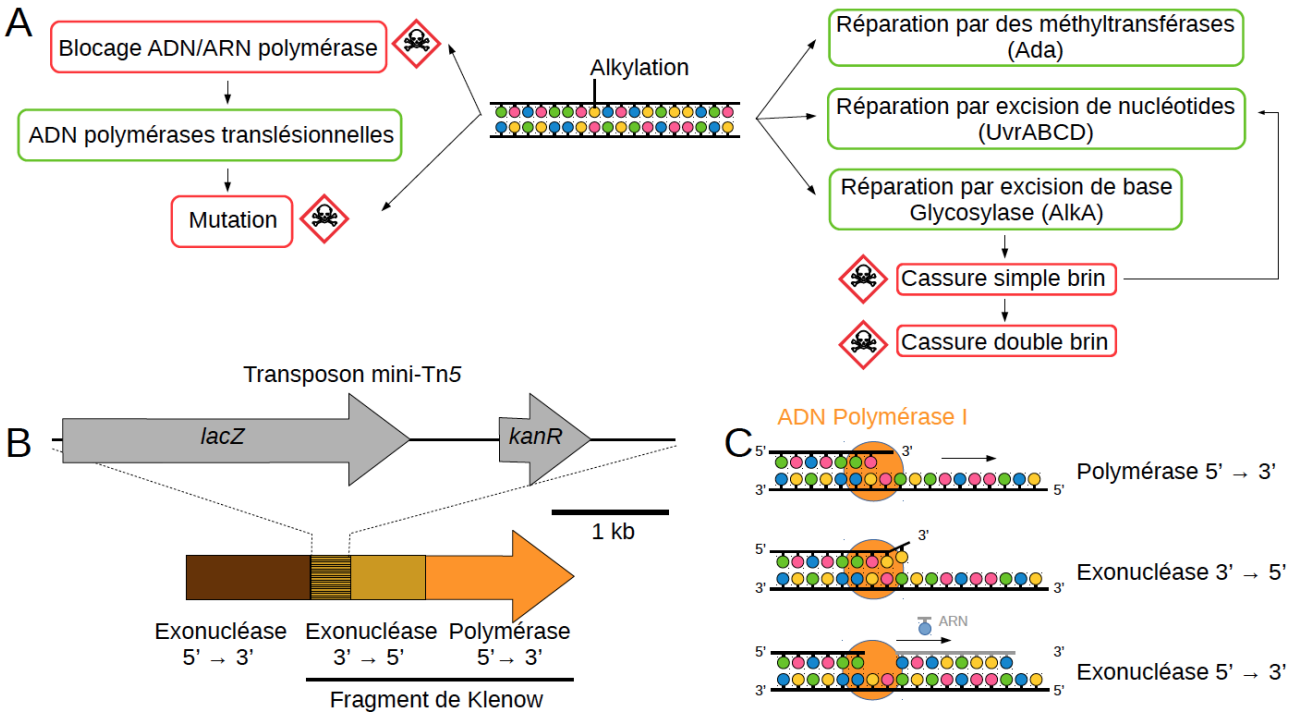


Figure 1.5: Alkylation de l'ADN et réparation via l'ADN polymérase I.
A Impacts des alkylations sur l'ADN et voies de réparations connues chez *E. coli* (Mielecki and Grzesiuk, 2014). Les alkylations peuvent conduire à la formation de mutations ou à des cassures doubles brins (Shee *et al.*, 2013). Elles peuvent également être réparées via la voie d'excision des nucléotides faisant appel à UvrABCD. La délétion de bases alkylées trop proches dans le génome peut conduire à la formation de cassure double brin. B Représentation du gène *polA* chez le mutant *polA*::miniTn*5* adapté de (Kayser, 2001). Une protéine avec juste l'activité 5' → 3' serait produite. C Différentes fonctions de l'ADN polymérase I. Elle intervient notamment pour synthétiser de courts fragments d'ADN après réparation par excision de nucléotides. Son activité exonucléase 3'→ 5' lui permet d'éliminer un nucléotide mal incorporé. Son activité exonucléase 5' → 3' lui permet d'éliminer les amorces d'ARN utilisées lors de la réplication de l'ADN.

1.3.2 Libération intracellulaire de chlorure

Pour chaque molécule de DCM dégradée par DcmA, deux ions chlorure et deux protons sont libérés dans le cytoplasme. La libération d'ion chlorure et de proton représente un double stress : osmotique et baisse du pH auquel une bactérie dégradant le DCM doit faire face. Lors de culture, en présence de DCM, des souches d' *alpha-proteobacteria* (*Albibacter helveticus* DM6, *A. methylovorans* DM10 et *M. extorquens* DM4), l'utilisation d'agents dissipant le gradient de protons ou inhibant l'ATPase montre que la sécrétion de chlorure est dépendante du gradient de proton à travers la membrane interne (Torgonskaya *et al.*, 2011). Des structures de nature inconnue contenant du chlorure et de phosphore ont été observées par microscopie électronique et analyse au rayons X, à la surface d'*A. helveticus* DM6 et *A. methylovorans* DM10 lors de la dégradation du DCM (Torgonskaya *et al.*, 2011). Ces structures de rôle inconnu n'ont pas été détectées chez *M. extorquens* DM4. Il est supposé qu'il s'agirait d'un mécanisme d'adaptation à ce stress. Par exemple par export du chlorure en dehors de la cellule ou pour fournir une densité de charge permettant de modifier la perméabilité membranaire (Torgonskaya *et al.*, 2011).

Les ions chlorure libérés dans le cytoplasme peuvent également être exportés hors de la cellule par des transporteurs membranaires. Un groupe de gènes, nommés $\underline{Cl^-}$ *induced* (*cliABC*), a été détecté chez *M. extorquens* DM4 par une approche de mutagenèse aléatoire (Muller *et al.*, 2011b). Ces gènes sont plus exprimés en présence de chlorure et codent un transporteur membranaire de la famille RND (*Resistance Nodulation cell Division*). Ces gènes ne sont pas essentiels à la croissance en présence de DCM (Muller, 2011). La délétion des gènes *cliABC* ralentit la croissance en présence de DCM mais pas en présence de méthanol. Ils sont régulés par le gène *cliR*, un facteur de transcription dont le gène est adjacent à *cliABC* et orienté de façon divergente. La délétion de *cliR* améliore la vitesse de croissance en présence de DCM. Cela peut s'expliquer par le rôle putatif de *cliR* comme répresseur de *cliABC :* sa délétion permet donc de produire d'avantage de ce transporteur (Muller, 2011). Curieusement, des résultats ultérieurs par une approche d'évolution expérimentale n'ont pas permis de valider l'implication du transporteur CliABC dans la réponse au stress chlorure (Michener *et al.*, 2014b). En effet, lorsque des souches de *Methylorubrum* pour lesquelles un plasmide exprimant *dcmA* a été introduit et ont pu évoluer au cours de 100 générations, des mutations ciblant un autre transporteur ont été identifiées comme conférant un trait d'adaptation à la croissance sur le DCM (Michener *et al.*, 2014b). Des mutations indépendantes au niveau du gène *clcA* codant un transporteur putatif d'échange H^+/Cl^- ont été trouvées chez plusieurs

clones évolués. L'apport de *clcA* en plus du gène *dcmA* sur un plasmide permet de faciliter la croissance en présence de DCM. D'autres mutations ont aussi été identifiées dans le gène *secY* codant une protéine qui intervient dans la translocation de protéines, avec un peptide signal, à travers la membrane interne (Saraogi & Shan, 2014). Il est supposé que des mutations dans ce gène permettraient d'agrandir le canal par lequel passent les protéines transloquées à travers la membrane. Cet agrandissement favoriserait également le passage d'ion chlorure, en même temps que les protéines, pour l'éliminer. Cela indique que la libération intracellulaire de chlorure constitue bien un stress lié au métabolisme du DCM.

1.3.3 Effet solvant et stress au niveau de la membrane

Le DCM est un composé lipophile, qui perturbe les membranes cellulaires par un effet dit « solvant » défini comme une perturbation de la membrane cytoplasmique par insertion de composés hydrophobes (Mukhopadhyay, 2015). Cette perturbation en modifiant la perméabilité de la membrane conduit à une perte de proton et donc diminue le gradient transmembranaire de proton. Alors que le DCM pourrait rentrer dans la cellule par diffusion, la cinétique de dégradation du DCM suggère plutôt un transport actif chez *M. extorquens* DM4 (Gisi *et al.*, 1998 ; Torgonskaya *et al.*, 2018). Le ou les gène(s) responsable(s) de ce transport ne sont pas encore connus. Il est peu probable que les protéines DcmB ou DcmC soient impliqués dans le transport car pour ces deux protéines aucun peptide signal ou domaine transmembranaire n'est prédits.

Pour s'adapter au stress engendré par l'effet solvant, différents mécanismes ont été décrits chez les bactéries (Segura *et al.*, 2012). La production de chaperon permet stabiliser des protéines tandis que la production de transporteurs membranaire intervient dans l'export des composants à effets solvant (Mukhopadhyay, 2015). Un autre mécanisme adaptatif consiste à modifier la composition de la membrane pour la rendre plus rigide et donc réduire les effets d'une exposition à un solvant. Par exemple, les hopanoïdes sont des composés organiques retrouvés dans la membrane lipidique de certaines bactéries et permettent de stabiliser la membrane et de diminuer sa perméabilité (Siedenburg & Jendrossek, 2011). Une mutation du gène *shc* (*squalene-hopene cyclases*), impliqué dans la synthèse d'hopanoïdes (Siedenburg & Jendrossek, 2011), empêche la croissance en présence de DCM (Muller *et al.*, 2011b ; Bradley *et al.*, 2017).

1.3.4 Utilisation du glutathion

Le glutathion est utilisé comme le cofacteur de l'enzyme DcmA mais également d'autres enzymes comme les GST ou GloAB (Deponte, 2013). Il possède d'autres fonctions cellulaires telle que maintenir le cytoplasme dans un état réduit (Sporer *et al.*, 2017). L'utilisation du glutathion par DcmA pourrait représenter un stress oxydatif, indirect, pour la cellule en réduisant le stock de glutathion disponible. Une quantification du glutathion indique que la quantité totale reste stable mais que le ratio glutathion oxydé (GS-SG ; glutathion disulfure) ou réduit (GSH) change vers un état plus oxydé lors du métabolisme du DCM (Torgonskaya *et al.*, 2018). À la fin de la réaction de déshalogénation catalysée par DcmA, le glutathion est recyclé sous sa forme de départ (glutathion réduit) (Stourman *et al.*, 2003 ; Mashiyama *et al.*, 2014). Il est même supposé qu'une seule molécule de glutathion puisse intervenir dans plusieurs réactions consécutives (Stourman *et al.*, 2003). Aussi, la modification de l'état d'oxydation du glutathion pourrait être due à une modification de l'état énergétique de la cellule. L'activité de la glutathion réductase, qui réduit le glutathion par une réaction faisant intervenir du NADPH, est également plus importante dans des cellules cultivées avec du DCM que celles avec du méthanol (Torgonskaya *et al.*, 2018). Cette augmentation de l'activité enzymatique permettrait de réduire le glutathion oxydé.

1.4 Régulation de la transcription de l'îlot *dcm* via DcmR

En plus des mécanismes d'adaptation aux stress causés par le DCM, les souches dégradantes doivent également produire la protéine DcmA en quantité suffisante pour permettre la déchloration de ce polluant. Des expériences ont montré qu'en présence de DCM, l'activité enzymatique d'extrait cellulaire de *M. extorquens* DM4 était plus importante que pour des cellules cultivées avec du méthanol (Roche & Leisinger, 1990). Il est supposé que cette régulation passerait par le gène *dcmR* présent dans l'îlot *dcm*. Le gène *dcmR* code une protéine qui serait composée de deux domaines (Fig. 6C). Il y a présence, en N terminal, d'un domaine hélice-tour-hélice (HTH) de la famille MerR qui lui permettrait de se fixer à l'ADN (La Roche & Leisinger, 1991). Il est supposé que ce domaine reconnaîtrait une séquence conservée de 12 pb localisée en amont des gènes *dcmA* et *dcmR* (de -14 à -25 par rapport au +1 de la transcription ; entre les boites -10 et -35 des promoteurs de *dcmR* et *dcmA)*. La protéine DcmR empêcherait ainsi la reconnaissance du promoteur par l'ARN polymérase. Néanmoins cette séquence d'ADN, qui n'a pas été validée expérimentalement, ne correspond pas à celles généralement reconnues par les facteurs de

transcriptions qui sont inversées et répétées. La Roche et Leisinger supposaient que ce site pourrait être reconnu par un facteur de transcription sous forme de monomère et non pas de dimère comme le cas classique (La Roche & Leisinger, 1991). Le facteur de transcription MerR qui possède un domaine HTH similaire à DcmR est connu pour se fixer entre les boîtes -10 et -35 des promoteurs qu'il régule (Hobman *et al.*, 2005 ; Couñago *et al.*, 2016). Cela fait sens avec l'hypothèse que DcmR se fixerait entre les boîtes -10 et -35 de *dcmA* et *dcmR* même si la séquence consensus ne correspond pas à celle reconnue par MerR. Un second domaine, en C terminal, a été prédit par bioinformatique mais n'est pas encore vérifié expérimentalement (Anantharaman & Aravind, 2005). Ce second domaine appelé *methanogen / methylotroph, DcmR sensory* (MEDS) est trouvé dans des protéines de plusieurs bactéries méthylotrophes ou méthanogènes. Les protéines portant ce domaine possèdent aussi des domaines de fixation à l'ADN ou impliqués dans le transfert de signaux cellulaires tels que l'histidine kinase et facteur anti-sigma (Fig. 6A). Il a donc été prédit que ce domaine pouvait fixer de petites molécules hydrocarbonées intervenant dans le métabolisme des composés en C_1. Cela permettrait de modifier la conformation de la protéine et ainsi permettre une régulation. Ce domaine protéique a aussi été trouvé chez d'autres bactéries telles que *Mycobacterium, Nocardia* et *Streptomyces* (Anantharaman & Aravind, 2005) ainsi qu'au niveau de systèmes à deux composants chez des archées (Galperin *et al.*, 2017). Toutes ces recherches sont basées sur l'étude de séquences et aucune donnée expérimentale n'est disponible.

Des protéines avec une composition similaire à DcmR sont trouvées chez plusieurs autres bactéries mais en étant peu conservées par rapport à *M. extorquens* DM4 (Fig. 6). Ces protéines sont trouvées notamment chez plusieurs *Nitrosomonas* (*β-proteobacteria*) avec 43 à 46 % d'identité au niveau protéiques avec DcmR de *M. extorquens* DM4 : *Nitrosomonas* ; *Nitrosomonas aestuarii* NM36 ; *N. cryotolerans* ATCC 49181 et *N. marina* NM71. Ces bactéries sont capables de dégrader le DCM par co-métabolisme avec l'enzyme ammonium mono-oxygénase (qui est homologue de l'enzyme méthane mono-oxygénase) (Vannelli *et al.*, 1990). Néanmoins ces gènes ne sont pas groupés ensemble dans le génome. Le gène *dcmA* est absent de ces génomes. Ces données de génomique n'indiquent pas forcément un lien entre ce co-métabolisme et ces protéines similaires à DcmR.

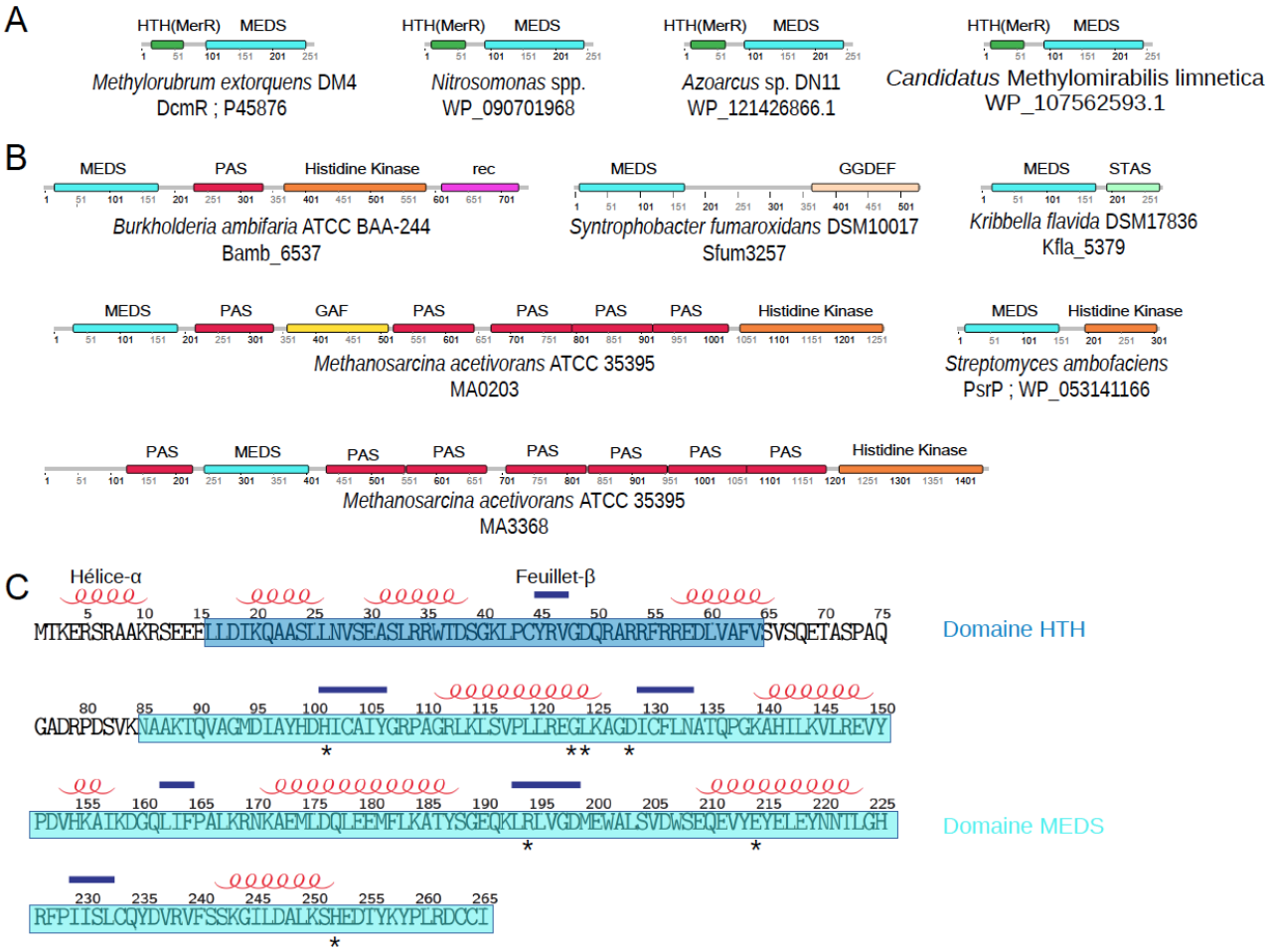


Figure 1.6: Présence du domaine MEDS dans de nombreuses protéines de signalisation cellulaire. Pour chaque exemple la référence Uniprot de la protéine est indiquée sous le nom de l'organisme. Le domaine MEDS peut être trouvé avec un domaine HTH comme pour *dcmR* (panneau A) ou avec un autre domaine de régulation (panneau B). Lorsque le domaine HTH est absent dans la protéine portant le domaine MEDS, les autres domaines sont : STAS (*Sulphate Transporter and AntiSigma factor antagonist* (Sharma *et al*., 2011)), GGDEF (nommé en fonction du motif conservé GG[DE][DE]F (Schirmer, 2016)), rec (domaine receveur de phosphorylation), GAF (nommé en fonction des protéines cGMP-regulated cyclic nucleotide PDEs, adenylyl cyclases et FhlA (Schirmer, 2016)), PAS (nommé en fonction des protéines Per, Arnt, Sim où il a été trouvé (Schirmer, 2016)). C Structure primaire et secondaire de DcmR. La structure secondaire est prédite par le logiciel PSIPRED 3.3 (Buchan *et al.*, 2013). Les acides aminés conservés dans le domaine MEDS sont indiqués par des étoiles (Anantharaman & Aravind, 2005).

Précédemment, pour comprendre comment se passe la régulation de ce transposon catabolique, plusieurs approches ont été tentées. Une étude par nucléase S1 a permis d'observer quatre ARNm transcrits pour *dcmR* chez une souche de *M. extorquens* DM4 sans le transposon catabolique *dcm* mais avec les gènes *dcmA* et *dcmR* sur un plasmide (La Roche & Leisinger, 1991). Cela laisse supposer la présence de quatre sites d'initiation de la transcription pour *dcmR* ce qui pourrait permettre une régulation de ce gène. Parmi ces quatre sites, deux (P_{R3} et P_{R4}) ont été détectés

uniquement en culture avec du DCM tandis que les autres (P_{R1} et P_{R2}) sont détectés à la fois en présence de méthanol ou DCM. Une autre hypothèse pourrait être que les quatre ARNm détectés pourraient sinon proviennent d'une maturation de l'ARNm de *dcmR*. Cette même étude a permis de détecter un seul ARNm pour *dcmA* (Fig. 7).

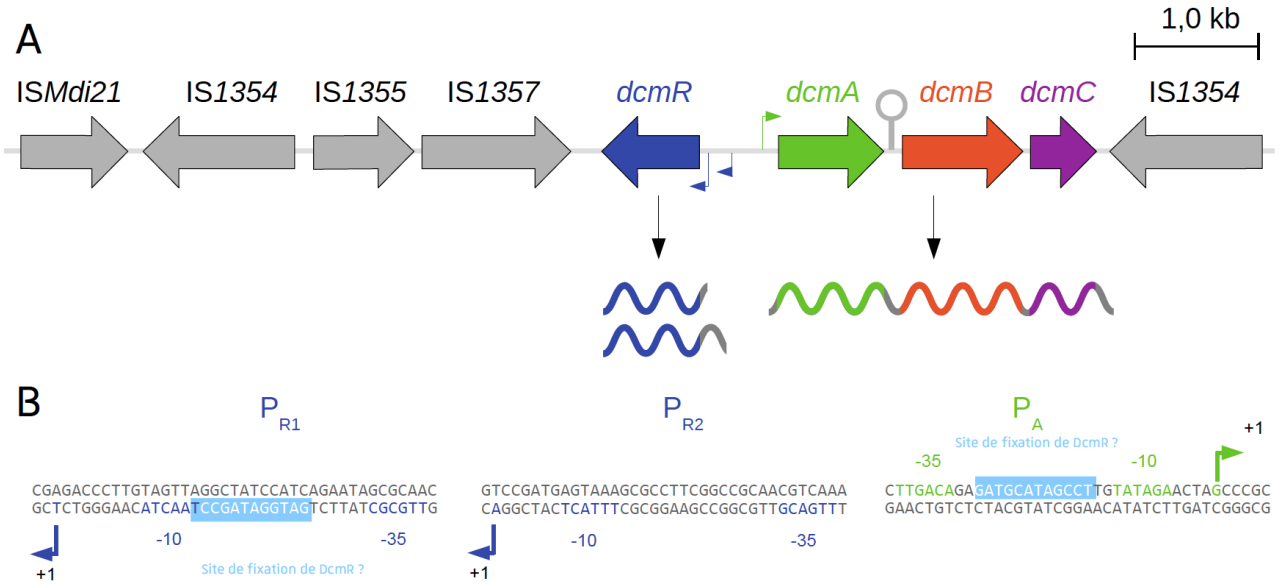


Figure 1.7: Transposon catabolique *dcm* chez *M. extorquens* DM4 et promoteurs de *dcmR* et *dcmA*
A. Organisation des gènes *dcm* et localisation des promoteurs dans la région intergénique *dcmR-dcmA*. B. Des expériences de nucléase S1 couplées à des expériences de complémentation de fonction ont permis de proposer la présence des promoteurs P_{R1} et P_{R2} du gène *dcmR*, avec le promoteur P_{R1} impliqué dans la régulation (La Roche & Leisinger, 1991). Seul le promoteur P_A a été caractérisé pour le gène *dcmA*. Ces trois promoteurs sont détectés dans les conditions de culture méthanol ou DCM. Un terminateur de la transcription est prédit en aval de *dcmA* (Roche & Leisinger, 1990). Il est représenté par une boucle.

La spécificité des composés halogénés capables de moduler la régulation de l'expression de *dcmA* a été étudiée à l'aide d'un plasmide rapporteur (Fig. 8). Un plasmide rapporteur a été construit en couplant un fragment d'ADN contenant les gènes *dcmR* et le promoteur de *dcmA* avec un autre fragment contenant les gènes *luxCDABE,* permettant la production de bioluminescence (Lopes *et al.*, 2012). Dans ce plasmide les gènes *luxCDABE* forment un opéron sous le contrôle du promoteur de *dcmA* ce qui conduit à une bioluminescence proportionnelle à l'expression de *dcmA* (Fig. 8). Une production de bioluminescence est observée chez *M. extorquens* DM4 en présence de DCM mais pas d'autres solvants chlorés comme le tétrachloroéthylène, trichloroéthylène, 1,1,1-trichloroéthane et le chloroforme. Cela indique que la régulation serait spécifique au DCM.

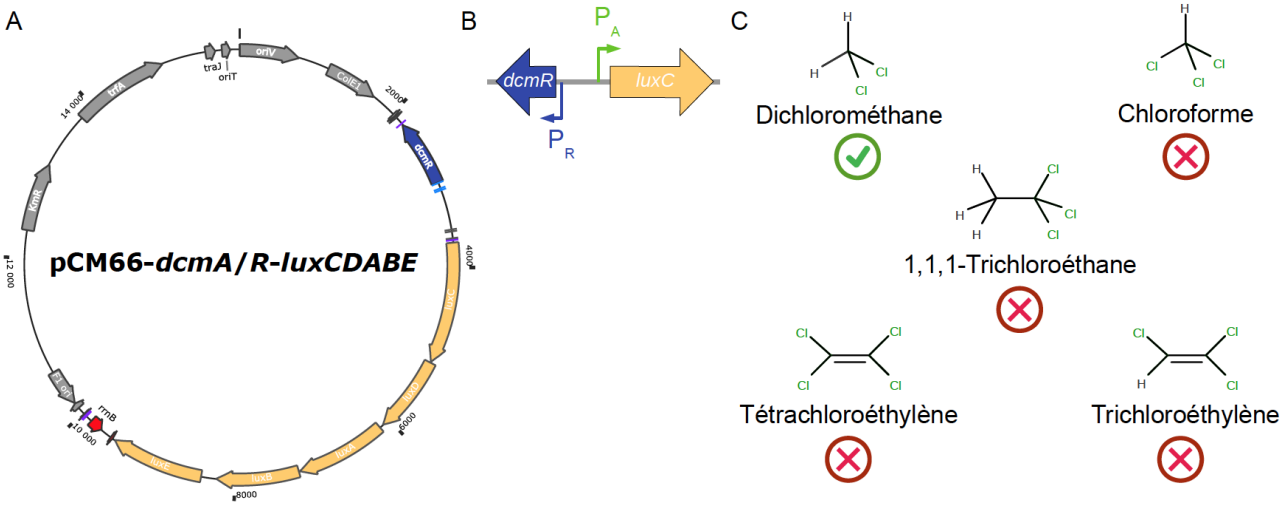


Figure 1.8: Plasmide rapporteur détectant le DCM.
A Carte plasmidique d'après les informations fournies par (Lopes *et al.*, 2012). Le plasmide produit la protéine DcmR sous contrôle de son propre promoteur ainsi que les protéines LuxABCDE sous le contrôle du promoteur de *dcmA* (Lopes *et al.*, 2012). B Représentation du gène *luxC* exprimé via le promoteur de *dcmA*. Il forme un opéron *luxCDABE*. Le gène *dcmR* est aussi présent et exprimé via son propre promoteur. C La réponse de ce plasmide est spécifique au DCM et ne donne pas de résultats en présence d'autres solvant chlorés.

Une étude de protéomique différentielle par électrophorèse bidimensionnelle, chez *M. extorquens* DM4, a permis de détecter la protéine DcmR en condition de culture avec du DCM mais pas en présence de méthanol (Muller *et al.*, 2011b). Ceci indiquerait que la répression de *dcmA* en condition méthanol ne passerait pas via la protéine DcmR, et réfuterait ainsi l'hypothèse de La Roche & Leisinger, 1991. Le mécanisme de régulation de *dcmA* est donc encore incompris à ce jour.

1.5 Hypothèses de travail et objectifs du projet de thèse

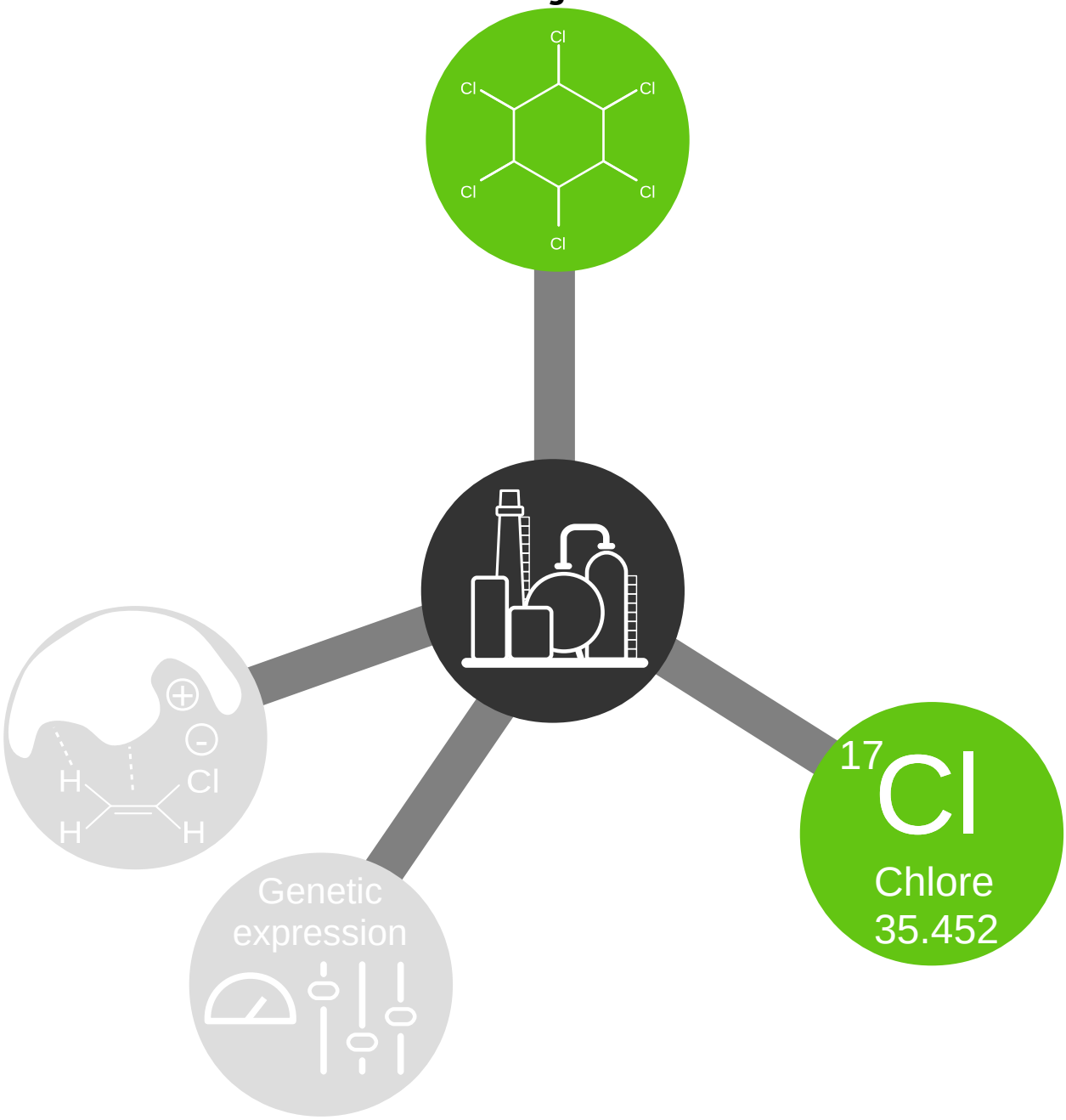
L'ensemble de ces informations a permis d'émettre plusieurs questions par rapport aux zones d'ombres du métabolisme et de l'adaptation au DCM chez *M. extorquens* DM4. À partir de ces questions, des hypothèses et des expériences pour les valider sont proposées.

1/ Comment sont régulés les gènes et protéines en dehors du transposon catabolique *dcm* lors d'exposition au DCM ? Au vu des différents stress provoqués par le métabolisme du DCM, il est supposé que l'adaptation passe par un changement d'expression de gènes situés en dehors du transposon catabolique *dcm*. Une hypothèse serait que l'expression de gènes impliqués dans l'export de chlorure ou de réparation de l'ADN soit plus importante en présence de DCM. Des approches « -omiques » seront utilisées pour étudier la réponse globale au DCM tant par RNAseq que par protéomique afin de comparer le niveau d'expression des gènes *dcm* avec celui d'autres gènes et protéines en dehors de l'îlot *dcm* lors d'exposition au DCM. Peut-on associer la régulation de certains gènes à la présence des motifs en amont du promoteur ou de structures présentes dans les séquences 5' non traduite des ARNm ? Pour cela, des approches détectant les sites d'initiation de la transcription et de la traduction ont été réalisées pour délimiter la région non traduite en 5' des ARNm. Des motifs communs aux gènes, régulés en présence de DCM, pourront être recherchés à ce niveau ainsi qu'à proximité des promoteurs.

2/ Comment *dcmR* régule-t-il l'expression du gène en présence de DCM ? Il a été démontré que la régulation de *dcmA* changeait selon la source de carbone (méthanol ou DCM ; (La Roche & Leisinger, 1991 ; Muller *et al.*, 2011b)). Cela impliquerait une répression par le facteur de transcription *dcmR*. Néanmoins, les mécanismes de régulation sont peu connus. Pour vérifier que cette régulation est due à DcmR, plusieurs mutants de délétion des gènes du transposon catabolique *dcm* ont été construits. Leur phénotype permettra d'observer s'il y a des modifications de l'expression des gènes *dcm* en présence ou absence de ces gènes. Cela permettra de voir si l'activité de DcmR est modulée par un autre gène du transposon catabolique *dcm*. La transcription des gènes *dcmRABC* sera étudiée en condition de culture avec plusieurs sources de carbone servant à la croissance de la souche sauvage et des mutants. L'expression de ces gènes sera étudiée par fusion transcriptionnelle et PCR quantitative après rétro-transcription (RT-qPCR). Cela permettra de confirmer s'il y a bien une répression de la transcription par DcmR. En plus des transcrits, les protéines du transposon *dcm* seront quantifiées par une approche de spectrométrie de masse ce qui permettra de comparer les quantités d'ARNm et de protéines pour observer s'il y a des modifications post-transcriptionnelles. Cette quantification permettra également de savoir si la protéine DcmR est bien produite en absence de DCM.

3/ Le gène *dcmR* est-il conservé dans l’environnement et chez les souches isolées de sites pollués ? Le gène *dcmA* a été détecté dans plusieurs sites pollués mais le gène *dcmR* lui n’a jamais été recherché. Au vu de la conservation du transposon catabolique, on peut supposer que les gènes *dcmA* et *dcmR* soient retrouvés en même quantité dans l’environnement. Une approche de bioinformatique sera utilisée pour rechercher le gène *dcmR* dans d’autres contextes génétiques. En parallèles, le gène *dcmR* sera recherché et quantifié par qPCR à partir d’ADN extrait de sites pollués au DCM.

2. État de l’art sur la régulation des gènes bactériens de déshalogénases



Transcriptional regulation of organohalogen pollutant utilisation in bacteria

Bruno Maucourt, Stéphane Vuilleumier and Françoise Bringel*

Affiliation: Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France

*Correspondence: Françoise Bringel; Email: francoise.bringel@unistra.fr; Phone: + 33 (0)3 68 85 18 15; Fax: + 33 (0)3 68 85 20 28; Address: 28 rue Goethe, 67000 Strasbourg, France.

Running title: Organohalogen pollutant degradation and regulation

Keywords: dehalogenation; gene expression; organohalogen pollutants; transcriptional regulators; regulation

Targeted journal: *FEMS Microbiology Reviews*; *Environmental Microbiology*

Summary

To metabolize toxic organohalogen pollutants into harmless compounds, bacteria need to efficiently express genes encoding dehalogenases that cleave carbon-halogen bonds, as well as other enzymes and accessory proteins of their catabolic pathways. The corresponding genes are often clustered in dehalogenation genetic modules specific for each organohalogen compound that are transcriptionally regulated by one or several transcription factors. Organohalogen compounds themselves or their metabolites can modulate the activity of such regulators. Transcriptomic studies have shown that, in addition to dehalogenation genetic modules, a much larger genome-wide gene modulation is involved in the adaptive response of bacteria. This allows dehalogenating microorganisms to cope with the metabolic stresses associated with pollutant degradation, such as halogen ion release and pH stress, as well as metabolic rerouting. This review focuses on the transcriptional regulation of dehalogenation genetic modules in bacteria studied *in labo* and *in situ*. Understanding transcriptional regulation during halogenated pollutant degradation may help to monitor pollutant bioremediation of contaminated sites with for example bioreporters and more effective degrading strains.

Highlights

- Dehalogenases, regulators, and associated proteins are often encoded by dehalogenation genetic modules.
- Dehalogenation genetic modules enable co-regulation of diverse gene assemblages.
- Dehalogenation genetic module regulation can be strain-dependent.
- Effective dehalogenase expression may rely on availability of non-halogenated growth substrates.

Introduction

Organohalogen compounds are organic molecules with one or several carbon-halogen bonds, which can be produced by both biotic and abiotic processes (Fig. 1). Organochlorine compounds are often toxic for humans (Jayaraj, Megha and Sreedev 2016), animals (Jayaraj, Megha and Sreedev 2016), fungi (Bernat *et a*l. 2018), plants (Clausen and Trapp 2017), and micro-organisms (Tobajas *et al.* 2016). When massively produced by industry and released into the environment, these compounds can threaten human health, especially at the high concentrations found at industrial sites of production or use. Some bacteria degrade and use organohalogen compounds for growth, or as electron acceptors for organohalogen respiration (Hug *et al*. 2013). Some bacteria degrade them by fortuitous co-metabolism without direct growth benefit (Mattes *et a*l. 2015). This review addresses a neglected aspect of dehalogenation metabolism: how bacteria regulate the degradation of organohalogen pollutants at the transcriptional level. This knowledge would help to identify the molecular mechanisms that inhibit or enhance the activity of bacterial-degrading populations active in the bioremediation of contaminated environments. This review focuses on current knowledge of regulated gene sets for growth with organohalogen pollutants, which are massively released by agricultural and industrial practices into the environment, and highlights common and substrate-specific processes involved in the detection and degradation of organohalogen pollutants. It also discusses cases of how environmental factors impact transcriptional regulation processes and the fate of *in situ* degradation. Most of the examples cited in this review are about organochlorine pollutants, but they can be extrapolated to other organohalogen compounds because studies have shown similar degradation, transport or regulation with different halogenated pollutants (Chae and Zylstra 2006; Ferreira *et al*. 2009; Haque *et al*. 2013). However, this review does not discuss regulation mechanisms of biosynthesis of organohalogen compounds (Yan *et al*. 2017) and resistance to pharmaceutical organohalogen compounds, such as antibiotics (e.g. chloramphenicol, vancomycin) and biocides (eg. triclosan) (Novotna, Kwun and Hong 2016).

While carbon-halogen bonds are recalcitrant to biodegradation, differences in chemical structure (e.g. aliphatic, aromatic or poly-aromatic) and composition (substitution with other halogenated atoms, nitrogen or sulphur) confer diverse chemical properties to organochlorine compounds, usage, toxicity, and biological half-life. Half-lives of organohalogen compounds range from short to long (e.g. 45 days for pentachlorophenol to 15 years for DTT) (Jayaraj, Megha and Sreedev 2016). Indeed, some organohalogen pesticides are still detected in the environment 10 years after their use

and production have been banned by the Stockholm convention on persistent organic pollutants (Wöhrnschimmel *et al.* 2016). Despite their toxicity for many organisms, some organisms, including several bacteria, can degrade them. The carbon–halogen bond cleavage is catalysed by a large variety of dehalogenase enzymes (Agarwal *et al.* 2017; Fincker and Spormann 2017; Ang *et al.* 2018), which allows for the degradation of the remaining dehalogenation products, such as with chlorophenols (Lillis, Clipson and Doyle 2010). Some dehalogenases are constitutively expressed, but in most cases, they are more abundant in presence of the pollutant, suggesting the overlooked regulatory processes in bacterial bioremediation. Some dehalogenases are even among the most abundant proteins in the bacterial cell when utilizing organohalogen compounds as growth substrates. For instance, dichloromethane dehalogenase represents 25 % of the total soluble proteins extracted from cultures grown with dichloromethane in *Methylorubrum extorquens D*M4 (formally Me*thylobacterium extorquens D*M4, (Green and Ardley 2018)) (Muller *et al.* 2011), haloalkane dehalogenase represents 30-40 % with epichlorohydrin in Ar*throbacter s*p. AD2, and 20 % with 1,2-dichloroethane in An*cylobacter aquaticus A*D20 (van den Wijngaard, Reuvekamp and Janssen 1991). Lower amounts (2-6 %) have also been reported (van den Wijngaard *et al.* 1992). Some authors have hypothesised that such high cellular levels of dehalogenases may be needed for faster growth with organohalogen pollutants, especially when enzymes have low turnover and substrate affinity, or when pollutants are toxic for cells (van den Wijngaard, Reuvekamp and Janssen 1991; Gisi *et al.* 1998). Therefore, to degrade organohalogen pollutants, bacteria need to efficiently regulate and express genes coding for dehalogenases and other associated catabolic enzymes.

Organization and regulation of genetic modules involved in organohalogen pollutant metabolisms

Dehalogenase-encoding genes are often found clustered with genes of transcriptional regulators, other genes involved in their catabolism, transport, and accessory genes, as recently reviewed for organohalogen-respiring bacteria (Kruse, Smidt and Lechner 2016). When found in the same orientation, adjacent genes may be co-expressed as operons. Gene clusters can be part of a regulon (group of transcriptionally co-regulated genes) or a genetic module (defined as a set of genes cooperating to perform a particular physiological function (Hartwell *et al*. 1999), here the metabolism of organohalogen compounds). A large variety of genetic modules for organohalogen pollutant degradation have been found, with different gene organisation, copy number including of regulatory genes, and localizations on chromosomes, plasmids or circular intermediates (Sentchilo

et al. 2003; Tabata *et al*. 2016; Nielsen *et al*. 2017) (Fig. 2). This patchwork architecture can be explained in part by the presence of insertion sequences (Schmid-Appert *et al*. 1997), transposases (Marzorati *et al.* 2007), and recombinases (Muller *et al*. 2004) that may have promoted gene shuffling and horizontal gene transfer of dehalogenation pathways between bacteria (Liang *et al*. 2012). Here, genetic modules have been classified as follows:

Dehalogenation genetic modules harboring a dehalogenase-encoding gene adjacent to other genes, such as a regulatory gene. These modules are greatly variable in gene number (from 2 up to 20). For lindane degradation in *Sphingobium japonicum* UT26, 16 genes (*lin*) are scattered over two chromosomes and one plasmid (Tabata *et al*. 2016). They are often plasmid-borne and associated with mobile elements in genomic islands, integrative and conjugative elements or near mobile genetic elements (Sentchilo *et al*. 2003; Liang *et al.* 2012). So far, a few synthetic dehalogenation genetic modules have been constructed by metabolic engineering, including modules for dichloromethane and 1,3 dichloropropene utilization (Nikel and de Lorenzo 2013; Michener *et al*. 2014a) (Fig. 2B).

Catabolic genetic modules lacking dehalogenase-encoding genes. These modules enable organohalogen degradation without halogen removal from the carbon skeleton, leading to halogenated by-product accumulation (Fig. 2C). For example, the *bph* gene cluster is involved in biphenyl and polychlorinated biphenyl degradation in *Rhodococcus jostii* RHA1 (Takeda *et al*., 2010) These genetic modules involve transcription factors that recognize halogenated pollutants (Takeda *et al*., 2010).

Both types of genetic modules, to be efficiently expressed in presence of organohalogen pollutants, involve sensing and regulation mechanism at the transcriptional and post-transcriptional levels.

Transcription factor and their ligands during organohalogen compound degradation

Among dehalogenation genetic modules, genes encoding transcription factors are often found associated with those of dehalogenases in opposite orientation, a common characteristic of transcriptional regulators in bacteria (Beck and Warren 1988) (Fig. 2). This facilitates concerted expression (Beck and Warren 1988) and the production of transcription factors in proximity of targeted regulated genes, thereby ensuring that transcription factors do not diffuse randomly in cell before recognizing their binding sites (Weng and Xiao 2014). Transcription factors harbour at least two domains: a DNA-binding domain recognizing a specific DNA sequence (often an inverted

repeat located upstream of regulated genes), and a sensor domain, often termed companion domain or effector binding domain (Perez-Rueda *et al*. 2018). In the transcription factors involved in organohalogen detection, the DNA binding domain is mainly a HTH (Helix-Turn-Helix) domain (Fig. 3). This HTH domain is found in about half of bacterial transcription factors (Rivera-Gomez, Segovia and Perez-Rueda 2011), and in the majority of regulators involved in biodegradation (de Lorenzo *et al.* 2010).

In addition to DNA binding domains, sensor domains in transcription factor often recognize an organohalogen pollutant or metabolite produced during dehalogenation. Electrophoretic mobility shift assays and DNase footprints are often used to test whether halogenated compounds bind to transcription factors (Whangsuk *et al.*, 2010; Starr *et al.*, 2012; Sallabhan *et al.*, 2013; Torii *et al.*, 2013; Wagner *et al.*, 2013). Ligand binding can also be inferred by comparing gene expression in cultures with organohalogen compounds compared to with other substrates, using transcriptional fusion and RT-qPCR (Tab. 1). For example, the binding of trichloroethylene by several proteins, including a dehalogenase and a two-component system, has been recently found using unbiased and high-throughput thermal proteome profiling (Türkowsky *et al*. 2018b). In several recent examples, the 3D structure of crystallized transcription factors revealed that the ligand binds by non-covalent interactions to a cavity of the sensor domain at the surface of the transcription factor (Tab. 1; Fig. 4). Crystallographic structures have showed that albeit some transcription factors have one cavity that can bind several different organohalogen compounds (Ray *et al.*, 2016), others have separate cavities that allow to bind two compounds for regulation fine-tuning (Fig. 4, PcpR) (Hayes *et al*. 2014). Cavity properties (volume, solvent accessibility, hydrophobicity) determine ligand binding specificity (Liang, Woodward and Edelsbrunner 1998), so that protein structures can be used to predict the binding of new ligands by molecular docking, as previously shown for dehalogenase substrates binding (Daniel *et al*. 2015).

Ligands can be organic halogenated compounds with chlorine, bromine or fluorine (Chae and Zylstra 2006; Sanchez and Gonzalez 2007; Hayes *et al*. 2014). In addition, metabolites without carbon-halogen bonds, produced during dehalogenation, are also recognized by some transcription factors (Platero *et al*. 2012). These results reveal that transcription factors are not restricted to organohalogens, but can also bind to other organic compounds such as catalytic metabolites. This binding of small ligands induces conformational changes of the transcription factor protein structure (Fig. 4) and subsequently modifies DNA-binding properties (Anantharaman and Aravind 2005;

Möglich, Ayers and Moffat 2009).

A few transcription factor families have been specifically associated with some organohalogen pollutant groups. For example, the transcription factor of the LysR family (CatR, ClcR, LinR, PcpR; Fig. 2; Tab. 1) is mainly involved in the detection of aromatic organohalogen compounds such as chlorobenzoate or chlorophenol. The transcription factor of NtrC family (DmpR, MopR, XylR; (Ray *et al.* 2016)) is also involved in the detection of aromatic compounds (halogenated or not). On the other hand, MarR-type transcription factors, such as RdhR in *Dehaloccocoides mccartyi,* and CRP-FNR type, such as CprK in *Desulfitobacterium* strains, are often involved in the regulation of halogenated solvents degradation such as chlorinated ethanes (Jugder *et al.* 2015). For other transcription factor families, in contrast, no trends have been found. For example, ChpR, that detects chlorpyrifos, is the only member of the transcription factor CadC family known to interact with organohalogen pollutants (Whangsuk *et al*. 2010). These families of transcription factors harbour different domains to detect organohalogen compounds and regulation processes.

Mechanisms of regulation of dehalogenation genetic modules

Most studies have focused on transcription factors able to activate or repress the expression of genetic modules in response to organohalogen pollutants, with cases of interference with other growth substrates when available. So far, studies of the regulation of organohalogen pollutants metabolism have described only one case at the post-transcriptional level.

Activation and repression of dehalogenase expression by transcription factors. The expression of dehalogenases can be increased when pollutants are detected. This activation can involve either the binding or unbinding of transcription factors on DNA, as well as the interaction with other proteins. For example, activation of expression of *mop* genes in the presence of chlorophenol (Schirmer, Ehrt and Hillen 1997; Ray *et al*. 2016) involves in addition to its DNA binding HTH domain, the RNA polymerase sigma 54 factor interaction domain of MopR (Fig. 3). This transcription factor of the NtrC family binds to the promoter region of the *mop* genes and favours the sigma 54 factor and RNA polymerase interactions for increased gene expression (Bush and Dixon 2012). On the other hand, some transcription factors bind to the promoter region to occlude access to RNA polymerase in absence of ligands. For instance, DcmR (La Roche and Leisinger 1991), HdhR (Sallabhan *et al*. 2013) and RdhR1625 (Krasper *et al*. 2016) have been identified as repressors of the expression of

genes involved in the degradation of dichloromethane, chloroacetic acid and trichlorobenzene, respectively. In these cases, no ligand has yet been identified. The switching function from repressor to activator occurs in chlorobenzoate degradation, when the ligand 2-chloro-cis,cis-muconate binds to ClcR (McFall, Chugani and Chakrabarty 1998) (Fig. 5). In addition to its ability to modulate transcription by DNA-bending, ClcR interacts directly with the alpha subunit of the RNA polymerase to increase transcription activation (McFall, Chugani and Chakrabarty 1998). These mechanisms involve one transcription factor. Other processes recruit several regulators in complex interactive regulation networks as described below.

Cross-regulation of dehalogenase-encoded genes by several transcriptional regulators. Many studies have shown that more than one transcription regulator can bind to promoter regions of dehalogenation genetic modules found at multiple copies within a genome. For instance, in *Rhodococcus jostii* RHA1, *bphST* duplication and evolution led to the occurrence of regulatory proteins with unchanged DNA binding sites but different ligand sensing specificities (Takeda *et al*. 2010). The duplicated *bphST* copies are highly similar (respectively 92 and 97 % of identity at the protein level between BphS1/BphS2 sensors histidine kinases and BphT1/BphT2 response regulators). BphS1T1 and BphS2T2 detect chlorobenzene and 1,2-dichlorobenzene but only BphS1T1 recognizes biphenyl (Takeda *et al*. 2010). In another example, *Pseudomonas putida* RB1 harbours two gene clusters involved in chlorobenzoate metabolism with a common origin: *catR-catBCA* and *clcR-clcABD* (McFall, Chugani and Chakrabarty 1998). Cross-activation of the *cat* and *clc* operons is possible by the transcriptional regulator CatR, whereas ClcR can only activate expression of the *clc* operon albeit being able to bind to the *cat* operon promoter. Thus, duplication of transcription factor genes is one evolutionary mechanism that secures new functions, including the sensing of new effectors (Perez and Groisman 2009). These effectors, in addition to organohalogen compounds and by-products, may also involve other compounds associated to the catabolism of alternative growth substrates.

Catabolic repression of dehalogenases by other sources of nutriments. Cases of catabolic repression of dehalogenation gene module expression have been reported in which bacteria utilise other carbon sources in addition to halogenated compounds. Simultaneous presence of two organochlorine substrates (TCE and PCB) reduces expression of dehalogenase gene *pcbA4* in *Dehalococcoides mccartyi* CG4 by an unknown mechanism (Chen and He 2018). In this situation, TCE is dehalogenated before PCB, suggesting a substrate preference. In another example, when the

tricarboxylic acid cycle intermediate fumarate competes with the natural ligand 2-chloro-cis,cis-muconate for binding to the transcription factor ClcR, *clcABC* gene activation for chlorobenzoate utilization by ClcR is inhibited (McFall *et al*. 1997) (Fig. 5). Intermediates of the tricarboxylic acid cycle and sugars also repress *bph* gene expression for PCB degradation, for example in *Acidovorax* sp. KKS102 (Ohtsubo *et al*. 2006). This system involves a two-component regulator-sensor (*bphPQ*) located outside the *bph* gene cluster. BphQ recognizes and binds to the pE promoter upstream of the *bph* gene cluster when other carbon sources than PCB are present (Ohtsubo *et al*. 2006). When organohalogen compounds serve as a source of nitrogen in addition to carbon, the expression of their catabolic genes can be repressed in presence of an alternative preferential nitrogen source, therefore inhibiting dehalogenation (Platero *et al*. 2012). In the case of atrazine, the dehalogenation gene module *atzRSTUVW* harbours a fixation site for the transcription factor (NtrC) in addition to a binding site of a nucleoid protein (integration host factor; IHF) (Platero *et al*. 2012) (Fig. 5A). Thus, in presence of ammonium, nitrate or urea, NtrC represses *atz* gene transcription. Organohalogen compounds can be used as a sulfur and carbon source, as for the pesticide endosulfan in *Arthrobacter* sp. KW, where other sulphur sources inhibit its metabolism by a yet unknown mechanism (Weir *et al*. 2006). Transcription regulation may favour preferential organohalogen compound pathways less toxic to the host, as found for ortho- and meta- aromatic compound cleavage pathways for which enzymes from one pathway can react with substrates from the other one and lead to the formation of "dead-end" products or toxic compounds (Pérez-Pantoja *et al*. 2008). The ortho pathway is mainly used for degradation of halogenated aromatic compounds, while the meta cleavage leads to the production of end-products that block the meta pathway by damaging the enzyme catechol 2,3-dioxygenase (Bartels 1984; Kumar, Trefault and Olaniran 2014). Thus, the balance between ortho and meta pathways for the degradation of 2,4-dichlorophenoxyacetic acid and 2-methyl-4-chlorophenoxyacetic acid is clearly regulated at the transcriptional level, for example in *Sphingobium* sp. HV3 (Sipilä *et al*. 2010). Regulation by compounds not related to organohalogen pollutants indicated that each molecule that enters in bacterial cell could have a direct or indirect effect on transcriptional regulation. This possibility is discussed in the next paragraph.

Uptake of pollutants is needed for regulation. The transport of organohalogen pollutants into cells may determine dehalogenase expression in most cases. This was clearly demonstrated in *Ralstonia pickettii* PKO1, in which gene *tbuX* encoding a transporter is essential for trichloroethylene detection and the production of enzymes involved in its degradation (Kahng *et al*. 2000) (Fig. 5). In

many dehalogenation genetic modules, transporter-encoding genes are often found near dehalogenase genes (Fig. 2). These transporters involved in organohalogen uptake are located in the outer membrane (e.g. beta-barrel transporter (Belchik *et al*. 2010)) or the inner membrane (e.g. ABC transporters (Platero *et al*. 2012), TRAP transporter (Chae and Zylstra 2006) or major facilitator superfamily transporters (Hoffmann and Müller 2006; Su and Tsang 2013)). Once organohalogen pollutants have entered in cells, they can modulate transcriptional gene expression, as described above, and post-transcriptional regulation processes targeting mRNA or proteins.

Beyond transcription? Post-transcriptional regulatory mechanisms.

First, translation can be modulated by mRNA leader sequence upon formation of stem-loop structures. For example, such putative regulatory loops were predicted in the 5'-untranslated region of the mRNA of the dehalogenase *clcA* (Coco, Parsek and Chakrabarty 1994). Some riboswitchs (metabolite-binding mRNA structures) have been shown to bind fluoride but not to other halogen anions, and thus trigger the translation of some halogen transporters (Speed *et al*. 2018), enzymes that are known to be inhibited by fluoride, and proteins of unknown function (Baker *et al*. 2012). Dehalogenase and transcription factor proteins can be targets for post-translational modifications that modulate their activity. For instance, acetylation of a two-component system regulates *pceA* transcription in *Sulfurospirillum halorespirans*. Upon exposure to PCE, acetylation of the DNA-binding response regulator component (SHALO_1502) appears to mimic phosphorylation by the histidine kinase (SHALO_1503) that recognizes PCE (Türkowsky *et al*. 2018a) (Fig. 5C). This acetylation confers a lower activity than phosphorylation, but allows the regulator component to have a "memory", for example so that *pceAB* gene remains transcribed and activated even when PCE is not detected by the histidine kinase. Protein modifications (phosphorylation or glycosylation) may also occur. In *Comamonas* sp. strain CNB-1, for example, variants of proteins involved in 4-chloronitrobenzene degradation displayed different p*I* in 2D-gel electrophoresis (Zhang *et al*. 2009) and may be involved in post-transcriptional process. However, studies have not investigated post-transcriptional regulatory very much, particularly the extent and diversity of post-transcriptional regulation mechanisms in the degradation of organohalogen pollutants. This represents an attractive, long-overlooked area for future study. Studies need to evaluate genome-wide transcriptional and post-transcriptional regulation processes of dehalogenation genetic module expression and other genes directly or indirectly modulated by dehalogenation metabolism.

Global transcription response to growth with organohalogen compounds

Cells that grow with organohalogen compounds undergo different stresses that affect global gene regulation beyond regulation of the dehalogenative genetic modules *per se*. Moreover, dehalogenase genes are often located on mobile genetic elements involved in their dissemination in the environment (Liang *et al*. 2012), so that when acquired by a host, such modules may have wide-ranging effects on gene expression not necessarily linked to the functions encoded by the element (Lang and Johnson 2015; San Millan *et al*. 2015). For example, even in absence of chlorobenzoate, after acquisition of the genomic island ICE*clc*, *Pseudomonas aeruginosa* PAO1 modifies the expression of genes outside ICE*clc* (Gaillard *et al*. 2008). In addition, the dehalogenation of organohalogen compounds can lead to major changes of host genome gene expression that help the host to respond to stresses (halogen ion release, pH, metabolic re-routing) upon growth with these pollutants. Transcripts with higher or lower abundance were found for from ten up to one hundred genes during growth with chlorinated versus non-halogenated substrates, as shown using cDNA micro-array and RNA-seq approaches (Tab. 2). As expected, the abundance of host-genome encoded proteins is also modified during growth with organohalogen pollutants (for proteomic studies see (Zhang *et al.* 2009; Rupakula *et al*. 2013; Zhang *et al*. 2014; Goris *et al.* 2015; Türkowsky *et al*. 2018a, 2018b)). Genome-wide regulation involves several response processes, with genes regulated directly by transcription factors that sense organohalogen compounds or their metabolites, while other genes are regulated by transcription factors that respond to stress conditions generated by halogenated pollutant exposure and degradation. Prior to their degradation, some chlorinated solvents disrupt membrane fluidity with the efflux of small molecules such as protons and loss of proton-motive force (Koenig *et al*., 2014). Toxicity can be a side effect of organohalogen compound degradation metabolites. For example, dichloromethane metabolism leads to the formation S-chloromethyglutathione, an unstable metabolite that reacts with DNA to form alkylated adducts (Kayser and Vuilleumier 2001). Pentachlorophenol degradation leads to the formation of reactive oxygen species by reaction of an intermediate metabolite (tetrachlorobenzoquinone) with hydrogen peroxide (Yadi*d et a*l. 2013). Several studies have shown that, upon organohalogen degradation, a large variety of housekeeping stress response genes were also differentially expressed (Gvakharia *et al*. 2007; Jennings *et al*. 2009; Puglisi *et a*l. 2010; Islam *et al*. 2014). Among them, commonly found chaperones DnaK, GroES and GroEL have a large spectrum of targeted proteins (Bhandari and Houry 2015), suggesting potential wide-major effects on global gene expression triggered by dehalogenation metabolism. In contrast, chaperones encoded

within operons for reductive dehalogenation were proposed to be active only in the contest of dehalogenases folding (Kruse, Smidt and Lechner 2016). Enzymes such as catalases or peroxidases are also often found more expressed to detoxify reactive oxygen species produced during co-metabolic organohalogen pollutants degradation by oxygenases (Jennings *et a*l. 2009; Puglisi *et al.* 2010). Energy retrieved from organohalogen compound degradation is sometimes smaller than the energy needed to produce dehalogenase and adapt to pollutant-associated stresses (Cases and de Lorenzo 2005a), so that degraders needed to overexpress enzymes from central metabolism and substrate transporters (Jennings *et al*. 2009; Zhang *et al*. 2014). The need for higher cellular pools of essential cofactors of dehalogenases (e.g. cobalamin, glutathione) often explains the higher gene expression of genes involved in cofactor biosynthesis within dehalogenation genetic modules or from the host genome. For instance, *Desulfitobacterium hafniense* Y51 has increased cobalamin biosynthetic gene expression during growth with tetrachlorethene (Peng *et al*. 2012). Examples of regulated genes involved in cobalamin metabolism associated with mobile elements include operons of reductive dehalogenase (Kruse, Smidt and Lechner 2016), and plasmid-borne chloromethane utilization pathway in *M. extorquens* CM4 (Roselli *et al*. 2013; Chaignaud *et al*. 2017). These genes may also be regulated by riboswitches that detect cobalamin (Choudhary *et al*. 2013). Other glutathione related examples include the glutathione reductase gene downstream of dehalogenase (glutathione S-transferase) *dsmH2* involved in dicamba degradation (Li *et al*. 2018), and the *tcpX* (FAD reductase) for 2,4,6-trichlorophenol degradation regulated expression to cope with the reduction of cofactors oxidized during dehalogenation (Sanchez and Gonzalez 2007). Halogen extrusion was proposed to be a major adaptive response during growth with halogenated pollutants. To cope with chloride stress during dichloromethane dehalogenation, high transcription expression of the ClcA chloride/proton antiporter was needed as demonstrated after acquisition of the dehalogenase *dcmA* gene using experimental evolution studies (Michener *et al*. 2014a). These experimental evolution experiments clearly highlight that post-refinement of host housekeeping genes may be needed after acquisition of dehalogenation genetic modules for efficient growth with organohalogen compounds (Michener *et al*. 2014a, 2014b).

Most transcriptomic studies have involved reproducible, low complexity conditions, such as comparing growth with one or two substrates. More realistic environmental conditions involving catabolic repression or simultaneous dehalogenation of several organohalogen pollutants will help to better understand bacteria physiology.

Assessing complex regulation of dehalogenation genetic modules in the environment

Transcriptional regulation of organohalogen degradation compounds pathways *in labo* studies are usually performed under controlled conditions for oxygen, temperature, pH, and carbon sources. However, in polluted environments, the dynamics of interactions of physical, chemical and biological parameters are constantly changing, suggesting that adaptive complex transcriptional regulation processes are needed to sustain bacterial growth (Cases and de Lorenzo 2005b). Indeed, many efficient pollutant-degrading bacteria showing promising results when cultivated in the laboratory were ineffective when tested *in situ* for bioremediation applications (Boopathy 2000; Lovley 2003; de Lorenzo 2009). For instance, for atrazine degraders, *in situ* nitrogen sources repressed the expression of genes involved in atrazine degradation (Govantes *et al.* 2009). Organohalogen pollutants are often found together with other contaminants, resulting in additive or synergistic toxic effects (Nirmalakhandan *et al.* 1997), unsuccessful bioremediation and inefficient bacterial growth. For example, this was observed with heavy metals (Kuo and Genthner 1996; Pardue, Kongara and Jones 1996; Olaniran, Balgobind and Pillay 2013) or fungicides (White, Potter and Culbreath 2010). Indeed, the combined presence of several unrelated pollutants can influence transcriptional regulation, for example if transcription factors recognize several molecules with antagonists effects (inducers, inhibitors) (Selifonova and Eaton 1996). Another difference between *in labo* and *in situ* experiments is the presence of microorganisms that can affect dehalogenation by predation (Cunningham, Kinner and Lewis 2009) and by competition for growth substrates *in situ* (Becker 2006). In contrast, microorganisms living in close proximity as aggregates (Mao *et al*. 2015) or biofilms (Yoshida *et al*. 2009) may exchange metabolites and cofactors, in some cases enhancing activity of dehalogenating populations (Mao *et al*. 2015; Chen *et al*. 2017) (Tab. 3). For example, many organohalogen-respiring bacteria are unable to synthesize the cobalamin cofactor, and thus depend on other bacteria for its production (Fincker and Spormann 2017).

In attempts to mimic *in situ* conditions, microcosms offer controlled growth conditions and monitoring facilities, so they have been widely used to study bacterial dehalogenation in water, sediment and soil samples (Amos *et al*. 2008; Lillis, Clipson and Doyle 2010; Xiu *et al.* 2010; Men *et al*. 2017). For example, in water microcosms exposed to hexachlorobenzene, the pollutant degradation was quantitatively correlated to the transcription of the dehalogenase gene *cbdbA* relative to 16S rRNA copy numbers at various temperatures (Tas *et al.* 2011). Microcosms have also enabled dehalogenation studies in bacteria-plant interactions (Liu *et al*. 2007; Chaignaud *et al*.

2018; Wang *et al*. 2018b).

Nevertheless, *in situ* studies of polluted sites are indispensable to collect enough data to monitor and improve bioremediation processes. Of the few *in situ* studies that have focussed on the regulation during organohalogen compound degradation, the detection of mRNA biomarkers by RT-qPCR in polluted environments is by far the preferred approach (Lee *et al*. 2008; Kranzioch, Ganz and Tiehm 2015; Mattes *et al*. 2015; Ismaeil, Yoshida and Katayama 2017; Liang *et al*. 2017). The presence of gene transcripts can be correlated with biodegradation (e.g. for vinyl chloride (Liang *et al.* 2017)), which revealed actively dehalogenating populations during bioremediation processes. For example, the presence of dehalogenase encoding genes *bvcA* and *vcrA* mesured at the DNA and RNA levels were found respectively in 99 % versus 58 % of the tested samples (95 samples from 6 contaminated environments) (Liang *et al.*, 2017). Indeed, dehalogenase genes can be harboured by dead cells or be unexpressed at the contaminated site. However, detection of transcripts is not necessarily correlated to biodegradation (Lee *et al.* 2006), especially when dehalogenase genes are constitutively expressed (Paulin, Nicolaisen and Sorensen 2010; Peng *et al*. 2012; T'Syen *et al*. 2015). Taken together, all these results highlight that the presence of DNA, mRNA or protein biomarkers of organohalogen metabolism needs to be carefully interpreted (Rahm, Morris and Richardson 2006). Currently used biomarkers at the DNA, mRNA or protein levels are mainly genes of dehalogenases (Werner *et al*. 2009; Kranzioch, Ganz and Tiehm 2015; Matturro and Rossetti 2015; Ibrahim *et al*. 2017; Liang *et al*. 2017; Hermon *et al*. 2018). We believe that future studies of dehalogenation gene module regulation will help to discover environmental relevant limiting parameters in bioremediation processes, select more efficient degraders and improve bioremediation protocols.

Potential of tools based on regulatory processes to improve bioremediation

Bioremediation strategies can be improved by discovering bacterial strains with better growth fitness, dehalogenase expression or new dehalogenation capacities. Metabolic engineering of dehalogenase can modify its activity and substrate spectrum (Pavlova *et al*. 2009; Dvořák *et al*. 2017; Ang *et al*. 2018), whereas that of transcription factors can expand ligand recognition ability (Wise and Kuske 2000; Mohn *et al.* 2006; Beggah *et al.* 2007; Lang and Ogawa 2009). Synthetic biology combined with metabolic engineering (Dvořák *et al.* 2017) has been used to build new dehalogenation pathways, called here synthetic dehalogenation genetic modules (Fig. 2B). When

acquired by naïve bacteria, synthetic dehalogenation genetic modules are expected to confer better degradation fitness of bacteria used in bioremediation (Zhang *et al.* 2010) or can also be produced in plants, by heterologous systems, for phytoremediation (Mena-Benitez *et al*. 2008; Abhilash, Jamil and Singh 2009). Such modules, constitutive or regulated by heterologous promoters (e.g. *lac*, *tac*, *nod*), have been added upstream of dehalogenase-encoding genes to target and enhance gene expression *in situ*. To increase *Pseudomonas* strain biodegradation capacity of polychlorinated biphenyls in the rhizosphere, the bacterial *bph* genes were expressed from the *nod* promoter that is induced by plant-produced flavonoids (Villacieros *et al*. 2005). Deletion of genes outside of the dehalogenation genetic modules may also enhance growth and degradation (Nikel and de Lorenzo 2013), such as the deletion of flagella-encoding genes to avoid wasting energy. Diauxic growth with other compounds can be prevented by gene deletion, as applied to prevent growth with glycerol and favour 1,2,3-trichloropropane utilization following *glpR* transcription factor-encoding gene deletion (Gong *et al*. 2017). Acquisition of genes involved in the adaptive response to growth with specific organochlorine compounds has been shown to improve bacterial growth with these toxic compounds. For example, expression of *clcA* gene of the chloride/proton antiporter in addition of a dehalogenase was a key factor for growth with dichloromethane but not with chloromethane (Michener *et al*. 2016).

Bioreporter strains with a fluorescent protein induced by organohalogen pollutants are useful tools to study the expression of dehalogenases and localize dehalogenating bacteria in microcosms (Zhang *et al*. 2010). This was successfully used to demonstrate the expression of polychlorinated biphenyls and chloromethane dehalogenase genes in the rhizosphere and the phyllosphere respectively (Liu *et al.* 2010; Farhan Ul Haque *et al*. 2017). Monitoring of bioreporter activity is often limited *in situ* by medium opacity and natural autofluorescence of compounds in the environment. Use of bioreporters coupled with flow cytometry sorting can help to overcome these problems, despite loss of information about bacteria localization (Norman, Hansen and Sørensen 2006; Beggah *et al*. 2007). Bioreporters can also be used to quantify organohalogen pollutants in environmental matrices (Lopes *et al*. 2012; Hua *et al*. 2015; Whangsuk *et al.* 2016; Farhan Ul Haque *et al*. 2017). Many examples involving different strategies have been explored. Two-component systems offer the advantage of directly sensing pollutants in the environmental conditions (or in the periplasm) compared to transcription factors with a cytoplasmic localization (Ravikumar *et al*. 2017). To avoid artefacts due to pollutant transport across membranes when bioreporters are used to quantify pollutants, chimeric proteins can be built by replacing the histidine

kinase domain of a chemoreceptor with the kinase domain of a two-component system (Lacal *et al*. 2013). All these genetic circuits need to be carefully constructed and optimised to prevent physiological processes such as catabolic repression from interfering with pollutant detection (Marqués, Aranda-Olmedo and Ramos 2006). Gene expression can also be regulated by halogen atoms released during dehalogenation metabolism with a promoter inducible by chloride (Geldart, Borrero and Kaznessis 2015) or a riboswitch regulated by fluoride (Speed *et al*. 2018). These can be useful to detect dehalogenation in cells or to activate gene expression proportionally to dehalogenation. All these approaches of synthetic biology will provide better tools to study transcriptional regulation and improve strain efficiency in bioremediation.

In conclusion, regulation pathways can be a limiting step in effective biodegradation of organohalogen pollutants. Although data about the genetics and biochemistry of dehalogenases has accumulated over the last 30 years, this review shows that investigating regulation is crucial to fully understand the dehalogenation processes.

Bibliography

Abhilash, P.C., Jamil, S., and Singh, N. (2009) Transgenic plants for enhanced biodegradation and phytoremediation of organic xenobiotics. *Biotechnol. Adv.* **27**: 474–488.

Adrian, L., Rahnenfuhrer, J., Gobom, J., and Holscher, T. (2007) Identification of a chlorobenzene reductive dehalogenase in *Dehalococcoides* sp. strain CBDB1. *Appl. Environ. Microbiol.* **73**: 7717–7724.

Agarwal, V., Miles, Z.D., Winter, J.M., Eustaquio, A.S., El Gamal, A.A., and Moore, B.S. (2017) Enzymatic halogenation and dehalogenation reactions: Pervasive and mechanistically diverse. *Chem. Rev.* **117**: 5619–5674.

Albers, P., Weytjens, B., De Mot, R., Marchal, K., and Springael, D. (2018) Molecular processes underlying synergistic linuron mineralization in a triple-species bacterial consortium biofilm revealed by differential transcriptomics. *MicrobiologyOpen*.

Amos, B.K., Ritalahti, K.M., Cruz-Garcia, C., Padilla-Crespo, E., and Loffler, F.E. (2008) Oxygen effect on *Dehalococcoides* viability and biomarker quantification. *Environ. Sci. Technol.* **42**: 5718–5726.

Anantharaman, V. and Aravind, L. (2005) MEDS and PocR are novel domains with a predicted role in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. *Bioinformatics* **21**: 2805–2811.

Ang, T.F., Maiangwa, J., Salleh, A., Normi, Y., and Leow, T. (2018) Dehalogenases: From improved performance to potential microbial dehalogenation applications. *Molecules* **23**: 1100.

Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B., and Breaker, R.R. (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**: 233–235.

Bartels, I. (1984) Suicide inactivation of catechol 2,3-dioxygenase from *Pseudomonas putida* mt-2 by 3-halocatechols. *Appl. Environ. Microbiol.* **47**: 500–505.

Beck, C.F. and Warren, R.A. (1988) Divergent promoters, a common form of gene organization. *Microbiol Rev* **52**: 318–326.

Becker, J.G. (2006) A modeling study and implications of competition between *Dehalococcoides ethenogenes* and other tetrachloroethene- respiring bacteria. *Environ. Sci. Technol.* **40**: 4473–4480.

Beggah, S., Vogne, C., Zenaro, E., and van der Meer, J.R. (2007) Mutant HbpR transcription activator isolation for 2-chlorobiphenyl via green fluorescent protein-based flow cytometry and cell sorting. *Microb. Biotechnol.* **0**: 071018044055002.

Belchik, S.M., Schaeffer, S.M., Hasenoehrl, S., and Xun, L. (2010) A β-barrel outer membrane protein facilitates cellular uptake of polychlorophenols in *Cupriavidus necator*. *Biodegradation* **21**: 431–439.

Bernat, Przemyslaw, Nykiel-Szymanska, J., Stolarek, P., Slaba, M., Szewczyk, R., and Rozalska, S. (2018) 2,4-dichlorophenoxyacetic acid-induced oxidative stress: Metabolome and membrane modifications in *Umbelopsis isabellina*, a herbicide degrader. *PLoS ONE* **13**: e0199677.

Bers, K., Leroy, B., Breugelmans, P., Albers, P., Lavigne, R., Sorensen, S.R., et al. (2011) A novel hydrolase identified by genomic-proteomic analysis of phenylurea herbicide mineralization by *Variovorax* sp. strain SRS16. *Appl. Environ. Microbiol.* **77**: 8754–8764.

Bhandari, V. and Houry, W.A. (2015) Substrate interaction networks of the *Escherichia coli* chaperones: Trigger Factor, DnaK and GroEL. In, Krogan,N.J. and Babu,M. (eds), *Prokaryotic Systems Biology*. Springer International Publishing, Cham, pp. 271–294.

Boopathy, R. (2000) Factors limiting bioremediation technologies. *Bioresour. Technol.* **74**: 63–67.

Bush, M. and Dixon, R. (2012) The role of bacterial enhancer binding proteins as specialized activators of sigma 54-dependent transcription. *Microbiol Mol Biol Rev* **76**: 497–529.

Cai, M. and Xun, L. (2002) Organization and regulation of pentachlorophenol-degrading genes in *Sphingobium chlorophenolicum* ATCC 39723. *J Bacteriol* **184**: 4672–4680.

Cases, I. and de Lorenzo, Victor (2005) Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. *Int. Microbiol.* **8**: 213–222.

Cases, I. and de Lorenzo, Victor (2005) Promoters in the environment: transcriptional regulation in its natural context. *Nat. Rev. Microbiol.* **3**: 105–118.

Chae, J.C. and Zylstra, G.J. (2006) 4-Chlorobenzoate uptake in *Comamonas* sp. strain DJ-12 is mediated by a tripartite ATP-independent periplasmic transporter. *J. Bacteriol.* **188**: 8407–8412.

Chaignaud, P., Maucourt, B., Weiman, M., Alberti, A., Kolb, S., Cruveiller, S., *et al.* (2017) Genomic and transcriptomic analysis of growth-supporting dehalogenation of chlorinated methanes in *Methylobacterium*. *Front Microbiol* **8**:.

Chaignaud, P., Morawe, M., Besaury, L., Krober, E., Vuilleumier, S., Bringel, F., and Kolb, S. (2018) Methanol consumption drives the bacterial chloromethane sink in a forest soil. *ISME J.*

Chen, G., Kleindienst, S., Griffiths, D.R., Mack, E.E., Seger, E.S., and Loffler, F.E. (2017) Mutualistic interaction between dichloromethane- and chloromethane-degrading bacteria in an anaerobic mixed culture: Mutualistic degradation of chlorinated methanes. *Environ. Microbiol.*

Chen, K., Mu, Y., Jian, S., Zang, X., Chen, Q., Jia, W., *et al.* (2018) Comparative transcriptome analysis reveals the mechanism underlying 3,5-dibromo-4-hydroxybenzoate catabolism via a new oxidative decarboxylation pathway. *Appl Env. Microbiol* **84**: e02467–17.

Choudhary, P.K., Duret, A., Rohrbach-Brandt, E., Holliger, C., Sigel, R.K.O., and Maillard, J. (2013) Diversity of cobalamin riboswitches in the corrinoid-producing organohalide respirer *Desulfitobacterium hafniense*. *J. Bacteriol.* **195**: 5186–5195.

Clausen, L.P.W. and Trapp, S. (2017) Toxicity of 56 substances to trees. *Environ. Sci. Pollut. Res.* **24**: 18035–18047.

Coco, W.M., Parsek, M.R., and Chakrabarty, A.M. (1994) Purification of the LysR family regulator, ClcR, and its interaction with the *Pseudomonas putida clcABD* chlorocatechol operon promoter. *J Bacteriol* **176**: 5530–5533.

Cunningham, J.J., Kinner, N.E., and Lewis, M. (2009) Protistan predation affects trichloroethene biodegradation in a bedrock aquifer. *Appl. Environ. Microbiol.* **75**: 7588–7593.

Daniel, L., Buryska, T., Prokop, Z., Damborsky, J., and Brezovsky, J. (2015) Mechanism-based discovery of novel substrates of haloalkane dehalogenases using *in silico* screening. *J. Chem. Inf. Model.* **55**: 54–62.

Diaz-Marrero, A.R., Rovirosa, J., Darias, J., San-Martin, A., and Cueto, M. (2002) Plocamenols A−C, novel linear polyhalohydroxylated monoterpenes from *Plocamium cartilagineum*. *J. Nat. Prod.* **65**: 585–588.

Dvořák, P., Nikel, P.I., Damborský, J., and de Lorenzo, V. (2017) Bioremediation 3.0: Engineering pollutant-removing bacteria in the times of systemic biology. *Biotechnol. Adv.* **35**: 845–866.

Farhan Ul Haque, M., Besaury, L., Nadalig, T., Bringel, F., Mutterer, J., Schaller, H., and Vuilleumier, S. (2017) Correlated production and consumption of chloromethane in the *Arabidopsis thaliana* phyllosphere. *Sci. Rep.* **7**: 17589.

Ferreira, M.I.M., Iida, T., Hasan, S.A., Nakamura, K., Fraaije, M.W., Janssen, D.B., and Kudo, T. (2009) Analysis of two gene clusters involved in the degradation of 4-fluorophenol by *Arthrobacter* sp. Strain IF1. *Appl. Environ. Microbiol.* **75**: 7767–7773.

Fincker, M. and Spormann, A.M. (2017) Biochemistry of catabolic reductive dehalogenation. *Annu. Rev. Biochem.* **86**: 357–386.

Gaillard, M., Pernet, N., Vogne, C., Hagenbuchle, O., and van der Meer, J.R. (2008) Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. *Proc Natl Acad Sci U A* **105**: 7058–7063.

Geldart, K., Borrero, J., and Kaznessis, Y.N. (2015) Chloride-inducible expression vector for delivery of antimicrobial peptides targeting antibiotic-resistant *Enterococcus faecium*. *Appl Env. Microbiol* **81**: 3889–3897.

Gisi, D., Willi, L., Traber, H., Leisinger, T., and Vuilleumier, S. (1998) Effects of bacterial host and dichloromethane dehalogenase on the competitiveness of methylotrophic bacteria growing with dichloromethane. *Appl. Environ. Microbiol.* **64**: 1194–1202.

Gong, T., Xu, X., Che, Y., Liu, R., Gao, W., Zhao, F., *et al*. (2017) Combinatorial metabolic engineering of *Pseudomonas putida* KT2440 for efficient mineralization of 1,2,3-trichloropropane. *Sci. Rep.* **7**:.

Goris, T., Schiffmann, C.L., Gadkari, J., Schubert, T., Seifert, J., Jehmlich, N., *et al*. (2015) Proteomics of the organohalide-respiring *Epsilonproteobacterium Sulfurospirillum multivorans* adapted to tetrachloroethene and other energy substrates. *Sci. Rep.* **5**:.

Govantes, F., Garcia-Gonzalez, V., Porrua, O., Platero, A.I., Jimenez-Fernandez, A., and Santero, E. (2010) Regulation of the atrazine-degradative genes in *Pseudomonas* sp. strain ADP. *FEMS Microbiol Lett* **310**: 1–8.

Govantes, F., Porrua, O., Garcia-Gonzalez, V., and Santero, E. (2009) Atrazine biodegradation in the lab and in the field: Enzymatic activities and gene regulation. *Microb. Biotechnol.* **2**: 178–185.

Green, P.N. and Ardley, J.K. (2018) Review of the genus *Methylobacterium* and closely related organisms: a proposal that some *Methylobacterium* species be reclassified into a new genus, *Methylorubrum* gen. nov. *Int. J. Syst. Evol. Microbiol.* **68**: 2727–2748.

Gribble, G.W. (2010) Naturally occurring organohalogen compounds - A comprehensive update Springer-Verlag, Wien.

Gvakharia, B.O., Permina, E.A., Gelfand, M.S., Bottomley, P.J., Sayavedra-Soto, L.A., and Arp, D.J. (2007) Global transcriptional response of *Nitrosomonas* europaea to chloroform and chloromethane. *Appl Env. Microbiol* **73**: 3440–3445.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999) From molecular to modular cell biology. *Nature* **402**: C47–C52.

Hayes, R.P., Moural, T.W., Lewis, K.M., Onofrei, D., Xun, L., and Kang, C. (2014) Structures of the inducer-binding domain of pentachlorophenol-degrading gene regulator PcpR from *Sphingobium chlorophenolicum*. *Int J Mol Sci* **15**: 20736–20752.

Hermon, L., Denonfoux, J., Hellal, J., Joulian, C., Ferreira, S., Vuilleumier, S., and Imfeld, G. (2018) Dichloromethane biodegradation in multi-contaminated groundwater: Insights from biomolecular and compound-specific isotope analyses. *Water Res.* **142**: 217–226.

Hoffmann, D. and Muller, R.H. (2006) 2,4-Dichlorophenoxyacetic acid (2,4-D) utilization by *Delftia acidovorans* MC1 at alkaline pH and in the presence of dichlorprop is improved by introduction of the *tfdK* gene. *Biodegradation* **17**: 263–273.

Hua, A., Gueuné, H., Cregut, M., Thouand, G., and Durand, M.J. (2015) Development of a bacterial bioassay for atrazine and cyanuric acid detection. *Front Microbiol* **6**:.

Hug, L.A., Maphosa, F., Leys, D., Loffler, F.E., Smidt, H., Edwards, E.A., and Adrian, L. (2013) Overview of organohalide-respiring bacteria and a proposal for a classification system for reductive dehalogenases. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120322–20120322.

Ibrahim, E.S., Kashef, M.T., Essam, T.M., and Ramadan, M.A. (2017) A degradome-based polymerase chain reaction to resolve the potential of environmental samples for 2,4-dichlorophenol biodegradation. *Curr. Microbiol.*

Islam, M.A., Waller, A.S., Hug, L.A., Provart, N.J., Edwards, E.A., and Mahadevan, R. (2014) New insights into *Dehalococcoides mccartyi* metabolism from a reconstructed metabolic network-based systems-level analysis of *D. mccartyi* transcriptomes. *PLoS ONE* **9**: e94808.

Ismaeil, M., Yoshida, N., and Katayama, A. (2017) Identification of multiple dehalogenase genes involved in tetrachloroethene-to-ethene dechlorination in a *Dehalococcoides*-dominated enrichment culture. *BioMed Res. Int.*

Jayaraj, R., Megha, P., and Sreedev, P. (2016) Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. *Interdiscip Toxicol* **9**: 90–100.

Jennings, L.K., Chartrand, M.M.G., Lacrampe-Couloume, G., Lollar, B.S., Spain, J.C., and Gossett, J.M. (2009) Proteomic and transcriptomic analyses reveal genes upregulated by cis-dichloroethene in *Polaromonas* sp. strain JS666. *Appl Env. Microbiol* **75**: 3733–3744.

Joyce, M.G., Levy, C., Gabor, K., Pop, S.M., Biehl, B.D., Doukov, T.I., *et al*. (2006) CprK crystal structures reveal mechanism for transcriptional control of halorespiration. *J. Biol. Chem.* **281**: 28318–28325.

Jugder, B.E., Ertan, H., Lee, M., Manefield, M., and Marquis, C.P. (2015) Reductive dehalogenases come of age in biological destruction of organohalides. *Trends Biotechnol.* **33**: 595–610.

Kahng, H.Y., Byrne, A.M., Olsen, R.H., and Kukor, J.J. (2000) Characterization and role of *tbuX* in utilization of toluene by *Ralstonia pickettii* PKO1. *J. Bacteriol.* **182**: 1232–1242.

Kayser, M.F. and Vuilleumier, S. (2001) Dehalogenation of dichloromethane by dichloromethane dehalogenase/glutathione S-transferase leads to formation of DNA adducts. *J. Bacteriol.* **183**: 5209–5212.

Kim, S.H., Harzman, C., Davis, J.K., Hutcheson, R., Broderick, J.B., Marsh, T.L., and Tiedje, J.M. (2012) Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiol* **12**: 21.

Koenig, J.C., Groissmeier, K.D., and Manefield, M.J. (2014) Tolerance of anaerobic bacteria to chlorinated solvents. *Microbes Env.* **29**: 23–30.

Kranzioch, I., Ganz, S., and Tiehm, A. (2015) Chloroethene degradation and expression of *Dehalococcoides dehalogenase* genes in cultures originating from Yangtze sediments. *Environ. Sci. Pollut. Res.* **22**: 3138–3148.

Krasper, L., Lilie, H., Kublik, A., Adrian, L., Golbik, R., and Lechner, U. (2016) The MarR-type regulator Rdh2R regulates *rdh* gene transcription in *Dehalococcoides mccartyi* strain CBDB1. *J. Bacteriol.* **198**: 3130–3141.

Kruse, T., Smidt, H., and Lechner, U. (2016) Comparative genomics and transcriptomics of organohalide-respiring bacteria and regulation of *rdh* gene transcription. In, *Organohalide-Respiring Bacteria*. Springer Berlin Heidelberg, pp. 345–376.

Kumar, A., Trefault, N., and Olaniran, A.O. (2014) Microbial degradation of 2,4-dichlorophenoxyacetic acid: Insight into the enzymes and catabolic genes involved, their regulation and biotechnological implications. *Crit. Rev. Microbiol.* 1–15.

Kuo, C. and Genthner, B. (1996) Effect of added heavy metal ions on biotransformation and biodegradation of 2-chlorophenol and 3-chlorobenzoate in anaerobic bacterial consortia. *Appl. Environ. Microbiol.* **62**: 2317–2323.

La Roche, S.D. and Leisinger, T. (1991) Identification of *dcmR*, the regulatory gene governing expression of dichloromethane dehalogenase in *Methylobacterium* sp. strain DM4. *J. Bacteriol.* **173**: 6714–6721.

Lacal, J., Reyes-Darias, J.A., Garcia-Fontana, C., Ramos, J. l., and Krell, T. (2013) Tactic responses to pollutants and their potential to increase biodegradation efficiency. *J. Appl. Microbiol.* **114**: 923–933.

Lang, G. and Ogawa, N. (2009) Mutational analysis of the inducer recognition sites of the LysR-type transcriptional regulator TfdT of *Burkholderia* sp. NK8. *Appl. Microbiol. Biotechnol.* **83**: 1085–1094.

Lang, K.S. and Johnson, T.J. (2015) Transcriptome modulations due to A/C2 plasmid acquisition. *Plasmid* **80**: 83–89.

Lee, P.K.H., Johnson, D.R., Holmes, V.F., He, J., and Alvarez-Cohen, L. (2006) Reductive dehalogenase gene expression as a biomarker for physiological activity of *Dehalococcoides* spp. *Appl. Environ. Microbiol.* **72**: 6161–6168.

Lee, P.K.H., Macbeth, T.W., Sorenson, K.S., Deeb, R.A., and Alvarez-Cohen, L. (2008) Quantifying genes and transcripts to assess the in situ physiology of *Dehalococcoides* spp. in a trichloroethene-contaminated groundwater site. *Appl. Environ. Microbiol.* **74**: 2728–2739.

Li, N., Tong, R.-L., Yao, L., Chen, Q., Yan, X., Ding, D.-R., *et al.* (2018) Identification of two glutathione-dependent 3,6-dichlorogentisate dehalogenases and their roles in the catabolism of the herbicide dicamba in *Rhizorhabdus dicambivorans* Ndbn-20. *Appl. Environ. Microbiol.* AEM.00623-18.

Li, N., Yao, L., He, Q., Qiu, J., Cheng, D., Ding, D., et al. (2017) 3,6-Dichlorosalicylate catabolism Is initiated by the DsmABC cytochrome P450 monooxygenase system in *Rhizorhabdus* dicambivorans Ndbn-20. *Appl. Environ. Microbiol.* **84**: e02133–17.

Liang, yi, Liu, X., Singletary, M.A., Wang, K., and Mattes, T.E. (2017) Relationships between the abundance and expression of functional genes from vinyl chloride (VC)-degrading bacteria and geochemical parameters at VC-contaminated sites. *Environ. Sci. Technol.*

Liang, B., Jiang, J., Zhang, J., Zhao, Y., and Li, S. (2012) Horizontal transfer of dehalogenase genes involved in the catalysis of chlorinated compounds: Evidence and ecological role. *Crit Rev Microbiol* **38**: 95–110.

Liang, J., Woodward, C., and Edelsbrunner, H. (1998) Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**: 1884–1897.

Lillis, L., Clipson, N., and Doyle, E. (2010) Quantification of catechol dioxygenase gene expression in soil during degradation of 2,4-dichlorophenol. *FEMS Microbiol Ecol* **73**: 363–369.

Liu, X., Germaine, K.J., Ryan, D., and Dowling, D.N. (2007) Development of a GFP-based biosensor for detecting the bioavailability and biodegradation of polychlorinated biphenyls (PCBs). *J. Environ. Eng. Landsc. Manag.* **15**: 261–268.

Liu, X., Germaine, K.J., Ryan, D., and Dowling, D.N. (2010) Whole-cell fluorescent biosensors for bioavailability and biodegradation of polychlorinated biphenyls. *Sensors* **10**: 1377–1398.

Lopes, N., Hawkins, S.A., Jegier, P., Menn, F.M., Sayler, G.S., and Ripp, S. (2012) Detection of dichloromethane with a bioluminescent (*lux*) bacterial bioreporter. *J. Ind. Microbiol. Biotechnol.* **39**: 45–53.

de Lorenzo, V. (2009) Recombinant bacteria for environmental release: what went wrong and what we have learnt from it. *Clin. Microbiol. Infect.* **15**: 63–65.

de Lorenzo, V., Silva-Rocha, R., Carbajosa, G., C Galvao, T., and Cases, I. (2010) Sensing xenobiotic compounds: Lessons from bacteria that face pollutants in the environment., p. 81.

Lovley, D.R. (2003) Cleaning up with genomics: applying molecular biology to bioremediation. *Nat. Rev. Microbiol.* **1**: 35–44.

Mao, X., Stenuit, B., Polasko, A., and Alvarez-Cohen, L. (2015) Efficient metabolic exchange and electron transfer within a syntrophic trichloroethene-degrading coculture of *Dehalococcoides mccartyi* 195 and *Syntrophomonas wolfei*. *Appl. Environ. Microbiol.* **81**: 2015–2024.

Marqués, S., Aranda-Olmedo, I., and Ramos, J.L. (2006) Controlling bacterial physiology for optimal expression of gene reporter constructs. *Curr. Opin. Biotechnol.* **17**: 50–56.

Marzorati, M., de Ferra, F., Van Raemdonck, H., Borin, S., Allifranchini, E., Carpani, G., *et al.* (2007) A novel reductive dehalogenase, identified in a contaminated groundwater enrichment culture and in *Desulfitobacterium dichloroeliminans* strain DCA1, is linked to dehalogenation of 1,2-dichloroethane. *Appl. Environ. Microbiol.* **73**: 2990–2999.

Mattes, T.E., Jin, Y.O., Livermore, J., Pearl, M., and Liu, X. (2015) Abundance and activity of vinyl chloride (VC)-oxidizing bacteria in a dilute groundwater VC plume biostimulated with oxygen and ethene. *Appl. Microbiol. Biotechnol.* **99**: 9267–9276.

Matturro, B. and Rossetti, S. (2015) GeneCARD-FISH: Detection of *tceA* and *vcrA* reductive dehalogenase genes in *Dehalococcoides mccartyi* by fluorescence *in situ* hybridization. *J. Microbiol. Methods* **110**: 27–32.

McFall, S.M., Abraham, B., Narsolis, C.G., and Chakrabarty, A.M. (1997) A tricarboxylic acid cycle intermediate regulating transcription of a chloroaromatic biodegradative pathway: fumarate-mediated repression of the *clcABD* operon. *J. Bacteriol* **179**: 6729–6735.

McFall, S.M., Chugani, S.A., and Chakrabarty, A.M. (1998) Transcriptional activation of the catechol and chlorocatechol operons: variations on a theme. *Gene* **223**: 257–267.

Men, Y., Feil, H., Verberkmoes, N.C., Shah, M.B., Johnson, D.R., Lee, P.K.H., *et al*. (2012) Sustainable syntrophic growth of *Dehalococcoides ethenogenes* strain 195 with *Desulfovibrio vulgaris* Hildenborough and *Methanobacterium congolense*: global transcriptomic and proteomic analyses. *ISME J* **6**: 410–421.

Men, Y., Yu, K., Baelum, J., Gao, Y., Tremblay, J., Prestat, E., *et al.* (2017) Metagenomic and metatranscriptomic analyses reveal the structure and dynamics of a dechlorinating community containing *Dehalococcoides mccartyi* and corrinoid-providing microorganisms under cobalamin-limited conditions. *Appl. Environ. Microbiol.* **83**: e03508–16.

Mena-Benitez, G.L., Gandia-Herrero, F., Graham, S., Larson, T.R., McQueen-Mason, S.J., French, C.E., *et al.* (2008) Engineering a catabolic pathway in plants for the degradation of 1,2-dichloroethane. *PLANT Physiol.* **147**: 1192–1198.

Michener, J.K., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014a) Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. *J. Bacteriol.* **196**: 2101–2107.

Michener, J.K., Camargo Neves, A.A., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014b) Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post-transfer refinement. *ELife* **3**:.

Michener, J.K., Vuilleumier, S., Bringel, F., and Marx, C.J. (2016) Transfer of a catabolic pathway for chloromethane in *Methylobacterium* strains highlights different limitations for growth with chloromethane or with dichloromethane. *Front. Microbiol.* **7**:.

Miyauchi, K., Lee, H.S., Fukuda, M., Takagi, M., and Nagata, Y. (2002) Cloning and characterization of *linR*, involved in regulation of the downstream pathway for hexachlorocyclohexane degradation in *Sphingomonas paucimobilis* UT26. *Appl. Environ. Microbiol.* **68**: 1803–1807.

Moglich, A., Ayers, R.A., and Moffat, K. (2009) Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* **17**: 1282–1294.

Mohn, W.W., Garmendia, J., Galvao, T.C., and De Lorenzo, V. (2006) Surveying biotransformations with *à la carte* genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into *lacZ*-based phenotypes. *Environ. Microbiol.* **8**: 546–555.

Muller, E.E.L., Hourcade, E., Louhichi-Jelail, Y., Hammann, P., Vuilleumier, S., and Bringel, F. (2011) Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4. *Environ. Microbiol.* **13**: 2518–2535.

Muller, J.A., Rosner, B.M., von Abendroth, G., Meshulam-Simon, G., McCarty, P.L., and Spormann, A.M. (2004) Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. strain VS and its environmental distribution. *Appl. Environ. Microbiol.* **70**: 4880–4888.

Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., and Senda, T. (2003) Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *J. Mol. Biol.* **328**: 555–566.

Nielsen, T.K., Rasmussen, M., Demanèche, S., Cecillon, S., Vogel, T.M., and Hansen, L.H. (2017) Evolution of *Sphingomonad* gene clusters related to pesticide catabolism revealed by genome sequence and mobilomics of *Sphingobium herbicidovorans* MH. *Genome Biol. Evol.* **9**: 2477–2490.

Nikel, P.I. and de Lorenzo, V. (2013) Engineering an anaerobic metabolic regime in *Pseudomonas putida* KT2440 for the anoxic biodegradation of 1,3-dichloroprop-1-ene. *Metab. Eng.* **15**: 98–112.

Nirmalakhandan, N., Xu, S., Trevizo, C., Brennan, R., and Peace, J. (1997) Additivity in microbial toxicity of nonuniform mixtures of organic chemicals. *Ecotoxicol. Environ. Saf.* **37**: 97–102.

Norman, A., Hansen, L.H., and Sørensen, S.J. (2006) A flow cytometry-optimized assay using an SOS–green fluorescent protein (SOS–GFP) whole-cell biosensor for the detection of genotoxins in complex environments. *Mutat. Res. Toxicol. Environ. Mutagen.* **603**: 164–172.

Novotna, G.B., Kwun, M.J., and Hong, H.-J. (2016) *In vivo* characterization of the activation and interaction of the VanR-VanS two-component regulatory system controlling glycopeptide antibiotic resistance in two related *Streptomyces* Species. *Antimicrob. Agents Chemother.* **60**: 1627–1637.

Ohtsubo, Y., Goto, H., Nagata, Y., Kudo, T., and Tsuda, M. (2006) Identification of a response regulator gene for catabolite control from a PCB-degrading *beta-proteobacteria, Acidovorax* sp. KKS102. *Mol. Microbiol.* **60**: 1563–1575.

Olaniran, A.O., Balgobind, A., and Pillay, B. (2013) Bioavailability of heavy metals in soil: impact on microbial biodegradation of organic compounds and possible improvement strategies. *Int. J. Mol. Sci.* **14**: 10197–10228.

Pardue, J.H., Kongara, S., and Jones, J.W. (1996) Effect of cadmium on reductive dechlorination of trichloroaniline. *Environ. Toxicol. Chem.* **15**: 1083–1088.

Parnell, J.J., Park, J., Denef, V., Tsoi, T., Hashsham, S., Quensen, J., and Tiedje, J.M. (2006) Coping with polychlorinated biphenyl (PCB) toxicity: Physiological and genome-wide responses of *Burkholderia xenovorans* LB400 to PCB-mediated stress. *Appl Env. Microbiol* **72**: 6607–6614.

Paulin, M.M., Nicolaisen, M.H., and Sorensen, J. (2010) Abundance and expression of enantioselective *rdpA* and *sdpA* dioxygenase genes during degradation of the racemic herbicide (R,S)-2-(2,4-dichlorophenoxy)propionate in soil. *Appl. Environ. Microbiol.* **76**: 2873–2883.

Pavlova, M., Klvana, M., Prokop, Z., Chaloupkova, R., Banas, P., Otyepka, M., *et al*. (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.* **5**: 727–733.

Peng, X., Yamamoto, S., Vertès, A.A., Keresztes, G., Inatomi, K., Inui, M., and Yukawa, H. (2012) Global transcriptome analysis of the tetrachloroethene-dechlorinating bacterium *Desulfitobacterium hafniense* Y51 in the presence of various electron donors and terminal electron acceptors. *J. Ind. Microbiol. Biotechnol.* **39**: 255–268.

Perez, J.C. and Groisman, E.A. (2009) Evolution of transcriptional regulatory circuits in bacteria. *Cell* **138**: 233–244.

Perez-Pantoja, D., De la Iglesia, R., Pieper, D.H., and Gonzalez, B. (2008) Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus* necator JMP134. *FEMS Microbiol. Rev.* **32**: 736–794.

Perez-Rueda, E., Hernandez-Guerrero, R., Martinez-Nunez, M.A., Armenta-Medina, D., Sanchez, I., and Ibarra, J.A. (2018) Abundance, diversity and domain architecture variability in prokaryotic DNA-binding transcription factors. *PLoS ONE* **13**: e0195332.

Platero, A.I., Garcia-Jaramillo, M., Santero, E., and Govantes, F. (2012) Transcriptional organization and regulatory elements of a *Pseudomonas* sp. strain ADP operon encoding a LysR-type regulator and a putative solute transport system. *J. Bacteriol.* **194**: 6560–6573.

Poelarends, G.J., Kulakov, L.A., Larkin, M.J., Vlieg, J.E.T. van H., and Janssen, D.B. (2000) Roles of horizontal gene transfer and gene integration in evolution of 1,3-dichloropropene- and 1,2-dibromoethane-degradative pathways. *J Bacteriol* **182**: 2191–2199.

Puglisi, E., Cahill, M.J., Lessard, P.A., Capri, E., Sinskey, A.J., Archer, J.A.C., and Boccazzi, P. (2010) Transcriptional response of *Rhodococcus* aetherivorans I24 to polychlorinated biphenyl-contaminated sediments. *Microb. Ecol.* **60**: 505–515.

Rahm, B.G., Morris, R.M., and Richardson, R.E. (2006) Temporal expression of respiratory genes in an enrichment culture containing *Dehalococcoides ethenogenes*. *Appl. Environ. Microbiol.* **72**: 5486–5491.

Ravikumar, S., Baylon, M.G., Park, S.J., and Choi, J. (2017) Engineered microbial biosensors based on bacterial two-component systems as synthetic biotechnology platforms in bioremediation and biorefinery. *Microb. Cell Factories* **16**:.

Ray, S., Gunzburg, M.J., Wilce, M., Panjikar, S., and Anand, R. (2016) Structural basis of selective aromatic pollutant sensing by the effector binding domain of MopR, an NtrC family transcriptional regulator. *ACS Chem. Biol.* **11**: 2357–2365.

Rivera-Gomez, N., Segovia, L., and Perez-Rueda, E. (2011) Diversity and distribution of transcription factors: their partner domains play an important role in regulatory plasticity in bacteria. *Microbiology* **157**: 2308–2318.

Roselli, S., Nadalig, T., Vuilleumier, S., and Bringel, F. (2013) The 380 kb pCMU01 plasmid encodes chloromethane utilization genes and redundant genes for vitamin B12- and tetrahydrofolate-dependent chloromethane metabolism in *Methylobacterium extorquens* CM4: A proteomic and bioinformatics study. *PLOS ONE* **8**: e56598.

Rupakula, A., Kruse, T., Boeren, S., Holliger, C., Smidt, H., and Maillard, J. (2013) The restricted metabolism of the obligate organohalide respiring bacterium *Dehalobacter* restrictus: lessons from tiered functional genomics. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120325–20120325.

Sallabhan, R., Kerdwong, J., Dubbs, J.M., Somsongkul, K., Whangsuk, W., Piewtongon, P., *et al.* (2013) The *hdhA* gene encodes a haloacid dehalogenase that is regulated by the LysR-type regulator, HdhR, in *Sinorhizobium meliloti*. *Mol. Biotechnol.* **54**: 148–157.

San Millan, A., Toll-Riera, M., Qi, Q., and MacLean, R.C. (2015) Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nat Commun* **6**: 6845.

Sanchez, M.A. and Gonzalez, B. (2007) Genetic characterization of 2,4,6-trichlorophenol degradation in *Cupriavidus necator* JMP134. *Appl. Environ. Microbiol.* **73**: 2769–2776.

Scheublin, T.R., Deusch, S., Moreno-Forero, S.K., Muller, J.A., van der Meer, J.R., and Leveau, J.H.J. (2014) Transcriptional profiling of Gram-positive *Arthrobacter* in the phyllosphere: induction of pollutant degradation genes by natural plant phenolic compounds. *Env. Microbiol* **16**: 2212–2225.

Schirmer, F., Ehrt, S., and Hillen, W. (1997) Expression, inducer spectrum, domain structure, and function of MopR, the regulator of phenol degradation in *Acinetobacter calcoaceticus* NCIB8250. *J. Bacteriol.* **179**: 1329–1336.

Schmid-Appert, M., Zoller, K., Traber, H., Vuilleumier, S., and Leisinger, T. (1997) Association of newly discovered IS elements with the dichloromethane utilization genes of methylotrophic bacteria. *Microbiol. Read. Engl.* **143 (Pt 8)**: 2557–2567.

Selifonova, O.V. and Eaton, R.W. (1996) Use of an *ipb-lux* fusion to study regulation of the isopropylbenzene catabolism operon of *Pseudomonas putida* RE204 and to detect hydrophobic pollutants in the environment. *Appl. Environ. Microbiol.* **62**: 778–783.

Sentchilo, V., Ravatn, R., Werlen, C., Zehnder, A.J.B., and van der Meer, J.R. (2003) Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. *J. Bacteriol.* **185**: 4530–4538.

Sipilä, T.P., Väisänen, P., Paulin, L., and Yrjälä, K. (2010) *Sphingobium* sp. HV3 degrades both herbicides and polyaromatic hydrocarbons using ortho- and meta-pathways with differential expression shown by RT-PCR. *Biodegradation* **21**: 771–784.

Smidt, H., van Leest, M., van der Oost, J., and de Vos, W.M. (2000) Transcriptional regulation of the *cpr* gene cluster in ortho-chlorophenol-respiring *Desulfitobacterium dehalogenans*. *J Bacteriol* **182**: 5683–5691.

Speed, M.C., Burkhart, B.W., Picking, J.W., and Santangelo, T.J. (2018) An archaeal fluoride-responsive riboswitch provides an inducible expression system for hyperthermophiles. *Appl Env. Microbiol* **84**: e02306-17.

Starr, L.M., Fruci, M., and Poole, K. (2012) Pentachlorophenol induction of the *Pseudomonas aeruginosa mexAB-oprM* efflux operon: involvement of repressors NalC and MexR and the antirepressor ArmR. *PLoS ONE* **7**: e32684.

Stierand, K. and Rarey, M. (2010) Drawing the PDB: Protein−ligand complexes in two dimensions. *ACS Med. Chem. Lett.* **1**: 540–545.

Su, X. and Tsang, J.S.H. (2013) Existence of a robust haloacid transport system in a *Burkholderia* species bacterium. *Biochim. Biophys. Acta BBA - Biomembr.* **1828**: 187–192.

Tabata, M., Ohhata, S., Nikawadori, Y., Kishida, K., Sato, T., Kawasumi, T., *et al.* (2016) Comparison of the complete genome sequences of four γ-hexachlorocyclohexane-degrading bacterial strains: insights into the evolution of bacteria able to degrade a recalcitrant man-made pesticide. *DNA Res* **23**: 581–599.

Takeda, H., Shimodaira, J., Yukawa, K., Hara, N., Kasai, D., Miyauchi, K., et al. (2010) Dual two-component regulatory systems are involved in aromatic compound degradation in a polychlorinated-biphenyl degrader, *Rhodococcus* jostii RHA1. *J. Bacteriol.* **192**: 4741–4751.

Tas, N., van Eekert, M.H.A., Wagner, A., Schraa, G., de Vos, W.M., and Smidt, H. (2011) Role of "*Dehalococcoides*" spp. in the aaerobic transformation of hexachlorobenzene in European rivers. *Appl. Environ. Microbiol.* **77**: 4437–4445.

Tittlemier, S.A., Blank, D.H., Gribble, G.W., and Norstrom, R.J. (2002) Structure elucidation of four possible biogenic organohalogens using isotope exchange mass spectrometry. *Chemosphere* **46**: 511–517.

Tobajas, M., Verdugo, V., Polo, A.M., Rodriguez, J.J., and Mohedano, A.F. (2016) Assessment of toxicity and biodegradability on activated sludge of priority and emerging pollutants. *Environ. Technol.* **37**: 713–721.

Torii, H., Machida, A., Hara, H., Hatta, T., and Takizawa, N. (2013) The regulatory mechanism of 2,4,6-trichlorophenol catabolic operon expression by HadR in *Ralstonia pickettii* DTP0602. *Microbiology* **159**: 665–677.

Trefault, N., Guzmán, L., Pérez, H., Godoy, M., González, B., *et al.* (2009) Involvement of several transcriptional regulators in the differential expression of *tfd* genes in *Cupriavidus necator* JMP134. *Int. Microbiol.* **12**: 97.

Tsai, W.-T. (2017) Fate of chloromethanes in the atmospheric environment: Implications for human health, ozone formation and depletion, and global warming impacts. *Toxics* **5**:.

T'Syen, J., Tassoni, R., Hansen, L., Sorensen, S.J., Leroy, B., Sekhar, A., *et al.* (2015) Identification of the amidase BbdA that initiates biodegradation of the groundwater micropollutant 2,6-dichlorobenzamide (BAM) in *Aminobacter* sp. MSH1. *Environ. Sci. Technol.* **49**: 11703–11713.

Türkowsky, D., Esken, J., Goris, T., Schubert, T., Diekert, G., Jehmlich, N., and von Bergen, M. (2018) A retentive memory of tetrachloroethene respiration in *Sulfurospirillum halorespirans* - involved proteins and a possible link to acetylation of a two-component regulatory system. *J. Proteomics*.

Türkowsky, D., Lohmann, P., Mühlenbrink, M., Schubert, T., Adrian, L., Goris, T., *et al.* (2018) Thermal proteome profiling allows quantitative assessment of interactions between tetrachloroethene reductive dehalogenase and trichloroethene. *J. Proteomics*.

Ul Haque, M.F., Nadalig, T., Bringel, F., Schaller, H., and Vuilleumier, S. (2013) A fluorescence-based bacterial bioreporter for the specific detection of methyl halide emissions in the environment. *Appl. Environ. Microbiol.* AEM.01738–13.

Villacieros, M., Whelan, C., Mackova, M., Molgaard, J., Sanchez-Contreras, M., Lloret, J., *et al.* (2005) Polychlorinated biphenyl rhizoremediation by *Pseudomonas fluorescens* F113 derivatives, using a *Sinorhizobium meliloti nod* system to drive *bph* gene expression. *Appl. Environ. Microbiol.* **71**: 2687–2694.

Wagner, A., Segler, L., Kleinsteuber, S., Sawers, G., Smidt, H., and Lechner, U. (2013) Regulation of reductive dehalogenase gene transcription in *Dehalococcoides mccartyi*. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120317–20120317.

Waller, A.S., Hug, L.A., Mo, K., Radford, D.R., Maxwell, K.L., and Edwards, E.A. (2012) Transcriptional analysis of a *Dehalococcoides*-containing microbial consortium reveals prophage activation. *Appl. Environ. Microbiol.* **78**: 1178–1186.

Wang, H., Hu, J., Xu, K., Tang, X., Xu, X., and Shen, C. (2018) Biodegradation and chemotaxis of polychlorinated biphenyls, biphenyls, and their metabolites by *Rhodococcus* spp. *Biodegradation* **29**: 1–10.

Wang, X., Teng, Y., Tu, C., Luo, Y., Greening, C., Zhang, N., *et al.* (2018) Coupling between nitrogen fixation and tetrachlorobiphenyl dechlorination in a rhizobium–legume symbiosis. *Environ. Sci. Technol.* **52**: 2217–2224.

Ward, W.O., Swartz, C.D., Porwollik, S., Warren, S.H., Hanley, N.M., Knapp, G.W., *et al.* (2007) Toxicogenomic analysis incorporating operon-transcriptional coupling and toxicant concentration-expression response: analysis of MX-treated *Salmonella*. *BMC Bioinformatics* **8**: 378.

Weir, K.M., Sutherland, T.D., Horne, I., Russell, R.J., and Oakeshott, J.G. (2006) A single monooxygenase, Ese, is involved in the metabolism of the organochlorides endosulfan and endosulfate in an *Arthrobacter* sp. *Appl. Environ. Microbiol.* **72**: 3524–3530.

Weng, X. and Xiao, J. (2014) Spatial organization of transcription in bacterial cells. *Trends Genet. TIG* **30**: 287–297.

Werner, J.J., Ptak, A.C., Rahm, B.G., Zhang, S., and Richardson, R.E. (2009) Absolute quantification of *Dehalococcoides* proteins: enzyme bioindicators of chlorinated ethene dehalorespiration. *Environ. Microbiol.* **11**: 2687–2697.

Whangsuk, W., Dubbs, J.M., Sallabhan, R., Somsongkul, K., Mongkolsuk, S., and Loprasert, S. (2010) ChpR Is a chlorpyrifos-responsive transcription regulator in *Sinorhizobium meliloti*. *J. Mol. Microbiol. Biotechnol.* **18**: 141–147.

White, P.M., Potter, T.L., and Culbreath, A.K. (2010) Fungicide dissipation and impact on metolachlor aerobic soil degradation and soil microbial dynamics. *Sci. Total Environ.* **408**: 1393–1402.

van den Wijngaard, A.J., van der Kamp, K.W., van der Ploeg, J., Pries, F., Kazemier, B., and Janssen, D.B. (1992) Degradation of 1,2-dichloroethane by *Ancylobacter aquaticus* and other facultative methylotrophs. *Appl. Environ. Microbiol.* **58**: 976–983.

van den Wijngaard, A.J., Reuvekamp, P.T., and Janssen, D.B. (1991) Purification and characterization of haloalcohol dehalogenase from *Arthrobacter* sp. strain AD2. *J. Bacteriol.* **173**: 124–129.

Wise, A.A. and Kuske, C.R. (2000) Generation of novel bacterial regulatory proteins that detect priority pollutant phenols. *Appl. Environ. Microbiol.* **66**: 163–169.

Wöhrnschimmel, H., Scheringer, M., Bogdal, C., Hung, H., Salamova, A., Venier, M., *et al.* (2016) Ten years after entry into force of the Stockholm convention: What do air monitoring data tell about its effectiveness? *Environ. Pollut. Barking Essex 1987* **217**: 149–158.

Xiu, Z., Gregory, K.B., Lowry, G.V., and Alvarez, P.J.J. (2010) Effect of bare and coated nanoscale zerovalent iron on *tceA* and *vcrA* gene expression in *Dehalococcoides* spp. *Environ. Sci. Technol.* **44**: 7647–7651.

Yadid, I., Rudolph, J., Hlouchova, K., and Copley, S.D. (2013) Sequestration of a highly reactive intermediate in an evolving pathway for degradation of pentachlorophenol. *Proc. Natl. Acad. Sci.* **110**: E2182–E2190.

Yan, Q., Philmus, B., Chang, J.H., and Loper, J.E. (2017) Novel mechanism of metabolic co-regulation coordinates the biosynthesis of secondary metabolites in *Pseudomonas protegens*. *eLife* **6**: e22835.

Yoshida, S., Ogawa, N., Fujii, T., and Tsushima, S. (2009) Enhanced biofilm formation and 3-chlorobenzoate degrading activity by the bacterial consortium of *Burkholderia* sp. NK8 and *Pseudomonas aeruginosa* PAO1. *J. Appl. Microbiol.* **106**: 790–800.

Zhang, H., Jiang, X., Xiao, W., and Lu, L. (2014) Proteomic strategy for the analysis of the polychlorobiphenyl-degrading cyanobacterium *Anabaena* PD-1 exposed to aroclor 1254. *PLoS ONE* **9**: e91162.

Zhang, H., Wan, H., Song, L., Jiang, H., Wang, H., and Qiao, C. (2010) Development of an autofluorescent *Pseudomonas nitroreducens* with dehydrochlorinase activity for efficient mineralization of γ-hexachlorocyclohexane (γ-HCH). *J. Biotechnol.* **146**: 114–119.

Zhang, Y., Wu, J.-F., Zeyer, J., Meng, B., Liu, L., Jiang, C.-Y., *et al.* (2009) Proteomic and molecular investigation on the physiological adaptation of *Comamonas* sp. strain CNB-1 growing on 4-chloronitrobenzene. *Biodegradation* **20**: 55–66.

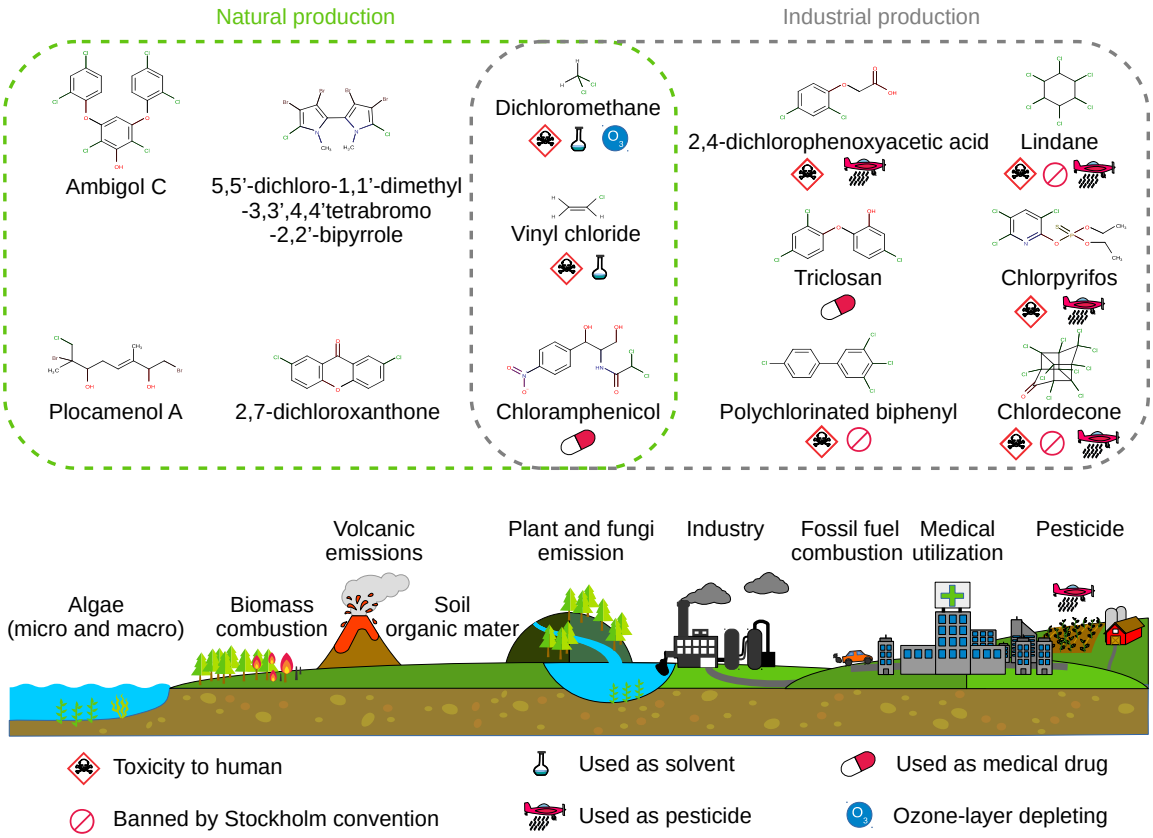


Fig. 1 Sources of organohalogen compounds production from natural or industrial processes. Some organohalogen molecules are naturally found in the environment as a result of abiotic production biomass combustion and volcano emissions (Gribble 2010), while biological emissions involve macro-algae and plants (Díaz-Marrero *et al*. 2002; Tittlemier *et al*. 2002). Some organohalogen compounds has been massively produced for use in agriculture as pesticides (Jayaraj, Megha and Sreedev 2016), or in industrial activities as refrigerants as solvents (Tsai 2017), as electrical insulators, and as polymers for plastic production.

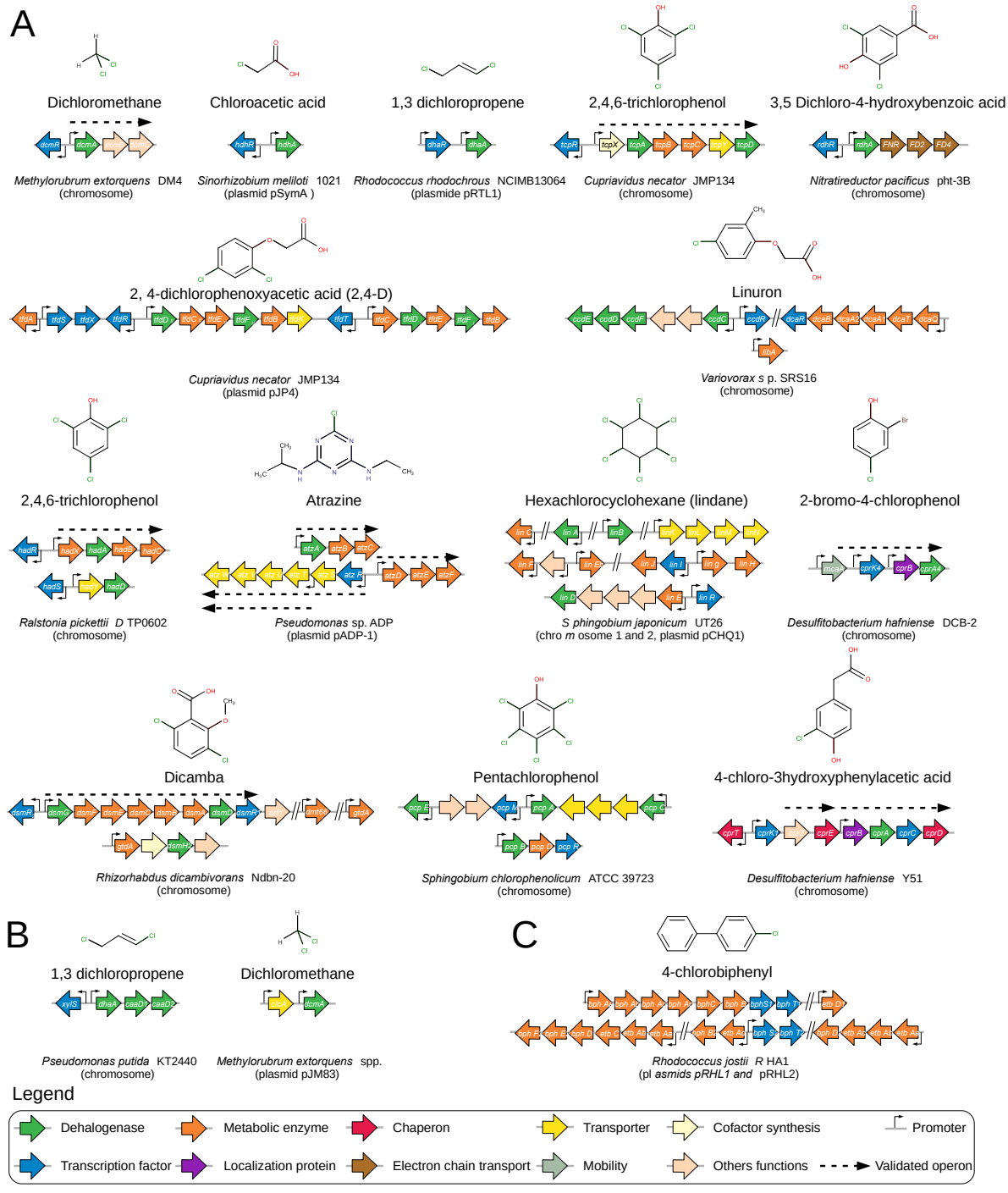


Fig. 2 Organohalogen compound chemical structure and corresponding regulated genetic modules in representative bacterial degraders. A) Dehalogenation genetic modules. B) Synthetic dehalogenation genetic modules. C) Catabolic genetic modules lacking dehalogenase-encoding genes. References are for genetic modules *atz* (Platero *et al*. 2012), *bph* (Takeda *et al*. 2010), *ccd* (Bers *et al*. 2011), *clc* (Coco, Parsek and Chakrabarty 1994), *cnb* (Zhang *et al*. 2009), *cpr* (Smidt *et al*. 2000), *dcm* (La Roche and Leisinger 1991), *dha* (Poelarends *et al*. 2000), *dsm* (Li *et al*. 2017), *lin* (Tabata *et al*. 2016), *pcp* (Cai and Xun 2002), *tcp* (Sanchez and Gonzalez 2007) and *tfd* (Trefault *et al*. 2009).

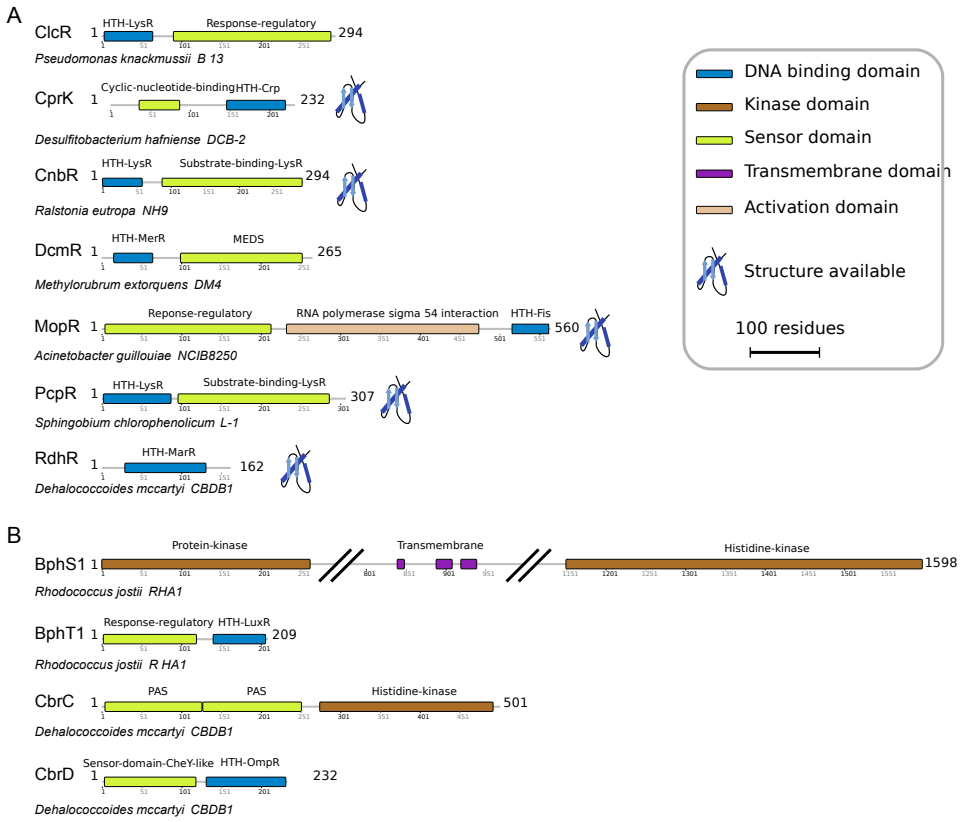


Fig. 3 Functional domains of a selection of organohalogen compound regulatory proteins. A) Transcription factors; B) Proteins of two component systems. HTH (helix-turn-helix) DNA binding domain (in blue) present in almost all transcription factors are fused to disimilar sensor domains (in green). Abbreviations: MEDS (MEthanogen / methylotroph DcmR Sensory; (Anantharaman and Aravind 2005)) and PAS (Per sensor domain, Arnt, Sim; (Möglich, Ayers and Moffat 2009)). Accession number for BphS1 (WP_011598994.1), BphT1 (WP_011598993.1), CbrC (CAI82343.1), CbrD (CAI82344.1), ClcR (WP_012248462), CprK (ACL18797.1), DcmR (AAB68953.1), PcpR (AAA68939), and RdhR (WP_011309983).

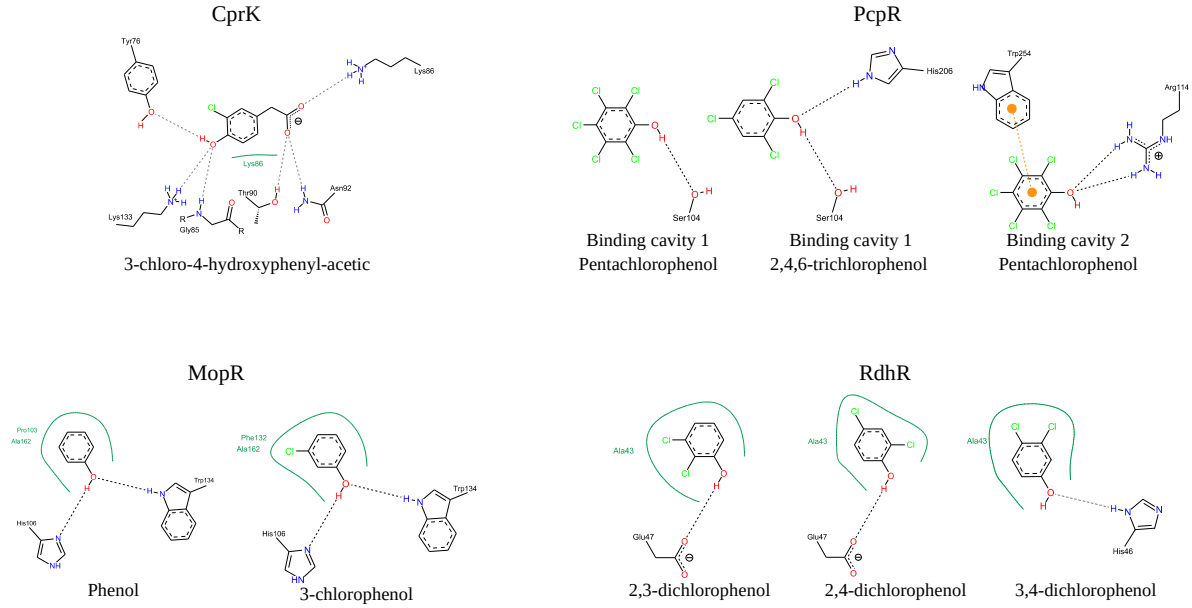


Fig. 4 Molecular interactions between organohalogen pollutants and the ligand cavity of transcription factors. For 2D representation, PoseView software (Stierand and Rarey 2010) was used. Hydrogen bounds and hydrophobic interactions are represented by dashed black and green lines, respectively. Interactions between aromatic rings (also named π-π stacking) are shown with orange dashed lines.

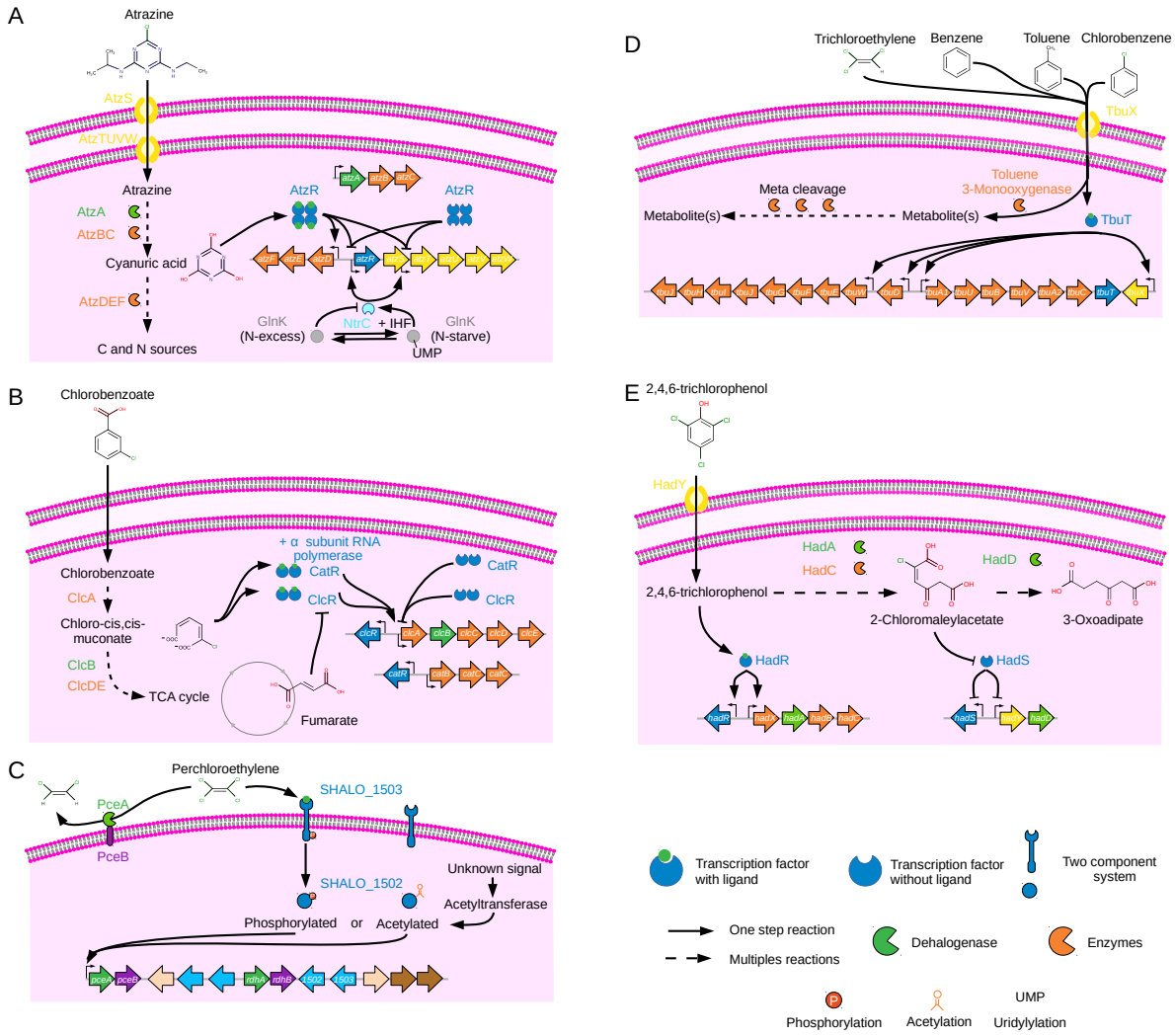


Fig. 5 Examples of transcriptional regulation mechanisms for organohalogen compounds degradation pathways. A) Atrazine metabolism in *Pseudomonas* sp. ADP. The expression of *atzDEF* is activated when AtzR binds to cyanuric acid. The expression of *atzR* and downstream genes at low and high nitrogen concentrations is activated and repressed, respectively (Platero *et al.* 2012). B) Chlorobenzoate metabolism in *Pseudomonas putida* PRS2000 is regulated by two homologous transcription regulators CatR and ClcR (McFall *et al*. 1997; McFall, Chugani and Chakrabarty 1998). C) PCE metabolism in *Sulfurospirillum multivorans.* A two-component system will activate expression of *pceAB* when PCE is detected. After exposure to PCE, an unknown signal activates an acetyltransferase that acetylates this two component system (Türkowsky *et al*. 2018a). D) Trichloroethylene metabolism in *Ralstonia pickettii* PKO1. TbuT transcription factor recognizes several organic and organochlorine compounds with increased expression of *tbuT* and catabolic

genes (Kahng *et al.* 2000). E) 2,4,6-trichlorophenol is metabolised by *Ralstonia pickettii* DTP0602 with proteins encoded by two regulons (*hadRXABC* and *hadSYD*) separated by 146 kb on the genome (Torii *et al*. 2013). These two regulons are differentiallly regulated and involve two transcription factors. HadR binds to 2,4,6-trichlorophenol for induced expression of *hadYABC*. HadS recognizes an intermediate metabolite, probably 2-chloromaleylacetate, to alleviate repression of *hadYD* (Torii *et a*l. 2013).

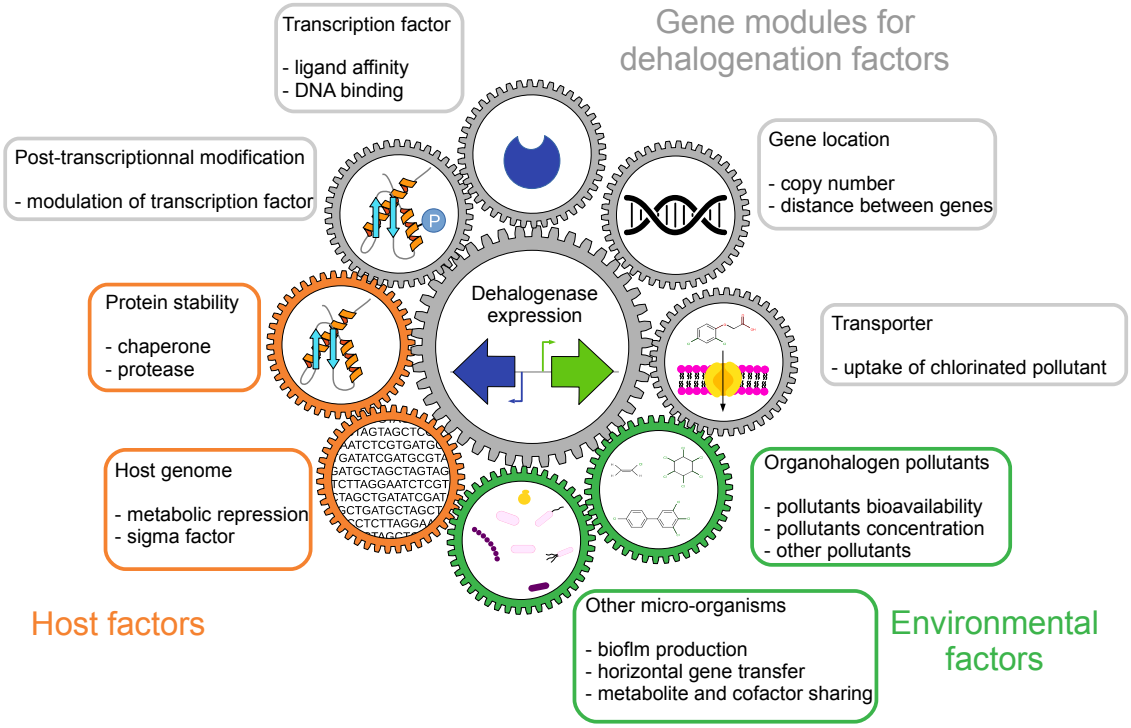


Fig. 6 Summary of factors known to modulate dehalogenase expression of organohalogen compounds degradation

Table 1: Transcription factors involved in organohalogen pollutant detection

Transcription factor	Pathway of degradation	Ligand recognized	DNA sequence recognized	Crystalographic structure	Organism	Reference
Directly organochlorine pollutants recognized as ligands						
BphS and BphR	Polychlorinated biphenyl	Biphenyl; dichlorobenzene [A]	TTCCGTACGTTTTCCCGGATGTTCG	No	*Rhodococcus jostii* RHA1	(Takeda *et al.* 2010)
HadR	2,4,6-trichlorophenol	2,4,6-trichlorophenol	ATGCCGCTGAGGAAT	No	*Ralstonia pickettii* DTP0602	(Torii *et al.* 2013)
MopR	3-chlorophenol	3-chlorophenol [B]	TTCATCAAATAATGGA -8nt- ATGCTGATTCATCAA	Yes	*Acinetobacter calcoaceticus*	(Ray *et al.* 2016)
PcpR	Pentachlorophenol; 2,4,6-trichlorophenol	Pentachlorophenol [B]; 2,4,6-trichlorophenol [B]	ATTC-7nt-GAAT	Yes	*Sphingobium chlorophenolicum*	(Hayes *et al.* 2014)
Metabolites products recognized as ligands						
AtzR	Atrazine	Cyanuric acid	GGTGCCG -3 nt- CGGCACC	No	*Pseudomonas* sp. ADP	(Platero *et al.* 2012)
ClcR	Chlorobenzoate	Chloro-cis,cis-muconate	TCCATCAGACCTCCAGGGTATGGTGG	No	*Pseudomonas knackmussii* B13	(McFall, Chugani and Chakrabarty 1998)
CbnR	3-chlorobenzoate	cis,cis-muconate	Unknown	Yes	*Ralstonia eutropa* NH9	(Muraoka *et al.* 2003)
CprK	Chlorophenols	orthochlorophenolacetic acid [B]	TTTA -8nt- CTAA	Yes	*D. hafniense*	(Joyce *et al.* 2006)
LinR	Lindane	2,5-dichlorohydroquinone; chlorohydroquinone	ATTCACAATCTGAAT	No	*D. dehalogenans* *Sphingomonas paucimobilis* UT26	(Miyauchi *et al.* 2002)
RdhR	Trichlorobenzene	2,3-dichlorophenol [B]; 2,4-dichlorophenol [B]; 3,4-dichlorophenol [B]	TAGTATACGAT -2nt- TAGTATACGATTA	Yes	*Dehalococcoides mccartyi* CBDB1	(Krasper *et al.* 2016)
Unknown ligands						
CbrC	Trichlorobenzene	Unknown	Unknown	No	*Dehalococcoides mccartyi* CBDB1	(Adrian *et al.* 2007)
DcmR	Dichloromethane	Unknown	Unknown	No	*Methylorubrum extorquens* DM4	(La Roche and Leisinger 1991)

[A] Other ligands are described in Takeda *et al.* 2010.

[B] ligand co-crystallized with the transcription factor

Table 2: Differentially regulated genes in bacteria utilizing organohalogen compounds for growth.

Bacteria	Substrates		Culture condition	Number of genes [A]		Method	Reference
	Halogenated pollutant	Reference substrate		up-regulated	down-regulated		
Arthrobacter chlorophenolicus A6	4-Chlorophenol	tryptone soy agar gelose	*in labo* (agar plate)	569 / 4,255	580 / 4,255	Micro-array	(Scheublin *et al.* 2014)
		N.A.	Microcosm with plant (phyllosphere low humidity)	390 / 4,255	412 / 4,255		
		N.A.	Microcosm with plant (phyllosphere high humidity)	430 / 4,255	474 / 4,255		
Burkholderia xenovorans LB400	Aroclor 1242 [B]	Biphenyl	*in labo* (liquid culture)	28 / 8,249	19 / 8,429	RNA-seq	(Parnell *et al.* 2006)
Dehalococcoides spp. (*Culture* KB1™)	Vinyl chloride	Methanol	Microcosm (liquid culture)	452/19,200 [C]	95/19,200 [C]	Micro-array	(Waller *et al.* 2012)
Desulfitobacterium hafniense DCB-2	3-chloro-4-hydroxybenzoate 3,5-dichlorophenol	Ortho-bromophenol, uranyl acetate, sodium selenate, ferric citrate	*in labo* (liquid culture)	N.A.	N.A.	Micro-array	(Kim *et al.* 2012)
Desulfitobacterium hafniense Y51	Trichloroethylene	MMYPF-medium	*in labo* (liquid culture)	24 / 5,060	N.A. / 5,060	Micro-array	(Peng *et al.* 2012)
Methylobacterium extorquens CM4	Chloromethane	Methanol	*in labo* (liquid culture)	137 / 6,262	13 / 6,262	RNA-seq	(Chaignaud *et al.* 2017)
Methylobacterium extorquens DM4	Dichloromethane	Methanol	*in labo* (liquid culture)	69 / 5,701	121 / 5,701	RNA-seq	(Chaignaud *et al.* 2017)
Nitrosomonas europaea	Chloromethane	Mineral medium + CO_2	*in labo* (liquid culture)	67 / 2,460	148 / 2,460	Micro-array	(Gvakharia *et al.* 2007)
	Chloroform	Mineral medium + CO_2	*in labo* (liquid culture)	175/ 2,460	501 / 2,460		
Pigmentiphaga sp. H8	3,5-Dibromo-4-Hydroxybenzoate	Non induced culture	*in labo* (liquid culture)	98 / 5,678	31 / 5,678	RNA-seq	(Chen *et al.* 2018)
Polaromonas sp. JS666	cDCE	Glycolate	*in labo* (liquid culture)	140 / 5,569	450 / 5,569	Micro-array & proteomic	(Jennings *et al.* 2009)
Rhodococcus aetherivorans I24	aroclor 1254 [A]	Glucose	*in labo* (liquid culture)	129 / 3,524	87 / 3,524	Micro-array	(Puglisi *et al.* 2010)
			Microcosm (sediments; contaminated Rotterdam harbor)	207 / 3,524	176 / 3,524		
			Microcosm (sediments; uncontaminated Rotterdam harbor)	44 / 3,524	67 / 3,524		
Salmonella typhimurium TA100	MX chlorinated furanone	dimethyl sulfoxide	*in labo* (liquid culture)	169 / 4,253		Micro-array	(Ward *et al.* 2007)

[A] Number of genes regulated / total number of genes in genome.

[B] PCB (poly-chlorinated biphenyl).

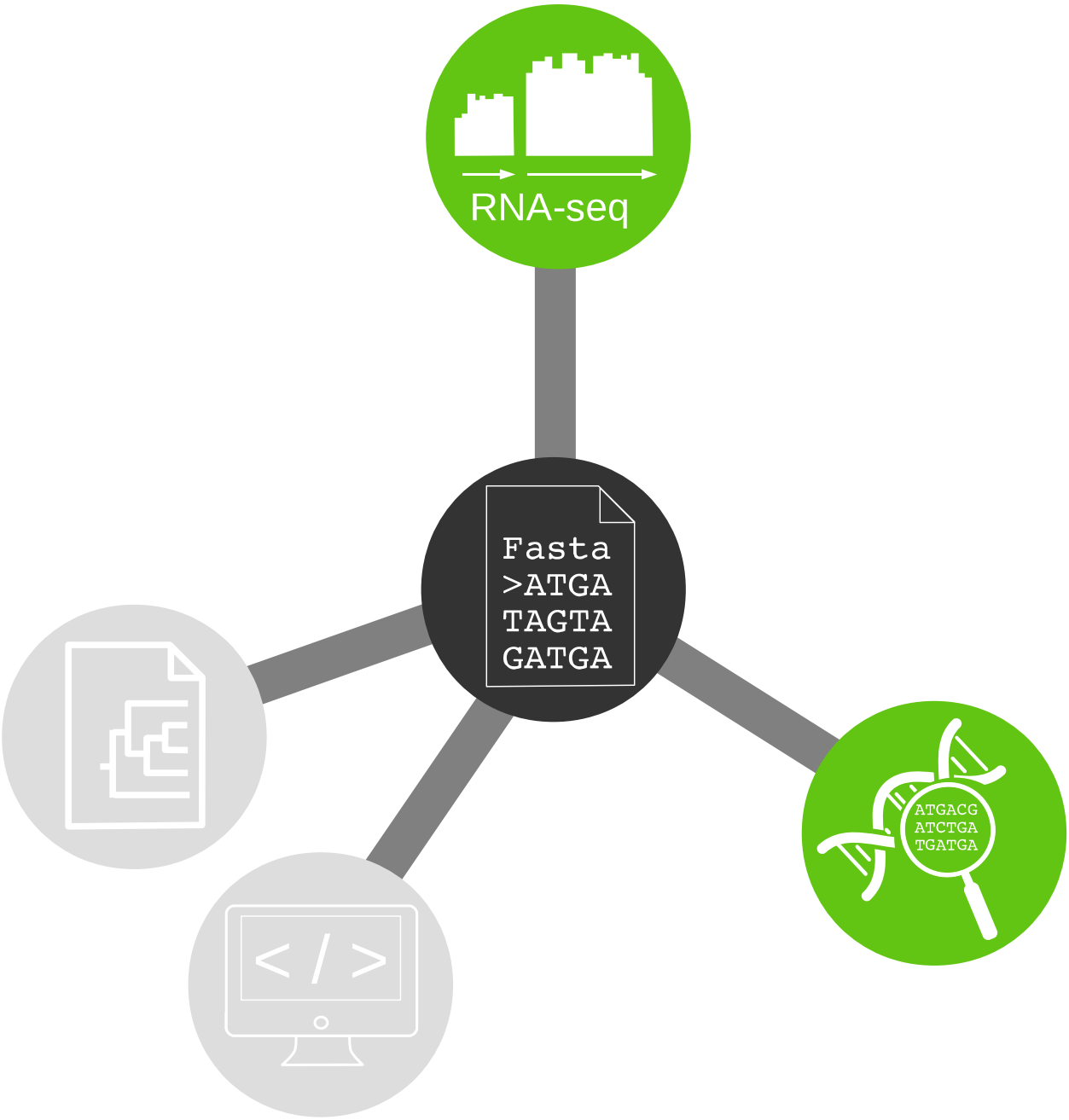
[C] Correspond to number of spots on micro-array with differential hybridization and not number of genes.

Table 3: Differentially regulated genes in bacteria in pure culture or consortium

Halogenated pollutant	Culture condition	Dehalogenating bacteria	Number of genes[A]		Other bacteria	Method	Reference
			up-regulated	down-regulated			
Linuron, 3,4-dichloroaniline	*in labo* (liquid culture)	*Variovorax* sp WDL1	1372		*Comamonas testosteroni* WDL7 *Hyphomicrobium sulfonivorans* WDL	RNA-seq	(Albers *et al.* 2018)
Trichloroethylene	*in labo* (liquid culture)	*Dehalococcoides mccartyi* 195	102		*Desulfovibrio vulgaris* Hildenborough *Methanobacterium congolense*	Micro-array	(Men *et al.* 2012)
Trichloroethylene	*in labo* (liquid culture with vitamin B12)	*Dehalococcoides mccartyi* 195	28	18	*Desulfovibrio vulgaris* Hildenborough *Pelosinus fermentans* R7	Micro-array	(Men *et al.* 2014)
Trichloroethylene	*in labo* (liquid culture without vitamin B12)	*Dehalococcoides mccartyi* 195	30	44	*Desulfovibrio vulgaris* Hildenborough *Pelosinus fermentans* R7	Micro-array	(Men *et al.* 2014)
Trichloroethylene	Microcosm (contaminated groundwater with or without cobalamin)	7 bin genomes	550		7 bin genomes	RNA-seq & micro-array	(Men *et al.* 2017)
Trichloroethylene	*in labo* (liquid culture)	*Dehalococcoides mccartyi* 195	18	196	*Syntrophomonas wolfei*	Micro-array	(Mao *et al.* 2015)

[A]Some publications do not provide the information of the number of up- and down-regulated genes. NA, Not Applicable, when no numbers at all were provided.

3. Une vision globale de l'adaptation au DCM par des approches « -omiques »



Chapitre 3 : Une vision globale de l'adaptation au DCM par des approches « -omiques »

3.1 Introduction générale

Pour faire face aux stress, aux réajustements métaboliques et énergétiques associés à la dégradation du DCM, il est supposé que l'expression de certains gènes de *M. extorquens* DM4 soit régulée (comme par exemple des gènes de réparation de l'ADN, d'excrétion de chlorure et de protons, ou de synthèse du glutathion, le cofacteur de l'enzyme DCM déshalogénase).

Un premier niveau de réponse adaptative rapide implique une régulation par modification de l'expression génétique par des facteurs de transcription reconnaissant des motifs présents en amont des gènes régulés ou de promoteurs alternatifs (Di Martino *et al.*, 2016), puis au niveau des ARNs, des protéines et des métabolites. Appréhender la complexité de ces régulations à différentes échelles peut se faire d'abord par une évaluation globale de l'expression, de l'ensemble du génome par des approches de transcriptomique, de protéomique et de métabolomique. Dans le cadre de de cette thèse, une analyse des résultats de RNA-seq (Chaignaud *et al.*, 2017) et de protéomique différentielle (Bibi-Triki *et al.*, 2018) a été réalisé pour la souche sauvage *M. extorquens* DM4. L'expression génomique en condition de croissance par l'utilisation du DCM a été comparée à celle du méthanol qui est le composé de référence du métabolisme d'utilisation de composés en C1 réduit et non chloré.

Ce chapitre sera axé sur trois grands aspects d'étude globale par des approches « -omiques », dont deux ont déjà donné lieu à des publications. Les différents sous-chapitres incluent les thématiques suivantes :

1. Analyse des transcrits (identification et quantification des transcrits) par approche RNA-seq et la présentation de l'article de Chaignaud *et al*. paru en 2017 ;
2. Analyse des protéines (annotation des N-ter, identification et quantification des protéines) par approche protéomique, notamment de type N-terminome et la présentation de l'article de Bibi-Triki *et al*. paru en 2018 ;

3. Intégration et analyse des premières quantifications à l'échelle globale des transcrits et des protéines en réponse à la croissance et la déshalogénation du DCM ;
4. Analyse des sites d'initiation de la transcription (détection et identification) par approche ciblée de TSS-seq. Cette analyse n'a pas encore été publiée. L'analyse des données « brutes » des séquences a été réalisée lors de cette thèse ;
5. Conclusions et perspectives

3.2 Étude de la transcription par RNA-seq chez *M. extorquens* DM4 durant la croissance avec le DCM

3.2.1 Article de Chaignaud *et al*.

L'article de Chaignaud *et al*., 2017 publié dans le journal *Frontiers in microbiology* se trouve dans les pages suivantes. Dans le cadre de cette thèse des analyses de RT-qPCR ont été réalisées pour étudier l'expression de plusieurs gènes puis les comparer avec celles de RNA-seq.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 September 2017
doi: 10.3389/fmicb.2017.01600

Genomic and Transcriptomic Analysis of Growth-Supporting Dehalogenation of Chlorinated Methanes in *Methylobacterium*

Pauline Chaignaud[1,2], *Bruno Maucourt*[1], *Marion Weiman*[3], *Adriana Alberti*[3], *Steffen Kolb*[2,4], *Stéphane Cruveiller*[3], *Stéphane Vuilleumier*[1] *and Françoise Bringel*[1]*

[1] *Department of Molecular Genetics, Genomics, and Microbiology, UMR 7156 Université de Strasbourg (UNISTRA)-Centre National de la Recherche Scientifique, Strasbourg, France,* [2] *Department of Ecological Microbiology, University of Bayreuth, Bayreuth, Germany,* [3] *UMR 8030 Centre National de la Recherche Scientifique-CEA, DSV/IG/Genoscope, LABGeM, Evry, France,* [4] *Institute of Landscape Biogeochemistry-Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany*

OPEN ACCESS

Edited by:
Ivan Berg,
Universität Münster, Germany

Reviewed by:
Nicholas Coleman,
University of Sydney, Australia
Jeremy Semrau,
University of Michigan, United States

***Correspondence:**
Françoise Bringel
francoise.bringel@unistra.fr

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *25 June 2017*
Accepted: *07 August 2017*
Published: *01 September 2017*

Citation:
Chaignaud P, Maucourt B, Weiman M, Alberti A, Kolb S, Cruveiller S, Vuilleumier S and Bringel F (2017) Genomic and Transcriptomic Analysis of Growth-Supporting Dehalogenation of Chlorinated Methanes in Methylobacterium. Front. Microbiol. 8:1600. doi: 10.3389/fmicb.2017.01600

Bacterial adaptation to growth with toxic halogenated chemicals was explored in the context of methylotrophic metabolism of *Methylobacterium extorquens*, by comparing strains CM4 and DM4, which show robust growth with chloromethane and dichloromethane, respectively. Dehalogenation of chlorinated methanes initiates growth-supporting degradation, with intracellular release of protons and chloride ions in both cases. The core, variable and strain-specific genomes of strains CM4 and DM4 were defined by comparison with genomes of non-dechlorinating strains. In terms of gene content, adaptation toward dehalogenation appears limited, strains CM4 and DM4 sharing between 75 and 85% of their genome with other strains of *M. extorquens*. Transcript abundance in cultures of strain CM4 grown with chloromethane and of strain DM4 grown with dichloromethane was compared to growth with methanol as a reference C_1 growth substrate. Previously identified strain-specific dehalogenase-encoding genes were the most transcribed with chlorinated methanes, alongside other genes encoded by genomic islands (GEIs) and plasmids involved in growth with chlorinated compounds as carbon and energy source. None of the 163 genes shared by strains CM4 and DM4 but not by other strains of *M. extorquens* showed higher transcript abundance in cells grown with chlorinated methanes. Among the several thousand genes of the *M. extorquens* core genome, 12 genes were only differentially abundant in either strain CM4 or strain DM4. Of these, 2 genes of known function were detected, for the membrane-bound proton translocating pyrophosphatase HppA and the housekeeping molecular chaperone protein DegP. This indicates that the adaptive response common to chloromethane and dichloromethane is limited at the transcriptional level, and involves aspects of the general stress response as well as of a dehalogenation-specific response to intracellular hydrochloric acid production. Core genes only differentially abundant in either strain CM4 or strain DM4 total 13 and 58

CDS, respectively. Taken together, the obtained results suggest different transcriptional responses of chloromethane- and dichloromethane-degrading *M. extorquens* strains to dehalogenative metabolism, and substrate- and pathway-specific modes of growth optimization with chlorinated methanes.

Keywords: dehalogenation, chloromethane, dichloromethane, GEI, genomic island, genome adaptation, methylotrophy

INTRODUCTION

Chlorinated one-carbon compounds chloromethane (CM, CH_3Cl) and dichloromethane (DCM, CH_2Cl_2) are toxic chemicals that occur in both pristine and polluted environments. CM is the most abundant halogenated trace gas in the atmosphere, where it contributes to ozone destruction (Carpenter et al., 2014). It is mainly produced by vegetation (Derendorp et al., 2012; Hardacre and Heal, 2013; Rhew et al., 2014; Bringel and Couée, 2015). DCM is one of the most commonly manufactured chlorinated chemicals. It is used for its solvent properties, and is a frequently found contaminant at polluted sites. Some methylotrophic microorganisms are able to utilize chlorinated methanes (such as CM and DCM) as their sole carbon and energy source for growth (Muller et al., 2011a; Nadalig et al., 2014). Bacterial CM and DCM utilization starts with dehalogenation, causing diverse physiological stresses such as production of intracellular HCl, which lowers pH and increases ionic strength, and formation of DNA adducts (Kayser and Vuilleumier, 2001; Torgonskaya et al., 2011; Michener et al., 2014b, 2016). The mechanisms allowing methylotrophic bacteria to cope with dehalogenation-induced stress are still poorly understood, despite potential use of such bacteria for bioremediation.

The aerobic alphaproteobacterium *Methylobacterium extorquens* is the most extensively studied methylotroph. It is found in a wide variety of habitats, including plants, soil, wastewater, and clouds (Amato et al., 2007; Kolb, 2009; Bringel and Couée, 2015). The well-studied *M. extorquens* strains CM4 and DM4, whose genomes have been sequenced (Vuilleumier et al., 2009; Marx et al., 2012), utilize the chlorinated compounds CM and DCM, respectively, as their sole source of carbon and energy. Degradation pathways for CM and DCM have been characterized in these strains, and are also found in many other species (Muller et al., 2011a; Nadalig et al., 2014). The ability to grow on chlorinated methanes has been demonstrated by genetic, biochemical and recent experimental evolution studies to require the expression of essential dehalogenation-associated genes that differ for CM and DCM. To our knowledge no strains able to metabolize both CM and DCM has been isolated so far. Genes *cmuA* and *cmuB* are essential for CM dehalogenation by the *cmu* (CM-utilization) pathway (Vannelli et al., 1999). The two-domain methyltransferase/corrinoid-binding CmuA protein catalyzes methyl transfer from CM to a cobalt-corrin cofactor (Studer et al., 2001), and the methylcobalamin:tetrahydrofolate methyltransferase CmuB transfers the resulting corrinoid-bound methyl group to tetrahydrofolate (H_4F) (Studer et al., 1999). For each molecule of CM, one methyl-H_4F and one HCl are produced. For DCM degradation, a glutathione-dependent dehalogenase of the glutathione *S*-transferase family encoded by *dcmA* converts DCM into formaldehyde and two molecules of HCl (Vuilleumier and Leisinger, 1996; Kayser et al., 2002).

Although protons and chloride ions are produced in both cases, the processing of carbon from chlorinated methanes for production of biomass and energy proceeds differently in CM- and DCM-utilizing *M. extorquens* strains (Michener et al., 2016). Thus, toxic chlorinated methanes CM and DCM may generate both similar and compound- and pathway-specific adaptive responses. In this work, we analyzed these responses in terms of global gene expression, by sequencing cDNA libraries of *M. extorquens* strains grown either with CM or with DCM.

Abbreviations: CM, chloromethane; DCM, dichloromethane; GEI, genomic island; H_4F, tetrahydrofolate.

MATERIALS AND METHODS

Bacterial Cultivation and RNA Purification

Strains CM4 and DM4 were grown aerobically in 1.2L Erlenmeyer flasks closed with gas-tight screw caps with Mininert® valves (Supelco) in *Methylobacterium* mineral medium (M3) (modified as described in Roselli et al., 2013), with shaking (120 rpm) at 30°C. For 220-mL cultures, one-carbon growth substrates were supplied at 10 mM final concentration, by adding either aqueous solutions of 2.75 mL of filter-sterilized aqueous stock solution of methanol (800 mM), 141 μL of neat dichloromethane, or 40 mL of gaseous chloromethane (assuming a Henry constant of 0.0106 $m^3 \cdot atm \cdot mol^{-1}$ at 30°C) (Chen et al., 2012). Upon reaching mid-exponential phase (OD_{600} ~0.15), growth was stopped by addition of 27.5 mL stabilization buffer. This buffer was prepared by mixing 5 mL of phenol and 5 mL of 1 M sodium acetate pH 5.5, then after centrifugation at 1,800 g for 3 min, 5 mL of the lower phase was mixed with 95 mL of absolute ethanol. Resulting cell suspensions were centrifuged at 5,000 rpm for 5 min at 4°C, and suspended in 10 mL TE containing 2 $mg \cdot mL^{-1}$ lysozyme (Euromedex). After 15 min incubation at 37°C, each cell suspension was centrifuged 10 min at 4°C, the obtained pellet resuspended in 10 mL of Trizol (Invitrogen), and 2.5 mL of chloroform was added. RNA was precipitated with isopropanol and washed with ethanol (70%), then resuspended in DEPC-water and treated with DNAse (Turbo DNAse, Invitrogen). DNA depletion was checked by PCR (see Table S1 for primers). RNA quality was checked with the RNA 6000 Nano kit on a Bioanalyzer 2100 (Agilent Technologies), and quantified with the Qubit RNA kit (Invitrogen). Depletion of rRNA was obtained by treating 5 μg

of total RNA with the Gram-negative RiboZero Magnetic kit (Tebu-Bio) according to the manufacturer's protocol.

cDNA Library Preparation, Sequencing and Data Normalization

rRNA-depleted RNA (50–60 ng in 5 μL) were fragmented by adding 13 μL of the "Fragment, prime, finish mix" of the "Purify and fragment mRNA" kit (Illumina). Then, cDNA libraries were constructed with the TruSeq stranded mRNA LT kit (Illumina) following the manufacturer's protocol. Obtained cDNA libraries were quality checked (Bioanalyzer, DNA 1000 kit, Agilent Technologies) prior to HiSeq2000 sequencing. Paired-end sequence mapping was used to limit mapping artifacts and remove mapping ambiguities between gene paralogs. Factor size normalization of the raw counts was performed using a set of 55 reference "housekeeping" genes instead of the standard method using the complete set of CDS (Anders et al., 2013), as a few genes had very high read numbers in growth with chlorinated methanes only. The 55 reference genes cover a wide range of expression levels but had unchanged transcript abundance in the tested biological replicates ($\log_2$fc value between −0.9 and 1.2) (Table S2). Normalization was validated by comparison with the complete set of CDS for the methanol dataset (Figure S1). Gene transcripts were defined as differentially abundant when the $\log_2$ of fold-change values ($\log_2$fc) between cultures grown with chlorinated methanes and methanol were $\geq$2.0 or ≤ -2 (Yang et al., 2015). Each condition was analyzed in duplicate, with the average of read sense and antisense for each condition. Data were considered significant when False Discovery Rate (i.e., FDR) was $\leq$0.1 as previously described (Benjamini and Hochberg, 1995). The complete RNA-Seq dataset is accessible online (https://www.genoscope.cns.fr/agc/microscope/transcriptomic/NGSProjectRNAseq.php?projType$=$RNAseq).

RT-qPCR

Retrotranscription was performed from 2.0 μg of DNA-depleted RNA extracted using the Nucleospin RNA plus kit (Macherey Nagel) combined with Turbo DNase and Turbo DNA free kit (Invitrogen) using Superscript III (Invitrogen) and random hexanucleotide primers (Invitrogen) following manufacturers' protocols, except that RNaseOUT™ was replaced by RNasin™ (40 U·μL^{-1}; Invitrogen), with inclusion of appropriate controls (reactions without RNA template or Superscript III enzyme). Amplification was performed in qPCR 96 wells plates (Agilent Technologies) covered with Microseal B adhesive seals (Bio-Rad) using the Thermocycler Stratagene Mx3005P (MxPro software v4.10; Agilent Technologies). Master mix containing 7.5 μL Brilliant III SYBR™ Green low ROX qPCR master mix (Agilent Technologies), 0.225 μL of each primer (Eurofins) at 20 μM and 3.3 μL molecular biology grade water was mixed with 3.75 μL template cDNA. After 3 min pre-denaturation at 95°C, 40 cycles of 20 s at 95°C and 20 s at 60°C, a final one-cycle step of 1 min at 95°C with ramping from 60 to 95°C was applied to obtain dissociation curves for quality control of PCR products. For each biological triplicate, technical duplicates were analyzed and compared to standard curves with gDNA ($0.1–1{\cdot}10^{-5}$ ng·μL^{-1}) and a no template control. SYBR™ green fluorescence data normalized with Rox fluorescence were analyzed using the package qpcR (v1.40) within R for Ct determination (Ritz and Spiess, 2008). Fold change values were calculated by the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008), with *rrsA* as internal control and *dcmA* in DM4 cultures grown with methanol as the reference condition.

Comparative Genomics Analysis

Comparative analysis of *M. extorquens* genomes was done in MaGe on the Genoscope MicroScope online platform (Vallenet et al., 2017), with the sequenced genomes of 5 strains of *M. extorquens,* i.e., the 2 dehalogenating strains CM4 (Genbank accession numbers CP001298, CP001299, CP001300) and DM4 (FP103042, FP103043, and FP103044), and 3 other strains AM1 (CP001511, CP001512, CP001513, and CP001514), PA1 (CP000908), and BJ001 (CP001029, CP001030, and CP001031) (Vuilleumier et al., 2009; Marx et al., 2012). All studied strains contain at least one plasmid, except for PA1. Proteins responsible for dehalogenation of CM or DCM are exclusively encoded by *M. extorquens* CM4 and DM4 genomes, respectively. The common genome called hereafter "core" was operationally identified using the MaGe Pan/Core-genome tool (https://www.genoscope.cns.fr/agc/microscope/compgenomics/pancoreTool.php?), by defining shared CDS as encoding proteins displaying at least 80% amino acid identity over 80% of CDS length in all considered genomes. Other genes were assigned either to the variable genome when present in at least 2 genomes, to the dehalogenation-associated genome when shared only by strains CM4 and DM4, and to the strain-specific genome when found only in one genome, respectively. Genomic islands of at least 5 kb were defined using the MaGe "Regions of Genomic Plasticity" tool (https://www.genoscope.cns.fr/agc/microscope/compgenomics/genomicIsland.php?) with the genomes of the 5 strains above, and applying a specificity score cutoff of 40 for all compared genomes.

RESULTS

Acquisition of specific dehalogenase genes by *M. extorquens* does not necessarily lead to growth with chlorinated methanes (Kayser et al., 2002; Michener et al., 2014a,b, 2016). Adaptation may require other specific genes associated with dehalogenative pathways, as well as modulation of expression of common "household" genes to optimize metabolic flux and responses to dehalogenation-associated stresses. In this study, the relative contribution of core, variable and strain-specific genomes in *M. extorquens* strains growing with halogenated methanes was investigated using a combination of comparative genomics and transcriptomics.

The Potentially Dehalogenation-Associated Genome of *M. extorquens* Is Limited

Beyond specific genes associated with dehalogenation [at least 6 genes of the *cmu* pathway for strain CM4 (Michener et al., 2016), and 4 genes of the *dcm* islet for strain DM4 (Muller et al., 2011a, respectively)], few strain-specific genes shared by

CM- and DCM-dehalogenating strains were identified through comparison with 3 other high quality assembled genomes from *M. extorquens* (**Figure 1A**). The strains AM1, PA1, and BJ001 were experimentally checked for their inability to grow on chloromethane or dichloromethane (data not shown). The gene content in the 5 *M. extorquens* genomes totals 12,273 unique CDS, representing 12.3 Mb. The core genome shared by all 5 *M. extorquens* genomes is extensive (3,489 CDS) ranging from 55 to 68% (55% for CM4; 61% for DM4). *M. extorquens* strains CM4 and DM4 share most of their genome content (75%, 4,424 CDS) and extensive gene synteny (**Figure 1B**). The specific genome for strain CM4 comprises 1,512 CDS (24% of total), and that of strain DM4 952 CDS (17%), respectively. In contrast, the variable genome shared only by the 2 dehalogenating strains was only 163 CDS, representing less than 3% of their total genome size, and in the range of the shared variable genome for any pair of the 5 strains considered (**Figure 1A**). A majority (97 CDS) had no predicted function and 42 genes were found in synteny on plasmids pCMU01 in CM4 and plasmid p1METDI in DM4 (Table S3).

Common Transcriptomic Response of *M. extorquens* Strains to Chlorinated Methanes as Growth Substrates

Methylobacterium extorquens strains CM4 and DM4 were grown with CM and with DCM, respectively. Longer generation times were observed for growth with CM and DCM compared to with methanol, as observed previously (Muller et al., 2011b; Roselli et al., 2013). Global profiles of gene expression were assessed by RNA-Seq with sequencing coverage exceeding 600X (**Table 1**). Only 29 CDS of strain CM4 and 36 CDS of strain DM4 showed no reads. Obtained data were normalized (Table S1; Figure S1) and verified by qPCR for a set of 16 genes spanning 2 orders of magnitude of gene expression (Table S2; Figure S2). Only genes with satisfactory false discovery rates (p-value < 0.1) were further

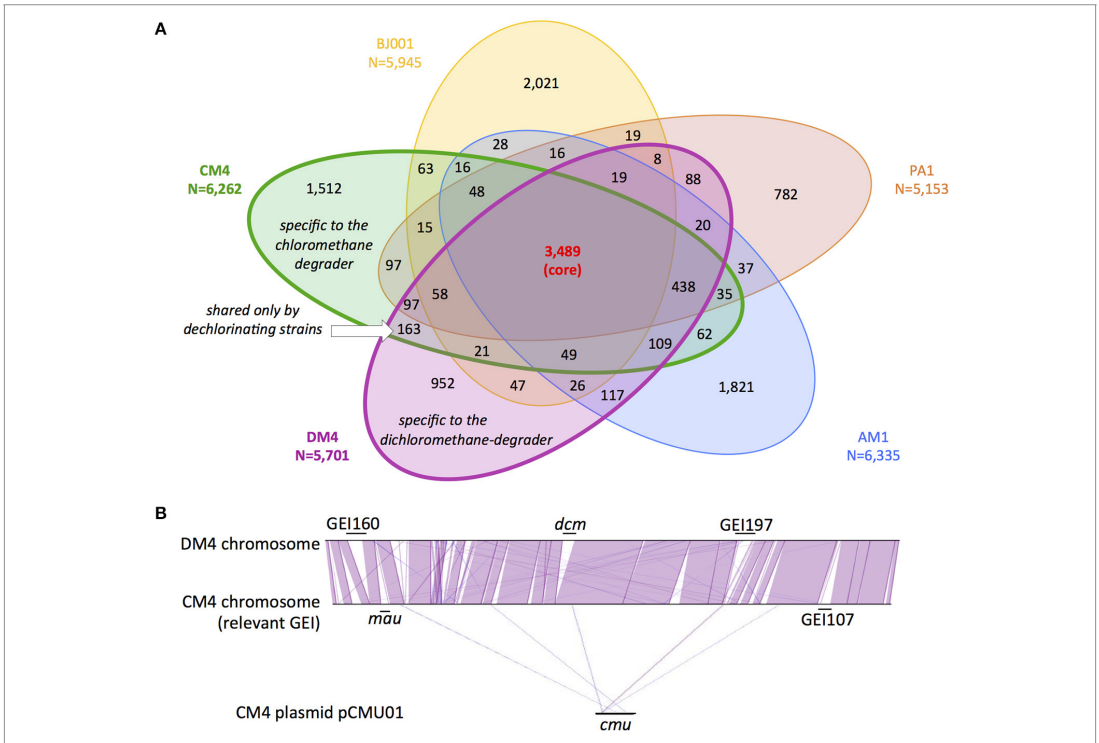


FIGURE 1 | Gene content and synteny in genomes of *M. extorquens* CM4 and DM4. Comparative genome analyses were performed on the sequenced and assembled genomes of *M. extorquens* dehalogenating strains CM4 and DM4, as well as strains AM1, PA1, and BJ001 (Vuilleumier et al., 2009; Marx et al., 2012), using MaGe (Vallenet et al., 2017) on the MicroScope online platform (see Materials and Methods). **(A)** Common homologous genes were defined and assigned to either core (found in all 5 strains), variable (found in 2–4 strains), shared (found only in CM4 and DM4) or specific (found in one strain only) genomes. **(B)** Synteny of CM4 and DM4 genomes indicating genomic islands (GEI) relevant to C_1 metabolism, i.e., *dcm* for dichloromethane utilization and *mau* for methylamine utilization (Vuilleumier et al., 2009), and *cmu* for CM utilization (380 kb plasmid pCMU01, Marx et al., 2012) (see **Table 3** for GEI details). Synteny involved groups of at least 6 homologous genes (purple lines, strand conservation; blue lines, strand inversion).

TABLE 1 | Overview of genomic properties of *M. extorquens* CM4 and DM4 and obtained RNA-Seq data.

Strain	Genome[a]	Growth condition generation time (h)[b]	Total reads[c]	Mapped reads (%)	rRNA (%)[d]
CM4	chromosome 5.8 Mb GC% = 68.2	Methanol	30,337,270	98.6	18.3
	pCMU01 380.2 kb GC% = 66.3	(3.0 ± 0.2)	37,414,003	98.2	14.3
	p2MCHL 22.6 kb GC% = 63.9	Chloromethane	53,790,411	96.3	18.6
		(5.4 ± 0.4)	36,665,352	97.5	25.4
DM4	chromosome 5.9 Mb GC% = 68.1	Methanol	48,154,448	99.1	12.3
	p1METDI 141.5 kb GC% = 65.3	(3.4 ± 0.4)	38,757,418	97.9	23.0
	p2METDI 38.6 kb GC% = 63.7	Dichloromethane	43,101,981	95.0	34.7
		(9.0 ± 0.7)	32,066,920	95.0	20.0

[a] *In strain CM4, the CM utilization pathway is encoded by cmu genes located on plasmid pCMU01 (Roselli et al., 2013). In strain DM4, dcm genes involved in DCM utilization are located on 5.5 kb dcm islet (Muller et al., 2011b) within a 126 kb genomic island on the chromosome (Vuilleumier et al., 2009).*
[b] *Aerobic growth in M3 medium with 10 mM one-carbon substrate provided as sole source of carbon and energy.*
[c] *Illumina HiSeq2000.*
[d] *Percentage of total reads.*

analyzed and included 1,245 and 301 genes for the genomes of strain CM4 and strain DM4, respectively.

Transcript abundance of strains grown with chlorinated methanes CM or DCM and with methanol, the reference methylotrophic growth substrate for *M. extorquens*, were then compared. A total number of 150 genes for CM4 and 190 genes for DM4 were detected as differentially abundant genes (**Table 2**). Overall gene expression of common genes during growth with methanol was similar in strains CM4 and DM4, as expected (Figure S1). More pronounced differences were observed between patterns of gene expression for common genes of *M. extorquens* strains CM4 and DM4 grown with CM or with DCM (Table S4), with only few genes showing the same trend with CM and DCM relative to methanol (**Figure 2**). Strikingly, essentially all 163 CDS only shared by the 2 dehalogenating strains and not found in any of the 3 other non-dehalogenating *M. extorquens* strains (**Figure 1A**) lacked differential expression between chlorinated methanes and methanol (with the exception of METDI4814, less abundant with DCM) (Table S3; Figure S3B). Among those, 42 genes were plasmid-borne and co-localized on the largest plasmids in both strains (Table S3), with some displaying high transcript abundance (Figure S3).

Of the 12 common genes encoded by the core genome more highly expressed during growth with either CM and DCM than with methanol (Table S4; **Figure 2**), only 2, *degP* and *hppA*, have predicted functions. Gene *degP* encodes a putative periplasmic serine protease whose *Escherichia coli* homolog HtrA (39% amino acid identity) is a central housekeeping molecular chaperone protein controlling the production of functional, properly folded outer-membrane proteins (Ge et al., 2014). In *Carboxydothermus hydrogenoformans*, the membrane-bound proton translocating pyrophosphatase HppA (48% amino acid identity to *M. extorquens* homologs) uses energy from pyrophosphate hydrolysis to build up a proton motive force by proton extrusion (Belogurov and Lahti, 2002). In *Rhodospirullum rubrum*, the closely homologous HppA (72% amino acid identity) is involved in stress bioenergetics and in particular salt stress (Lopez-Marques et al., 2004).

TABLE 2 | Differential expression of core, variable, shared and strain-specific CDS during growth with chlorinated methanes.

Strain	Genome[a]	CDS number	Differential transcript abundance with chlorinated methanes		
			Ratio (%)	Higher[b]	Lower[b]
CM-degrading *M. extorquens* CM4	Core	3,489	1.6	45	11
	Variable	1,098	1.8	19	1
	Shared only with DM4	163	0	0	0
	Specific to CM4	1,512	4.9	73	1
DCM-degrading *M. extorquens* DM4	Core	3,489	1.8	31	32
	Variable	1,097	5.9	28	46
	Shared only with CM4	163	0	0	0
	Specific to DM4	952	5.6	10	43

[a] *Common, variable, shared only by dehalogenating strains CM4 and DM4, and strain-specific genomes, as defined in Material and Methods.*
[b] *Number of CDS with higher ($log_2fc > 2$) or lower ($log_2fc < -2$) transcript abundance in cultures grown with chlorinated methanes compared to with methanol.*

The number of genes of the core genome with significantly lower expression on chlorinated methanes is also low (9 genes, **Figure 2**). These may rather be methanol-induced rather than genes repressed by chlorinated methanes, since 7 of these genes belong to predicted operons associated with the alternative alcohol dehydrogenase ExaF to the paradigmatic methanol dehydrogenase encoded by *mxa* genes (Good et al., 2016; Table S4).

The 10 most highly transcribed genes during growth with chlorinated methanes in strains CM4 and DM4 (**Figure 3**) included 3 common to both strains, albeit with values of $log_2fc < 2$ compared to growth with methanol. Of the other 7 genes highly transcribed on chlorinated methanes, only strain-specific genes directly associated with dehalogenation (5 for strain CM4, 2 for strain DM4), and not common genes, showed log_2fc values

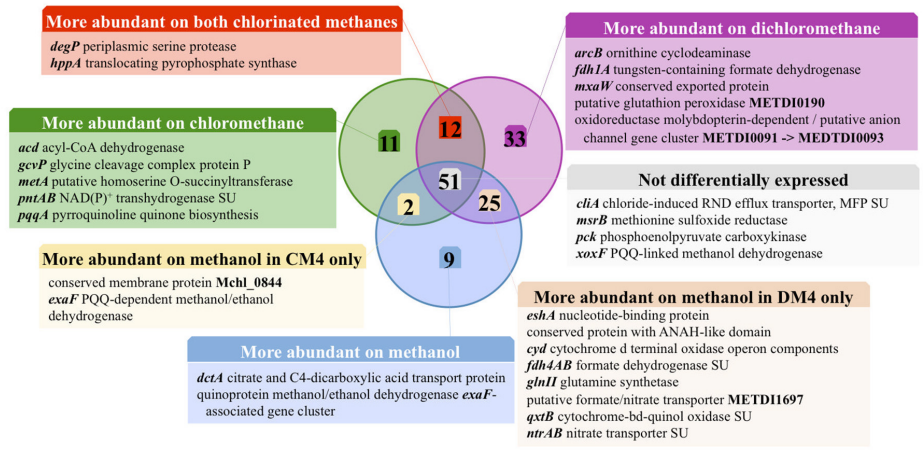


FIGURE 2 | Carbon source-dependent transcript abundance of common core genes in *M. extorquens* CM4 and DM4. Different categories were defined on the basis of the $\log_2$ fold-change ($\log_2$fc) values of RNA-Seq reads for *M. extorquens* CM4 grown with CM (green circle) and for *M. extorquens* DM4 grown with DCM (purple circle) versus methanol (blue cercle). Complete gene names, $\log_2$fc values for each category are detailed in Table S4.

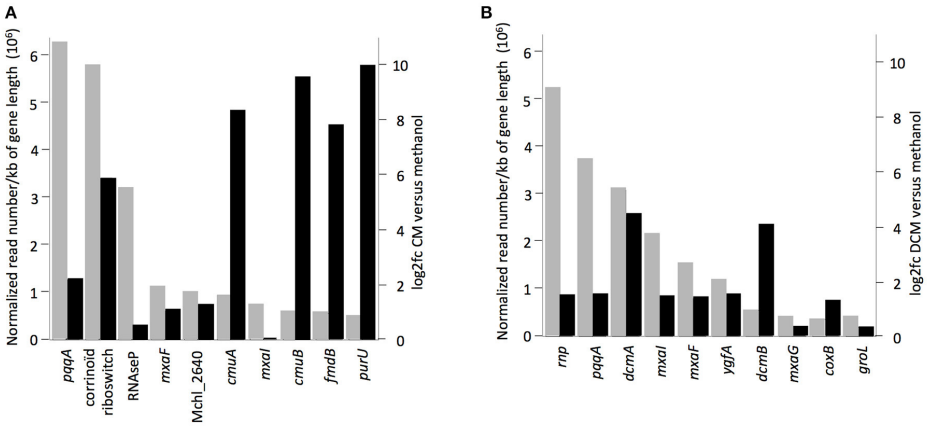


FIGURE 3 | Genes with highest transcript abundance in cultures grown with chlorinated methanes compared to with methanol. Black rectangles indicate values of $\log_2$fc (values higher than 2.0 or lower than −2.0 mean normalized reads more abundant in cultures grown with chlorinated methanes or with methanol, respectively). Gray rectangles represent normalized read numbers divided by gene length in kb. RNA-Seq data from cultures of **(A)** *M. extorquens* CM4, **(B)** *M. extorquens* DM4.

>2. Transcription responses for growth with CM and DCM were analyzed in more detail.

The Chloromethane Transcriptome of Strain CM4

Under the conditions tested, a total of 137 genes (2% of the CM4 transcriptome) had higher transcript abundance with CM than with methanol. Only 43 of these belong to the core genome defined here for *M. extorquens* ($\log_2$fc $\geq$ 2; Table S5; **Figure 4A**). Among these, 11 core genes only differentially expressed in strain CM4 had not previously been associated with *M. extorquens* growth with CM. They include the *pnt* gene cluster encoding a NADH/NADPH transhydrogenase with cross-membrane proton translocation activity (Chou et al., 2015); two consecutive paralogs of *pqqA*, a precursor of the redox active dehydrogenase cofactor PQQ (Ochsner et al., 2015); *ykuD*, encoding a transpeptidase of a large enzyme family associated with cell wall biology (Bielnicki et al., 2005); *ada* encoding

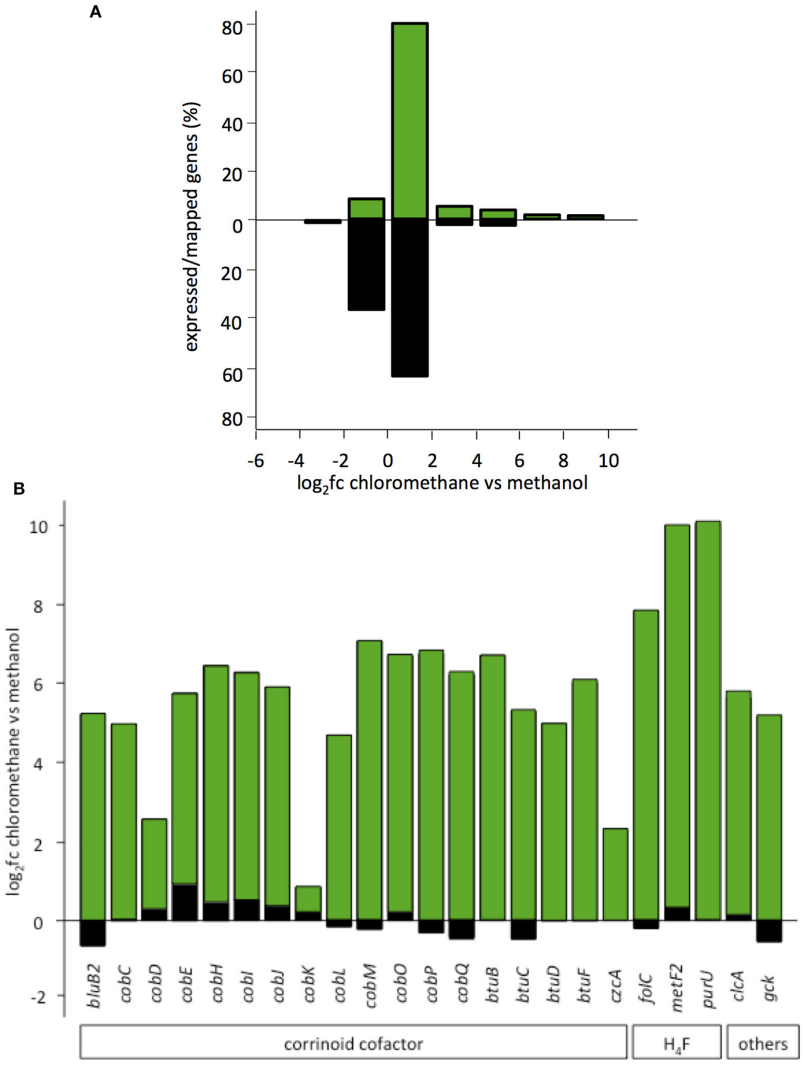


FIGURE 4 | Involvement of plasmid pCMU01 in the chloromethane transcriptome. **(A)** Expressed percentage of total genes per log_2fc range. Genes encoded by plasmid pCMU01 (in green) and by the *M. extorquens* CM4 chromosome (in black) were grouped according to their log_2fc values. Chromosome and plasmid pCMU01 encode 6,262 and 361 genes, and the total percentage of expressed genes is 1.4 and 17.5%, respectively. **(B)** Differential expression of pCMU01- and chromosome-encoded paralogs.

a bifunctional transcriptional activator which acts in response to alkylation damage (51% amino acid identity with the well-characterized *E. coli* homolog) (McCarthy and Lindahl, 1985); *ibpA* encoding a small heat shock protein (sHSP) with 57% amino acid identity to *E. coli* IbpA, which protects enzymes against oxidative stress (Kitagawa et al., 2002).

The major contribution in the CM transcriptome involves strain-specific genes either directly or indirectly associated with dehalogenation (**Figure 5**). First, 40 genes known from previous work to be associated with *cmu* gene clusters in CM-degrading strains, and found on the pCMU01 plasmid in strain CM4, showed highest relative expression (log_2fc values 7.7–10) during

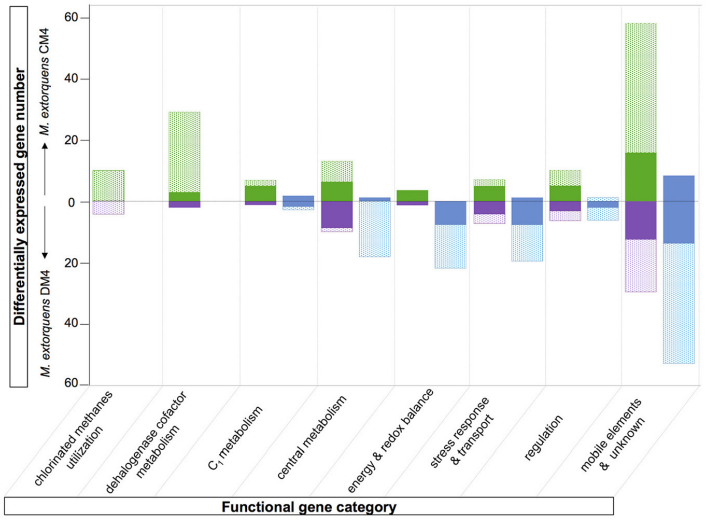


FIGURE 5 | Overview comparison of the chloromethane and dichloromethane transcriptomes. Differentially abundant genes with predicted function are clustered in different functional categories (see Tables S6, S7). The fractions of genes of the core and variable genomes are indicated in plain and dotted rectangles, respectively. Transcript with higher abundance in chloromethane, dichloromethane or methanol are in green, purple or blue, respectively.

growth with CM (Table S5; also see **Figure 4A**). Most of these genes are involved in metabolism and transport of the corrinoid and H_4F cofactors essential for CM dehalogenation by the *cmu* pathway (Studer et al., 2001) (Table S5; **Figure 4B**). A gene cluster specifically shared between sequenced genomes of CM degrading isolates [acxABC, Roselli et al., 2013 >81% amino acid identity with the characterized aerobic bacterium *Xanthobacter autotrophicus* acetone carboxylase components (Sluis et al., 2002)] was also more expressed in cells grown with CM (see Figure S3A for log_2fc values and expression level) although its functional connection with the *cmu* pathway remains to be characterized. Second, half of the plasmid pCMU01-borne genes more highly expressed on CM had homologous copies located on the chromosome (31 genes out of 62 plasmid-borne genes with higher transcript abundance; Table S5; **Figure 3B**). All the corresponding chromosomal paralogs were, in contrast, not differentially expressed. Again, a majority of plasmid-borne homologs were associated with the essential corrinoid and tetrahydrofolate cofactors of CM dehalogenation by the *cmu* pathway (**Figure 3B**). Similarly, the pCMU01 plasmid homolog of the *clc* H^+/Cl^- antiporter gene was more highly expressed on CM (log_2fc value of 5.5), unlike its chromosomal homolog (68% identity at the protein level; Figure S3A). Taken together, this argues strongly for a key role of plasmid pCMU01-borne genes in adaptation to growth with CM in strain CM4.

The Dichloromethane Transcriptome of Strain DM4

Under the conditions tested, only 3% of the DM4 transcriptome was altered in growth with DCM relative to methanol. Among these, 69 genes (1% of the transcriptome) showed higher transcript abundance in cultures grown with DCM (**Table 2**). These included the 4 genes of the *dcm* islet, i.e., *dcmA* coding for the DCM dehalogenase essential for growth with DCM; *dcmR* involved in its own transcription as well as that of *dcmA* (Leisinger et al., 1994; Kayser et al., 2002); and *dcmB* and *dcmC* genes of unknown function (Table S6). This is the first reported evidence for transcription of *dcmC* and its modulation by DCM, and confirms previous results for *dcmB* (Muller et al., 2011b). Transcripts of *dcmA* and *dcmB* were among the most abundant with DCM (**Figure 3B**).

The majority of genes showing differential abundance with DCM only (33 genes) belong to the core genome of *M. extorquens*, and significantly exceeds the number of *M. extorquens* genes of the core genome also more highly abundant during growth with CM (12 genes), or those only more highly expressed during growth with CM but not DCM (11 genes) (**Figure 2**). The specific functions of most of these genes remain elusive, although an association with redox status is suggested, with a putative glutathione peroxidase (METDI0190) and 2 sets of genes associated with uncharacterized molybdenum-dependent enzymes (METDI2693, METDI0091- METDI0093) showing increased transcript abundance (Table S6). Of those, the gene encoding the glutathione peroxidase was highly expressed (Figure S3B).

The contribution of the 2 plasmids of DCM-degrading strain DM4 (**Table 1**) in the DCM transcriptome is limited (2 uncharacterized genes among 186 CDS were differentially expressed). This suggests that unlike for the CM-degrading

strain, DM4 plasmids do not play a role in adaptation to DCM in strain DM4.

The set of genes from strain DM4 with higher expression on methanol is also larger than that of strain CM4 (**Figure 5**). The majority of these genes is found in gene clusters with redundant gene content (Figure S4), and often associated with genomic islands (GEIs) (**Table 3**, and see next subsection).

Expression of Genomic Island-Borne Genes in Dehalogenating *M. extorquens* Strains

As plasmids, GEIs are major agents of rapid genome evolution and adaptation of catabolic pathways in bacteria, and are often found integrated in the chromosome (van der Meer and Sentchilo, 2003). The potential role of GEIs in adaptation of *M. extorquens* to growth with chlorinated methanes was investigated. Out of the 14 and 11 specific GEIs detected in strains CM4 and DM4, respectively, only 3 GEIs contained genes with differential expression (**Table 3**). Such genes only represented a minor part of GEIs, and their roles beyond that of genes associated with dehalogenation remain unknown. It is striking that no other genes of the 126 kb *dcm* GEI of strain DM4 were differentially transcribed beyond those of the 5.5 kb *dcm* islet, unlike many genes associated with *cmu* gene clusters on plasmid pCMU01 (**Table 3**).

DISCUSSION

Dehalogenation-Specific Gene Complement

The number of strain-specific genes, as such potentially associated with dehalogenative metabolism, was relatively small (**Figure 1A**) and genes with an annotation suggestive of a role in this context were not detected (Table S3). Moreover, apart from dehalogenase genes, such genes were not especially prone to differential expression with chlorinated methanes (Tables S5, S6; **Figures 3–5**). Further, none of the few genes shared only by CM- and DCM-degrading strains among a small group of 5 *M. extorquens* strains were differentially expressed during growth with chlorinated methanes. Nevertheless looking for upregulated genes in CM/DCM-grown cultures vs. methanol-grown cultures will not detect constitutively expressed genes that support dehalogenation growth, and other complementary approaches need to be assessed (Ochsner et al., 2017). Among highly transcribed genes, a few had predicted functions in plasmid transfer and replication (genes *icmL*, *repA*, *repB*, *traD*; Figure S3). Thus, the part of the transcriptional response common to both chlorinated methanes was limited (Table S3). Adaptation to growth with chlorinated methanes seems to be largely a function of fine-tuning gene content/expression rather than large-scale changes as discussed in the following.

TABLE 3 | GEIs containing genes with differential transcript abundance.

GEI[a]	Length (kb)	CDS (start-end)	Structure[a]	Number		% CDS of GEI	Genes and potential relevant functions
				CM or DCM	Methanol		
STRAIN DM4							
dcm	126	METDI2550–METDI2682	(none) tRNA-int-mob-misc_RNA-SIGI-AH (none)	4	0	3	DCM utilization (*dcmRABC*)
GEI160	160	METDI0225–METDI0426	(tRNA)-int-SIGI-AH (int)	1	35	18	Carbon metabolism (*ackA*-like, *adh*-like, *xfp, fabI*, *phbC*); energy (ATPase, cyt. c, cyt. c oxidase); stress (*clpB*, *DnaJ*)
GEI197	197	METDI4320–METDI4570	(none) tRNA-int-mob-SIGI-AH (IS)	1	34	14	Carbon metabolism (*ackA*-like, *adh*-like, *xfp*); energy (ATPase, cyt. c, cyt. c oxidase); members of ANAH-like superfamily
STRAIN CM4							
GEI107	107	Mchl4758–Mchl4844	tRNA-int-SIGI-IVOM-Specific_Region	4	0	5	Putative porin; others without predicted function
pCMU01[b]	194	Mchl5386–Mchl5736	Unknown	63	0	18	CM utilization (*cmu*); corrinoid cofactor biosynthesis (*cob*, *czc*, *bluB2*); H_4F cofactor metabolism (*bluB2*, *folC2*); carbon assimilation (*acx*)

[a]Identified with "Regions of Genomic Plasticity" tool (MaGe plateform, Vallenet et al., 2017 using M. extorquens CM4, DM4, AM1, PA1, and BJ001 genomes). Int, integrase; mob, mobility determinant; AH, region rich in genes detected by Allen Hunter based on variable-length k-mers bias (Vernikos and Parkhill, 2006); SIGI, region with biased codon usage (Waack et al., 2006).

[b]Plasmid pCMU01 (380 kb, 350 predicted CDS) was only found so far in M. extorquens CM4 (Roselli et al., 2013), precluding analysis with the RGP tool except to remove extensive regions of synteny with DM4 plasmid p1METDI (Mchl5546-Mchl5579, Mchl5591-Mchl5621 in plasmid pCMU01). Hence, the 239 CDS without hits to the other 4 M. extorquens genomes were conservatively defined maximal GEI.

Adaptive Stress Response

Indications for a transcriptional adaptive stress response to the utilization of chlorinated methanes in *M. extorquens* obtained here confirm previous suggestions from mutant growth phenotypes in the case of DCM (Muller et al., 2011b) and proteomic data in the case of CM (Roselli et al., 2013). Whereas expression of several genes associated with the general stress response was modulated (Tables S3, S5, S6; **Figure 5**), genes potentially associated with dehalogenation-specific stress were of particular interest. Since bacteria that grow with chlorinated methanes have to cope with production of intracellular hydrochloric acid production, one key question to address was which of intracellular proton or chloride buildup represents a larger stress for *M. extorquens*.

Intracellular generation of protons during growth with chlorinated methanes is expected to be detrimental for growth through intracellular acidification and also through dissipation of the proton-motive force and associated ATP production. It thus seems significant that the membrane-bound proton translocating pyrophosphatase *hppA* was one of only 12 genes of the core genome with higher expression on both chlorinated methanes. HppA-driven proton extrusion and restoration of the proton-motive force involves an additional energy expense through pyrophosphatase hydrolysis (Belogurov and Lahti, 2002).

Similarly, the H^+/Cl^- antiporter ClcA involved in adaptation to chloride stress affords chloride efflux at a cost for the proton-motive force. The corresponding gene was expressed at high constitutive levels in strain DM4 (Figure S3B), as shown recently for DCM-degrading isolates including strain DM4 (Michener et al., 2014b). In case of strain CM4, *clcA* displayed moderate constitutive expression, but a pCMU01 plasmid-borne paralog *clcA2* (69% amino acid identity) was significantly more highly expressed in cultures on CM (Table S5; Figure S3A). This suggests that *clcA*-driven chloride extrusion is crucial for growth with both chlorinated methanes, and may involve different paralogs of this large gene family. Recent work showed that *clcA*, when transcribed from its native promoter cloned from strain DM4, confers higher fitness for growth with DCM but not with CM in *Methylobacterium* strains not previously exposed to chlorinated methanes (Michener et al., 2014a, 2016). It is noteworthy that the uncharacterized RND efflux transporter CliABC previously identified as chloride-induced in strain DM4 (Muller et al., 2011b) was found here to be even more expressed than gene *clcA* (Figure S3B). Taken together, the obtained data suggest that an increase in intracellular chloride levels represents more of the problem for growing strains of *M. extorquens* than intracellular proton production.

Regulation under Dehalogenative Methylotrophic Conditions

Strain-specific genes were among the most transcribed genes, especially those within gene clusters involved in chlorinated methane dehalogenation, i.e., *dcmA* and *cmuAB* genes (**Figure 3**). Here, CM-dependent transcription involved much larger changes in expression compared to DCM-dependent transcription (log_2fc of 10 vs. 4, respectively). This confirms previous studies with RT-qPCR and transcriptional fusions of the *cmuA* promoter (Farhan Ul Haque et al., 2013). Among genes with increased expression levels with chlorinated methanes identified in this work, a significant number are annotated as putative regulatory genes (9 for CM and 6 for DCM respectively; Tables S5, S6). The transcriptional regulator of *dcmA* has been preliminarily characterized (La Roche and Leisinger, 1991; Muller et al., 2011b). In contrast, the genetic determinants involved in regulating the expression of the *cmu* pathway remain to be identified (Roselli et al., 2013).

The contribution of genes less expressed during growth with CM than with methanol was limited in the case of the core genome (11 genes) and even more so for the CM4 strain-specific genes (2 genes; **Figure 5**). This suggests that gene downregulation mediated by CM in strain CM4 is limited. On the other hand, a striking finding was that many plasmid pCMU01-borne paralogs but not their chromosomal homologs had higher transcript abundance in CM cultures (**Figure 4B**). This suggests the existence of as yet uncharacterized regulation mechanisms favoring expression of plasmid-encoded over chromosomally encoded gene paralogs in response to growth with CM.

Effects of Dehalogenation Metabolism on Transcription of Methylotrophy Genes

High levels of transcripts of genes *mxaFI* for subunits of the canonical methanol dehydrogenase (MDH) of *M. extorquens* (Amaratunga et al., 1997) were observed in methanol cultures, as expected, but also in cultures grown with chlorinated methanes (**Figure 3**). Since dehalogenation of DCM leads to formaldehyde, one explanation might be that transcription of methanol dehydrogenase MxaFI, which also converts formaldehyde (Nunn and Lidstrom, 1986), is also induced by formaldehyde. However, this does not strictly apply to growth with CM, as formaldehyde is not a direct product of CM dehalogenation by the *cmu* pathway (Vannelli et al., 1999; **Figure 6**) and chemical equilibrium between free formaldehyde and methylene-tetrahydrofolate ($CH_2{=}H_4F$) does not favor free formaldehyde (Kallen and Jencks, 1966). This suggests that methanol dehydrogenase expression may be induced by other downstream metabolites common to CM and methanol catabolism, such as other folate-bound C_1 compounds or formate, or that *mxaFI* expression is not switched off in strain CM4. On the other hand, other recently identified dehydrogenases active with methanol did not show the same transcriptional profile. In particular, the expression of genes for the XoxF-type enzymes was not modulated by chlorinated methanes. As for the PQQ-dependent ExaF-type dehydrogenase active with ethanol, methanol and formaldehyde (Good et al., 2016), seven genes directly upstream of gene *exaF* showed higher transcript levels in methanol-grown cultures in both CM4 and DM4 strains, including for gene *exaF* in strain CM4 (Table S5). Our results suggest complex modes of cross-regulation of key enzymes of methylotrophic metabolism in response to different C_1 substrates that mostly involve genes that belong to the core genome (**Figure 5**).

Metabolic Fine-Tuning for Dehalogenative Methylotrophic Growth

Several genes of *M. extorquens* strains CM4 and DM4 that show significant modulation of expression with chlorinated

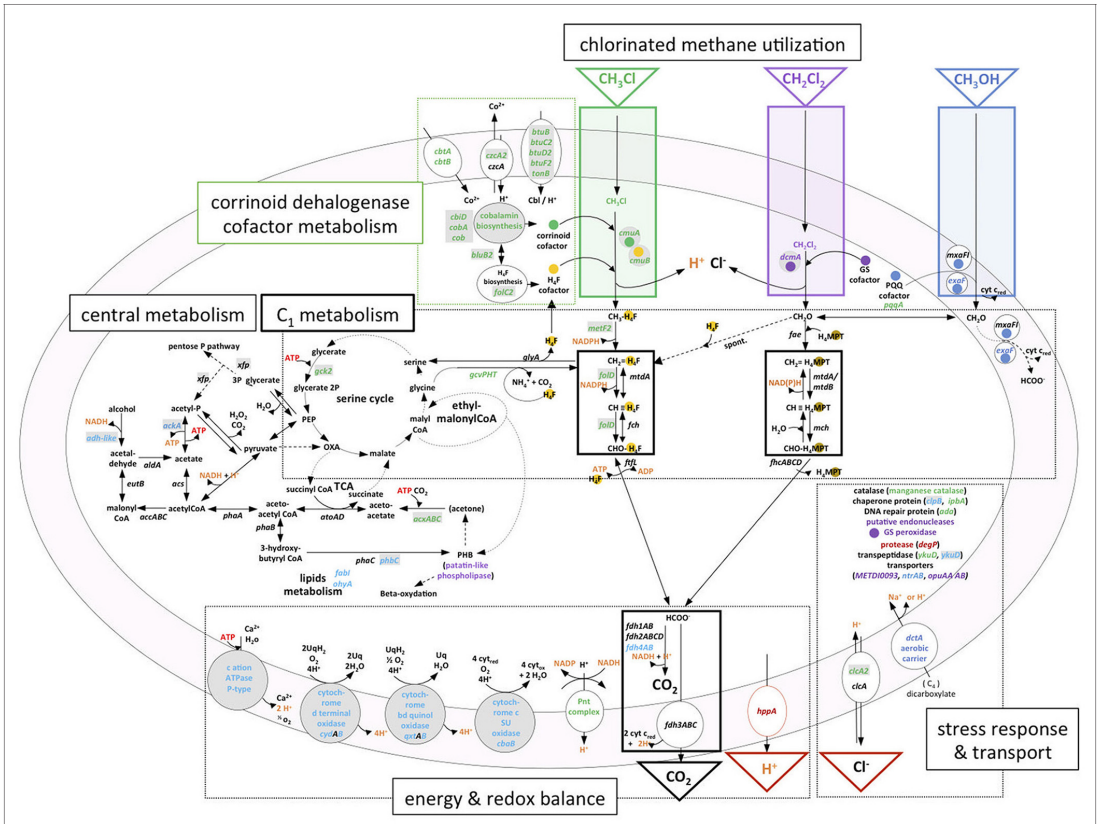


FIGURE 6 | Transcriptional response and metabolism of chlorinated C_1 compounds chloromethane and dichloromethane in *M. extorquens*. Chloromethane dehalogenation and methyl transfer to tetrahydrofolate (H_4F) reactions are boxed in green; dichloromethane dehalogenation to formaldehyde in purple; methanol oxidation to formaldehyde in blue. Within C_1 metabolism, black boxes delimit common methylotrophic modules (Chistoserdova et al., 2009): H_4F and tetrahydromethanopterin (H_4MPT)-dependent C_1 transfer reactions, and formate oxidation for C_1 dissimilation with $C0_2$ production. Serine and ethylmalonyl-CoA cycles for C_1 assimilation (Ochsner et al., 2015) are schematized with two connected circles. Reactions related to essential co-factors for chlorinated methane utilization are highlighted with colored spots; H_4F (yellow), corrinoid cofactor (green), glutathione (purple), and pyrroquinoline quinone (blue). GEI-located genes in *M. extorquens* DM4 or in *M. extorquens* CM4 (plasmid pCMU01) are shown in gray-shaded boxes. RNA-Seq data are summarized using gene color-coding for more abundant transcripts on CM (green), DCM (purple), both CM and DCM (dark red), methanol [both *M. extorquens* CM4 and DM4 (dark blue), DM4 only (light blue)]. Compounds associated with energy-dependent transformations are highlighted in orange. Cbi, cobinamide; Cbl, cobalamin; CoA, coenzyme A; PEP, phosphoenolpyruvate; OXA, oxaloacetate; TCA, tricarboxylic acid cycle.

methanes have predicted functions in energy homeostasis and redox processes. However, both the types of genes involved and the differential transcription patterns often differed between CM- and DCM-degrading strains. In the case of the DCM-degrading strain DM4, several gene clusters containing cytochrome oxidase genes displayed lower transcript counts in DCM cultures (Table S6). In particular, a substantial set of GEI-associated genes (**Table 3**), belonging to the variable genome of *M. extorquens* and potentially linked to cytochrome electron transport and thereby the transmembrane H^+ gradient, had enhanced transcript abundance in cultures grown with methanol (Figure S4). Other potentially energy-associated adjustments of *M. extorquens* DM4 include very high transcript levels for various genes involved in metabolism of fatty acids and polyhydroxyalkanoates (Table S6), which are a storage form of carbon and energy (Escapa et al., 2012), suggesting that growth on DCM vs. methanol may involved different carbon and energy spillage control processes.

For the CM-degrading strain CM4, in contrast, obtained data suggest that energy balance adjustments as a function of growth with CM or methanol involve NAD(P)-linked metabolism. Whereas only one enzymatic step from methanol to formate involves one molecule of $NAD(P)^+$ (MtdA/MtdB), for each molecule of CM transformed to formate, two molecules

of NAD(P)H are potentially generated from $NAD(P)^+$ in two successive reactions specifically catalyzed by MetF2 and FolD (**Figure 6**), These enzymes can use both NAD^+ and $NADP^+$ as cofactors, although *M. extorquens* CM4 FolD activity had only been assessed so far with $NADP^+$ (Marx and Lidstrom, 2004). The corresponding genes *metF2* and *folD* are two of the most differentially expressed genes in the CM transcriptome (Table S5). Importantly, the *cmu* pathway specifically features FolD rather than MtdA and Fch for oxidation of C_1 carbon to formate (Studer et al., 2002) unlike growth on methanol (Marx and Lidstrom, 2004). MetF2 and FolD may thus have different preferences than MtdA/MtdB for using NAD^+ or $NADP^+$ as cofactors. Should this be the case, significant altering of the intracellular pools of NAD^+ and $NADP^+$ during growth with CM compared to methanol could become growth limiting if levels of oxidized cofactors are limiting. A candidate for metabolic fine-tuning in this context is the membrane-bound transhydrogenase encoded by the chromosomal *pnt* gene cluster. All three *pnt* genes were significantly more expressed on CM at both transcription (Table S5) and protein (Roselli et al., 2013) levels. Transhydrogenase catalyzes the reversible reaction $NADPH + NAD^+ + H^+_{in} <=> NADP^+ + NADH + H^+_{out}$ (Carroll et al., 2015). In addition to $NADP^+$ regeneration via the membrane-bound transhydrogenase, concomitant proton efflux could thus potentially also help maintain internal cellular pH during CM dehalogenation.

During growth with CM, the high demand for oxidized $NAD^+/NADP^+$ cofactors may alter $CH_2{=}H_4F$ flux toward the serine cycle by limiting the carbon flux toward formate production. A compensatory metabolic rerouting of $CH_2{=}H_4F$ toward formate formation could be needed and may be achieved by the components of the so-called glycine cleavage complex (**Figure 6**), which were more highly abundant at both the transcript (GcvPHT, Table S5) and protein levels (GcvT; Roselli et al., 2013).

In conclusion, utilization of horizontally transferred genes for growth of *M. extorquens* with CM and DCM presumably involved required several post-transfer adjustments, as shown in recent experimental evolution experiments (Michener et al., 2014a, 2016). The new data obtained in the present study highlight potential global adjustments at the transcriptional level and more generally, the importance of substrate- and pathway-dependent genome adaptation following acquisition of new growth-supporting abilities such as degradation of toxic halogenated compounds. Clearly, *M. extorquens* represents a model of choice to address these issues in the future.

AUTHOR CONTRIBUTIONS

PC, BM, and FB prepared RNA. PC and AA constructed cDNA libraries. PC and FB performed RNA-Seq analysis. BM carried out RT-qPCR assays and data processing. FB designed the study. PC, MW, SV, FB, and SC participated in bioinformatic analysis. PC, BM, SK, SV, SC, and FB were involved in data analysis and interpretation. PC, SV, and FB wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was partially funded by a Ph.D grant to PC from Région Alsace (France) and DFG grant Ko2912/5-1 (Germany), the German Academic Exchange Service (DAAD), the University of Bayreuth, and French-German mobility funds PROCOPE (Deutscher Akademischer Austausch Dienst), the French Agence Nationale de la Recherche (ANR grant 14CE35-005-01), as well as support for RNA-Seq analysis by a GIS IBiSA grant-in-kind to FB. The funding bodies had no impact on the design of the study, data collection, analysis, interpretation of data, nor in writing the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge Beatrice Segurens (CEA CNRS-UMR 8030, France) for RNA extraction protocol optimization, David A. C. Beck (University of Washington, Seattle, USA) and Anne Francez-Charlot (ETH Zurich, Switzerland) for checking proposed reference genes in their unpublished *M. extorquens* RNA-Seq libraries.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01600/full#supplementary-material

REFERENCES

Amaratunga, K., Goodwin, P. M., O'Connor, D., and Anthony, C. (1997). The methanol oxidation genes *mxaFJGIR(S)ACKLD* in *Methylobacterium extorquens*. *J. Bacteriol.* 146, 31–38. doi: 10.1016/S0378-1097(96)00399-0

Amato, P., Parazols, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.-M. (2007). Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: major groups and growth abilities at low temperatures: microorganisms from the water phase of tropospheric clouds. *FEMS Microbiol. Ecol.* 59, 242–254. doi: 10.1111/j.1574-6941.2006.00199.x

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., et al. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols* 8, 1765–1786. doi: 10.1038/nprot.2013.099

Belogurov, G. A., and Lahti, R. (2002). A lysine substitute for K^+. A460K mutation eliminates K^+ dependence in H^+-pyrophosphatase of *Carboxydothermus hydrogenoformans*. *J. Biol. Chem.* 277, 49651–49654. doi: 10.1074/jbc.M210341200

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.* 57, 289–300.

Bielnicki, J., Devedjiev, Y., Derewenda, U., Dauter, Z., Joachimiak, A., and Derewenda, Z. S. (2005). *Bacillus subtilis* YkuD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. *Proteins* 62, 144–151. doi: 10.1002/prot.20702

Bringel, F., and Couée, I. (2015). Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. *Front. Microbiol.* 6:486. doi: 10.3389/fmicb.2015.00486

Carpenter, L. J., Reimann, S., Burkholder, J. B., Clerbaux, C., Hall, B. D., Hossaini, R. et al. (2014). "Ozone-depleting substances (ODSs) and other gases of interest to the Montreal protocol," in *Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project*, eds A. Engel and S. A. Montzka (Geneva: World Meteorological Organization), 1.1–1.101.

Carroll, S., Chubiz, L., Agashe, D., and Marx, C. (2015). Parallel and divergent evolutionary solutions for the optimization of an engineered central metabolism in *Methylobacterium extorquens* AM1. *Microorganisms* 3, 152–174. doi: 10.3390/microorganisms3020152

Chen, F., Freedman, D. L., Falta, R. W., and Murdoch, L. C. (2012). Henry's law constants of chlorinated solvents at elevated temperatures. *Chemosphere* 86, 156–165. doi: 10.1016/j.chemosphere.2011.10.004

Chistoserdova, L., Kalyuzhnaya, M. G., and Lidstrom, M. E. (2009). The expanding world of methylotrophic metabolism. *Annu. Rev. Microbiol.* 63, 477–499. doi: 10.1146/annurev.micro.091208.073600

Chou, H. H., Marx, C. J., and Sauer, U. (2015). Transhydrogenase promotes the robustness and evolvability of *Escherichia coli* deficient in NADPH production. *PLoS Genet.* 11:e1005007. doi: 10.1371/journal.pgen.1005007

Derendorp, L., Wishkerman, A., Keppler, F., McRoberts, C., Holzinger, R., and Rockmann, T. (2012). Methyl chloride emissions from halophyte leaf litter: dependence on temperature and chloride content. *Chemosphere* 87, 483–489. doi: 10.1016/j.chemosphere.2011.12.035

Escapa, I. F., García, J. L., Bühler, B., Blank, L. M., and Prieto, M. A. (2012). The polyhydroxyalkanoate metabolism controls carbon and energy spillage in *Pseudomonas putida*: PHA metabolism in *Pseudomonas putida*. *Environ. Microbiol.* 14, 1049–1063. doi: 10.1111/j.1462-2920.2011.02684.x

Farhan Ul Haque, M., Nadalig, T., Bringel, F., Schaller, H., and Vuilleumier, S. (2013). Fluorescence-based bacterial bioreporter for specific detection of methyl halide emissions in the environment. *Appl. Environ. Microbiol.* 79, 6561–6567. doi: 10.1128/aem.01738-13

Ge, X., Wang, R., Ma, J., Liu, Y., Ezemaduka, A. N., Chen, P. R., et al. (2014). DegP primarily functions as a protease for the biogenesis of β-barrel outer membrane proteins in the Gram-negative bacterium *Escherichia coli*. *FEBS J.* 281, 1226–1240. doi: 10.1111/febs.12701

Good, N. M., Vu, H. N., Suriano, C. J., Subuyuj, G. A., Skovran, E., and Martinez-Gomez, N. C. (2016). Pyrroloquinoline quinone-containing ethanol dehydrogenase in *Methylobacterium extorquens* AM1 extends lanthanide-dependent metabolism to multi-carbon substrates. *J. Bacteriol.* 198, 3109–3118. doi: 10.1128/JB.00478-16

Hardacre, C. J., and Heal, M. R. (2013). Characterization of methyl bromide and methyl chloride fluxes at temperate freshwater wetlands. *J. Geophys. Res. Atmos.* 118, 977–991. doi: 10.1029/2012jd018424

Kallen, R. G., and Jencks, W. P. (1966). The mechanism of the condensation of formaldehyde with tetrahydrofolic acid. *J. Biol. Chem.* 241, 5851–5863.

Kayser, M. F., Ucurum, Z., and Vuilleumier, S. (2002). Dichloromethane metabolism and C1 utilization genes in *Methylobacterium* strains. *Microbiology* 148, 1915–1922. doi: 10.1099/00221287-148-6-1915

Kayser, M. F., and Vuilleumier, S. (2001). Dehalogenation of dichloromethane by dichloromethane dehalogenase/glutathione *S*-transferase leads to the formation of DNA adducts. *J. Bacteriol.* 183, 5209–5212. doi: 10.1128/JB.183.17.5209-5212.2001

Kitagawa, M., Miyakawa, M., Matsumura, Y., and Tsuchido, T. (2002). *Escherichia coli* small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants: *E. coli* small heat shock proteins. *Eur. J. Biochem.* 269, 2907–2917. doi: 10.1046/j.1432-1033.2002.02958.x

Kolb, S. (2009). Aerobic methanol-oxidizing bacteria in soil. *FEMS Microbiol. Lett.* 300, 1–10. doi: 10.1111/j.1574-6968.2009.01681.x

La Roche, S. D., and Leisinger, T. (1991). Identification of *dcmR*, the regulatory gene governing expression of dichloromethane dehalogenase in *Methylobacterium* sp. strain DM4. *J. Bacteriol.* 173, 6714-6721. doi: 10.1128/jb.173.21.6714-67

Leisinger, T., Bader, R., Hermann, R., Schmid-Appert, M., and Vuilleumier, S. (1994). Microbes, enzymes and genes involved in dichloromethane utilization. *Biodegradation* 5, 237–248. doi: 10.1007/BF00696462

Lopez-Marques, R. L., Perez-Castineira, J. R., Losada, M., and Serrano, A. (2004). Differential regulation of soluble and membrane-bound inorganic pyrophosphatases in the photosynthetic bacterium *Rhodospirillum rubrum* provides insights into pyrophosphate-based stress bioenergetics. *J. Bacteriol.* 186, 5418–5426. doi: 10.1128/JB.186.16.5418-5426.2004

Marx, C. J., Bringel, F., Chistoserdova, L., Moulin, L., Farhan Ul Haque, M., Fleischman, D. E., et al. (2012). Complete genome sequences of six strains of the genus *Methylobacterium*. *J. Bacteriol.* 194, 4746–4748. doi: 10.1128/jb.01009-12

Marx, C. J., and Lidstrom, M. E. (2004). Development of an insertional expression vector system for *Methylobacterium extorquens* AM1 and generation of null mutants lacking *mtdA* and/or *fch*. *Microbiology* 150, 9–19. doi: 10.1099/mic.0.26587-0

McCarthy, T. V., and Lindahl, T. (1985). Methyl phosphotriesters in alkylated DNA are repaired by the Ada regulatory protein of *Escherichia coli*. *Nucl. Acids Res.* 13, 2683–2698. doi: 10.1093/nar/13.8.2683

Michener, J. K., Camargo Neves, A. A., Vuilleumier, S., Bringel, F., and Marx, C. J. (2014a). Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post-transfer refinement. *eLife* 3:4279. doi: 10.7554/eLife.04279

Michener, J. K., Vuilleumier, S., Bringel, F., and Marx, C. J. (2014b). Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. *J. Bacteriol.* 196, 2101–2107. doi: 10.1128/jb.00034-14

Michener, J. K., Vuilleumier, S., Bringel, F., and Marx, C. J. (2016). Transfer of a catabolic pathway for chloromethane in *Methylobacterium* strains highlights different limitations for growth with chloromethane or with dichloromethane. *Front. Microbiol.* 7:1116. doi: 10.3389/fmicb.2016.01116

Muller, E. E. L., Bringel, F., and Vuilleumier, S. (2011a). Dichloromethane-degrading bacteria in the genomic age. *Res. Microbiol.* 162, 869–876. doi: 10.1016/j.resmic.2011.01.008

Muller, E. E. L., Hourcade, E., Louhichi-Jelail, Y., Hammann, P., Vuilleumier, S., and Bringel, F. (2011b). Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4. *Environ. Microbiol.* 13, 2518–2534. doi: 10.1111/j.1462-2920.2011.02524.x

Nadalig, T., Greule, M., Bringel, F., Keppler, F., and Vuilleumier, S. (2014). Probing the diversity of chloromethane-degrading bacteria by comparative genomics and isotopic fractionation. *Front. Microbiol* 5:523. doi: 10.3389/fmicb.2014.00523

Nunn, D. N., and Lidstrom, M. E. (1986). Phenotypic characterization of 10 methanol oxidation mutant classes in *Methylobacterium* sp. strain AM1. *J. Bacteriol.* 166, 591–597. doi: 10.1128/jb.166.2.591-597.1986

Ochsner, A. M., Christen, M., Hemmerle, L., Peyraud, R., Christen, B., and Vorholt, J. A. (2017). Transposon sequencing uncovers an essential regulatory function of phosphoribulokinase for methylotrophy. *Curr. Biol.* 27, 1–10. doi: 10.1016/j.cub.2017.07.025

Ochsner, A. M., Sonntag, F., Buchhaupt, M., Schrader, J., and Vorholt, J. A. (2015). *Methylobacterium extorquens*: methylotrophy and biotechnological applications. *Appl. Microbiol. Biotechnol.* 99, 517–534. doi: 10.1007/s00253-014-6240-3

Rhew, R. C., Whelan, M. E., and Min, D. H. (2014). Large methyl halide emissions from south Texas salt marshes. *Biogeoscience* 11, 6427–6434. doi: 10.5194/bg-11-6427-2014

Ritz, C., and Spiess, A. N. (2008). *qpcR*: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics* 24, 1549–1551. doi: 10.1093/bioinformatics/btn227

Roselli, S., Nadalig, T., Vuilleumier, S., and Bringel, F. (2013). The 380 kb pCMU01 plasmid encodes chloromethane utilization genes and redundant genes for vitamin B12- and tetrahydrofolate-dependent chloromethane metabolism in *Methylobacterium extorquens* CM4: a proteomic and bioinformatics study. *PLoS ONE* 8:e56598. doi: 10.1371/journal.pone.0056598

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73

Sluis, M. K., Larsen, A. R., Krum, J. G., Anderson, R., Metcalf, W. W., and Ensign, S. A. (2002). Biochemical, molecular, and genetic analyses of the acetone carboxylases from *Xanthobacter autotrophicus* strain Py2 and *Rhodobacter capsulatus* strain B10. *J. Bacteriol.* 184, 2969–2977. doi: 10.1128/JB.184.11.2002

Studer, A., McAnulla, C., Büchele, R., Leisinger, T., and Vuilleumier, S. (2002). Chloromethane induced genes define a third C1 utilization pathway in

Methylobacterium chloromethanicum CM4. *J. Bacteriol.* 184, 3476–3484. doi: 10.1128/JB.184.13.3476-34

Studer, A., Stupperich, E., Vuilleumier, S., and Leisinger, T. (2001). Chloromethane:tetrahydrofolate methyl transfer by two proteins from *Methylobacterium chloromethanicum* strain CM4. *Eur. J. Biochem.* 268, 2931–2938. doi: 10.1046/j.1432-1327.2001.02182.x

Studer, A., Vuilleumier, S., and Leisinger, T. (1999). Properties of the methylcobalamin: H_4 folate methyltransferase involved in chloromethane utilization by *Methylobacterium* sp. strain CM4. *Eur. J. Biochem.* 264, 242–249. doi: 10.1046/j.1432-1327.1999.00629.x

Torgonskaya, M. L., Doronina, N. V., Hourcade, E., Trotsenko, Y. A., and Vuilleumier, S. (2011). Chloride-associated adaptive response in aerobic methylotrophic dichloromethane-utilising bacteria. *J. Basic Microbiol.* 51, 296–303. doi: 10.1002/jobm.201000280

Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., et al. (2017). MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucl. Acids Res.* 45, D517–D528. doi: 10.1093/nar/gkw1101

van der Meer, J. R., and Sentchilo, V. (2003). Genomic islands and the evolution of catabolic pathways in bacteria. *Curr. Opin. Biotechnol.* 14, 248–254. doi: 10.1016/S0958-1669(03)00058-2

Vannelli, T., Messmer, M., Studer, A., Vuilleumier, S., and Leisinger, T. (1999). A corrinoid-dependent catabolic pathway for growth of a *Methylobacterium* strain with chloromethane. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4615–4620. doi: 10.1073/pnas.96.8.4615

Vernikos, G. S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22, 2196–2203. doi: 10.1093/bioinformatics/btl369

Vuilleumier, S., Chistoserdova, L., Lee, M.-C., Bringel, F., Lajus, A., Zhou, Y., et al. (2009). *Methylobacterium* genome sequences: a reference blueprint to investigate microbial metabolism of C1 compounds from natural and industrial sources. *PLoS ONE* 4:e5584. doi: 10.1371/journal.pone.0005584

Vuilleumier, S., and Leisinger, T. (1996). Protein engineering studies of dichloromethane dehalogenase /glutathione S-transferase from *Methylophilus* sp. strain DM11 Ser12 but not Tyr6 is required for enzyme activity. *Eur. J. Biochem.* 239, 410–417. doi: 10.1111/j.1432-1033.1996.0410u.x

Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W. F., et al. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinform.* 7:142. doi: 10.1186/1471-2105-7-142

Yang, Y.-X., Wang, M.-M., Yin, Y.-L., Onac, E., Zhou, G.-F., Peng, S., et al. (2015). RNA-seq analysis reveals the role of red light in resistance against *Pseudomonas syringae* pv. tomato DC3000 in tomato plants. *BMC Genom.* 16:120. doi: 10.1186/s12864-015-1228-7

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.

Copyright © 2017 Chaignaud, Maucourt, Weiman, Alberti, Kolb, Cruveiller, Vuilleumier and Bringel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

3.2.2 Comparaison de l'adaptation aux méthanes chlorés entre *M. extorquens* CM4 et DM4

L'approche de RNA-seq a permis d'étudier l'adaptation de deux souches de *M. extorquens* aux méthanes chlorés (Chaignaud *et al.*, 2017). Les souches *M. extorquens* DM4 et CM4 sont capables de se développer en présence de DCM ou de chlorométhane, respectivement. Le but de l'étude était d'analyser l'adaptation à deux polluants chlorés proches chez deux souches de la même espèce pour voir si des mécanismes communs étaient présents. Par comparaison des génomes de cinq souches de *Methylorubrum extorquens*, une liste de gènes partagés systématiquement dans le génome de chaque souche, appelée *core genome*, a été établie. Parmi ces gènes, 12 gènes du *core genome* de *M. extorquens* ont été trouvés comme étant différentiellement exprimés : six en présence de méthanol et six en présence de méthane chlorés (Fig. 1). Parmi ceux-ci *degP* et *hppA* sont les seuls ayant une fonction connue.

Parmi les 163 gènes partagés uniquement par *M. extorquens* CM4 et DM4, aucun n'a été détecté comme différentiellement exprimé chez les deux souches (Fig. 1 ; ces gènes sont représentés par les points rouges). La réponse aux méthanes chlorés repose donc principalement sur l'expression des gènes spécifiques de chaque souche. Cela peut être corrélé avec des études précédentes qui indiquent que l'apport du gène *dcmA* chez des souches de *Methylorubrum* ne permet pas systématiquement une croissance avec ce polluant (Michener *et al.*, 2014).

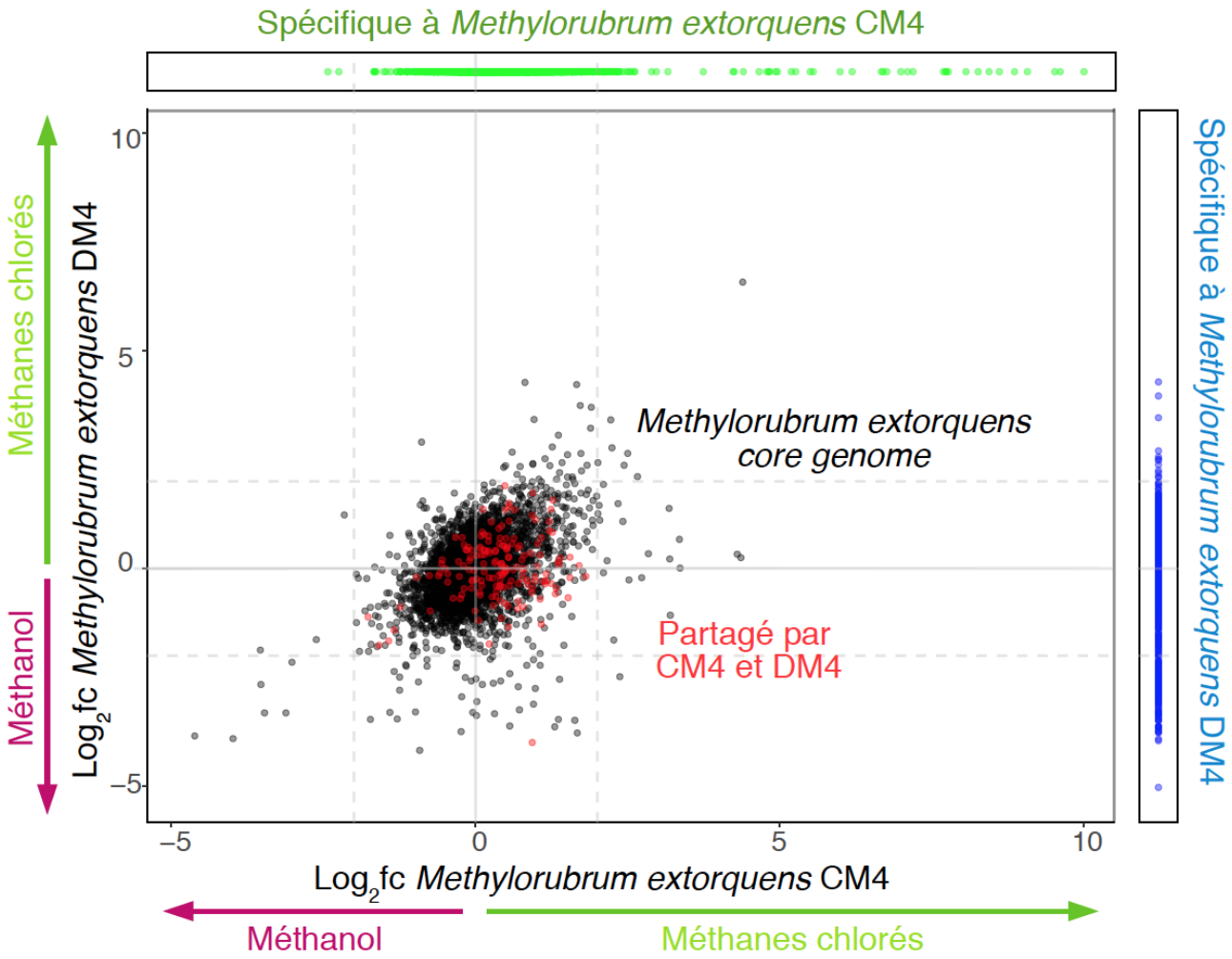


Figure 3.1: Comparaison des résultats de RNA-seq pour les souches *M. extorquens* DM4 et CM4. Les souches ont été cultivées en présence de méthanol ou d'un méthane chloré (DCM pour DM4 et chlorométhane pour CM4). Les lignes en pointillées indiquent la limite entre les gènes différentiellement exprimés ($\log_2$ fc ≥ 2) ou non différentiellement exprimés ($\log_2$ fc ≤ 2). Les gènes spécifiques à *M. extorquens* DM4 et CM4 sont indiqués en bleu et en vert respectivement.

Au cours de cette thèse, les données de transcriptomique de *M. extorquens* DM4 ont été validées par une approche de RT-qPCR avec des réplicats indépendants d'ARN. Une bonne corrélation est observée entre les deux méthodes, présentée en Fig S2 des données supplémentaires de l'article (Chaignaud *et al.*, 2017). Cette approche a également permis de vérifier que certains gènes prédits comme étant régulés par RNA-seq mais avec une mauvaise p-value était bien différentiellement exprimés.

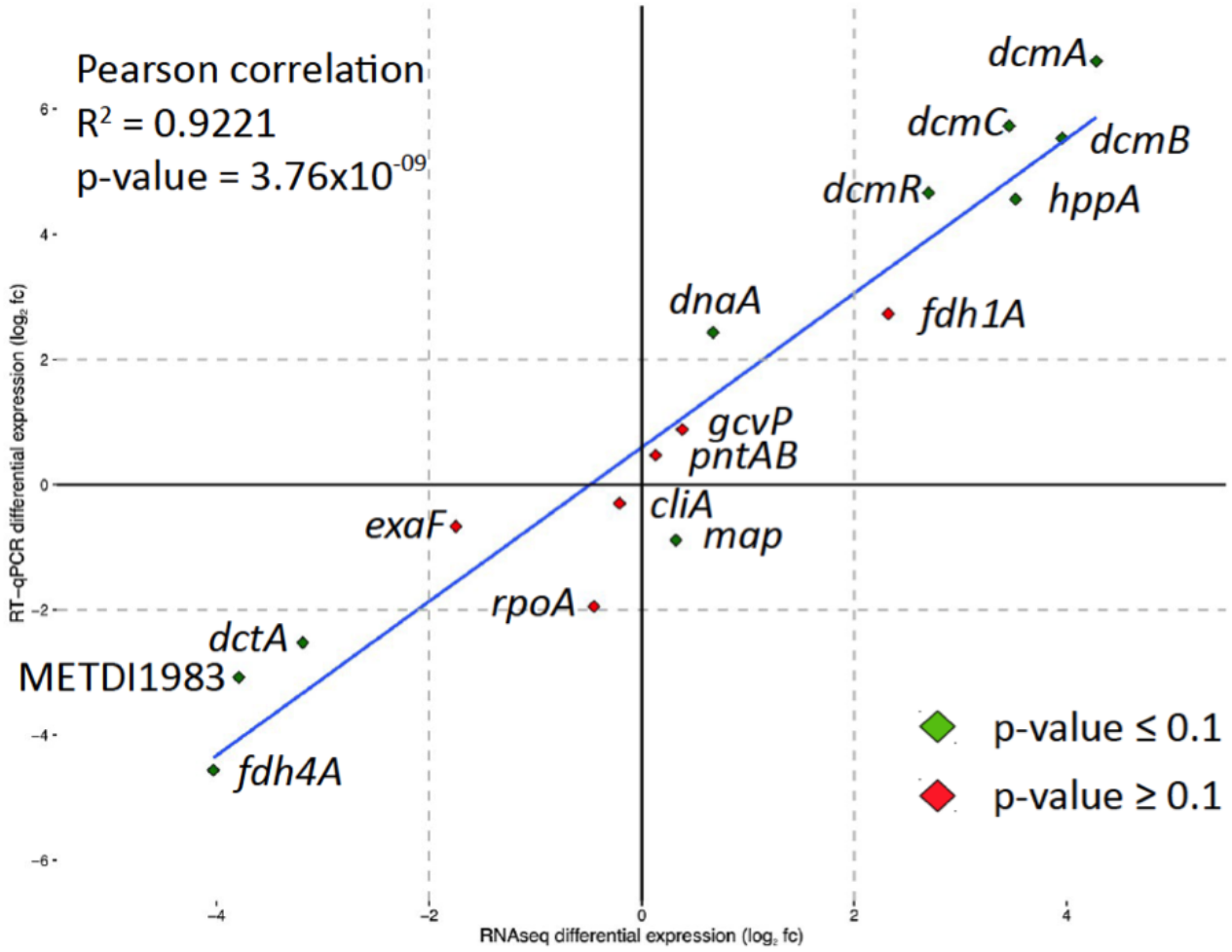


Figure 3.2: Comparaison des résultats de RNA-seq et RT-qPCR

3.3 Approche protéomique pour étudier les protéines chez *M. extorquens* DM4 en présence de DCM

3.3.1 Article de Bibi-Triki *et al*. paru en 2018

L'article de Bibi-triki *et al*., 2018 publié dans le journal *Journal of Proteomics* se trouve dans les pages suivantes. Dans le cadre de cette thèse l'extraction de protéines a été réalisée pour l'approche de protéogénomique. La présence de peptides signaux a été recherchée parmi les protéines dont l'extrémité NH_2 terminale a été validée. L'analyse des différences d'abondances des protéines entre deux conditions de culture a aussi été réalisée.

Journal of Proteomics 179 (2018) 131–139

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

N-terminome and proteogenomic analysis of the *Methylobacterium extorquens* DM4 reference strain for dichloromethane utilization

Sabrina Bibi-Triki[a,1], Gauthier Husson[b,1], Bruno Maucourt[a], Stéphane Vuilleumier[a], Christine Carapito[b,**], Françoise Bringel[a,*]

[a] Université de Strasbourg, CNRS, GMGM UMR 7156, Department of Microbiology, Genomics and the Environment, Strasbourg, France
[b] Université de Strasbourg, CNRS, IPHC UMR 7178, Laboratoire de Spectrométrie de Masse BioOrganique, Strasbourg, France

ARTICLE INFO

Keywords:
Proteogenomics
N-terminome
Gene discovery
Genome annotation
One-carbon metabolism
Dichloromethane

ABSTRACT

Methylobacterium strains can use one-carbon compounds, such as methanol, for methylotrophic growth. In addition to methanol, a few strains also utilize dichloromethane, a major industrial chlorinated solvent pollutant. With a fully assembled and annotated genome, *M. extorquens* DM4 is the reference bacterium for aerobic dichloromethane degradation. The doublet N-terminal oriented proteomics (dN-TOP) strategy was applied to further improve its genome annotation and a differential proteomics approach was performed to compare *M. extorquens* DM4 grown either with methanol or dichloromethane as the sole source of carbon and energy. These approaches led to experimental confirmation of 259 hypothetical proteins, correction of 78 erroneous predicted start codons, discovery of 39 new proteins and annotation of 66 signal peptides, including essential enzymes involved in methylotrophic growth.

Significance: Dichloromethane (methylene chloride, CH_2Cl_2, DCM) is one of the most widely used industrial halogenated solvents and a potential carcinogen. Microbial rehabilitation of worldwide-contaminated sites involves DCM breakdown by bacteria that are able to grow using this pollutant as their sole carbon and energy source. The most-studied methylotrophic DCM degrader is *Methylobacterium extorquens* strain DM4. Proteomic studies of the *Methylobacterium* genus have been performed previously, but genome-wide investigation of N-termini of expressed proteins has not yet been performed. Differential quantitative proteomic analysis also opens new research perspectives to better monitor and understand bacterial growth with DCM.

1. Introduction

Automatic genome annotation pipelines are error-prone. Use of experimental-based protein analysis data, when available, allows for a fine curation of the overall annotation quality [1]. In particular, up to 10% estimated error rates are anticipated for translational start codon localization (see references within [2]). In this context, the development of high-throughput N-terminomic strategies offers a unique opportunity to increase the efficiency of genome curation (reviewed by [3]). Among these, the doublet N-terminal oriented proteomics (dN-TOP) strategy is based on the specific labelling of free N-terminal alpha amine function of proteins with light and heavy trimethoxyphenyl phosphonium (TMPP) [4]. This method detects both free N-terminal peptides with an acute sensitivity and internal peptides in a single experiment [5]. TMPP labelling generally allows identifying protein N-termini (N-ter) representing 5 to 8% of the total identified polypeptides. dN-TOP can curate bacterial genomes, such as *Mycobacterium smegmatis* [6], the arsenite-oxydizing bacterium *Herminiimonas arsenicoxydans* [4] and the human mitochondrial proteome [4,7,8]. N-terminomic approaches can correct erroneously predicted translational start sites, and identify proteins having their initiator methionine (iMet) cleaved or not by the N-terminal methionine excision (NME) process, an ubiquitous process involved in regulation of protein turnover [9]. It also experimentally confirms protein precursors subjected to signal peptide cleavage by signal peptidases [10] of both the Sec-type allowing the translocation of unfolded proteins across the membrane, and that of the Tat-type allowing folded protein translocation [11].

Methylobacterium constitutes an attractive candidate for methanol-based product biosyntheses [12] and for bioremediation strategies of polluted sites contaminated by chlorinated methanes [13]. Within the genus *Methylobacterium*, the plasmidless *M. extorquens* strain PA1 has emerged as the methylotroph model, notably for growth with the

* Correspondence to: F. Bringel, 28 rue Goethe, 67000 Strasbourg, France.
** Correspondence to: C. Carapito, 25 rue Becquerel, 67087 Strasbourg, France.
E-mail addresses: ccarapito@unistra.fr (C. Carapito), francoise.bringel@unistra.fr (F. Bringel).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.jprot.2018.03.012
Received 26 January 2018; Received in revised form 28 February 2018; Accepted 16 March 2018
Available online 19 March 2018
1874-3919/ © 2018 Elsevier B.V. All rights reserved.

Table 1
Overview of *M. extorquens* DM4 proteogenomic analysis and discovery of new proteins.

Proteogenomic data	Number of identified CDS (discovered)		
	Chromosome 5.94 Mb, 68% GC	p1METDI 142 Kb, 65% GC	p2METDI 39 Kb, 64% GC
Previously annotated CDS[a]	5958	149	41
Differential proteomics	2079 (9)	46 (8)	11 (0)
dN-TOP[b]	2737 (37)	56 (11)	16 (0)
TMPP-labelled	742 (10)	10 (1)	4 (0)
Internal unlabelled	2700 (27)	55 (10)	16 (0)
Combined proteome coverage[c] [% of total protein identified]	2798 (39) [47%]	63 (16) [42%]	17 (0) [42%]

[a] [17].
[b] 718 proteins were identified by both TMPP-labelled peptides and internal unlabelled peptides by the dN-TOP approach.
[c] Total non-redundant CDS identified using both differential proteomics and dN-TOP approaches.

reduced one-carbon compound methanol (CH_3OH) [14]. Nonetheless, this reference strain is unable to grow with chlorinated methanes, including dichloromethane (CH_2Cl_2) [15]. On the other hand, *M. extorquens* DM4 is the best-characterized dichloromethane-degrading isolate. This strain was isolated from industrial wastewater sludge in Switzerland [16] and its high GC containing genome was manually annotated by an international consortium [17] using the Genoscope platform [18]. Its genome of 6.12 Mb is composed of a chromosome of 5.94 Mb and two plasmids p1METDI of 141 kb and p2METDI of 38 kb. It is composed of 6148 annotated CDS, with 5958 CDS encoded on the chromosome and 149 and 41 annotated CDS encoded on the p1METDI and p2METDI plasmids, respectively (Table 1). Its ability to grow with dichloromethane as sole source of carbon and energy is due to the presence of the dichloromethane dehalogenase-encoding *dcmA* gene located within a catabolic transposon highly conserved in other sequenced genomes of dichloromethane-utilizing strains, including those of different genus [13,19]. Genome-wide random mutagenesis, comparative proteomics determined by 2D-gel analysis [20], comparative transcriptomics [15] and laboratory evolution studies [21] have revealed the molecular adaptive response to dichloromethane utilization.

This study aimed to curate the genome annotation of *M. extorquens* DM4 for proteins produced in growth with dichloromethane and the non-chlorinated reference methylotrophic substrate, methanol. To discover new proteins, the mass spectrometry (MS)-datasets collected using two complementary proteomic approaches (dN-TOP and differential proteomics) were searched with several databases, including a six-frame-translated genome database. In addition to proteins involved in C1 assimilation and dissimulation, we focused on the global response to growth with dichloromethane compared to that with methanol.

2. Material and methods

2.1. Growth, harvesting of bacteria and protein extraction

M. extorquens DM4 was grown aerobically at 30 °C on a rotary shaker at 120 rpm (Multitron, Infors HT) in 1 L Erlenmeyer flasks containing 220 mL of M3Cl0 medium, as published in [22] except for the substitution of $CuCl_2$ by $CuSO_4$ in the trace element solution. The carbon and energy source dichloromethane or methanol was added at 10 mM in the sterile M3Cl0, as previously described [15]. Erlenmeyer used for growth with dichloromethane was sealed with gas-tight screw caps (Supelco mininert valve, Sigma-Aldrich). When cultures reached mid-exponential growth phase (O.D. at 600 nm of 0.15), bacteria were centrifuged 10 min at 4 °C at 6000*g* (Heraeus Multifuge X1R, Thermo Fischer Scientific) and pellets were stored at −20 °C until use. Cell pellets were re-suspended in 800 µL cold DNase, RNase, protease-free water (Sigma-Aldrich) containing anti-protease cocktail (Complete Protease Inhibitor Cocktail, Roche). Glass beads (700 mg; 0.1 mm, ThermoFischer Scientific) were added prior to mechanical disruption (RETSCH MM2 Pulverizer Mixer Mill, Hamilton Instrument). Three shaking cycles at maximum speed for 1 min interspaced by on ice incubation for 1 min, were performed. Cell lysates were centrifuged at 11,228*g* at 4 °C. Proteins were precipitated overnight at −20 °C by mixing 600 µL of supernatant with 4 volumes of ice-cold acetone, harvested by centrifugation at 14,000*g* for 30 min at 4 °C (Heraeus Multifuge X1R, Thermo Fischer Scientific) and re-suspended in derivatization buffer (50 mM Tris-HCl, 6 M urea, 2 M thiourea, 1% SDS, pH 8.2). Protein concentration was determined by RC DC Protein Assay kit (Bio-Rad) following manufacturer's instructions.

2.2. N-terminal protein derivatization

A total of 100 µg of proteins were reduced during 1 h under shaking in presence of 0.9 µL of 200 mM tributylphosphine. Proteins were then alkylated by incubating for 1 h in the dark while shaking in presence of 50 mM iodoacetamide (2.8 µL and 1.2 µL of 700 mM iodoacetamide for dichloromethane and methanol growth conditions, respectively). 4 µL of an equimolar solution of light and heavy TMPP (trimethoxyphenyl phosphonium, 100 mM, 80% acetonitrile, 200:1 reagent:protein molar ratio, 1 h shaking) were added. Light- and heavy-labelled TMPP induce mass shifts on peptides of 572.18 Da and 581.21 Da, respectively. The derivatization reaction was stopped by adding 100 mM hydroxylamine (2.8 µL and 1.3 µL of 16.3 M hydroxylamine for dichloromethane and methanol growth conditions, respectively).

2.3. SDS-PAGE

100 µg of TMPP-labelled proteins were loaded on SDS-PAGE gel and separated in 8 bands for the dN-TOP analysis, while 30 µg of unlabelled proteins were loaded on SDS-PAGE gel in technical triplicates and stacked in a single band for the differential proteomic analysis. Then, gel bands of proteins were excised, cut in small pieces, in-gel reduced (10 mM dithiothreitol in 25 mM NH_4HCO_3), alkylated (55 mM iodoacetamide in 25 mM NH_4HCO_3), and digested overnight using modified porcine trypsin (Promega) at 37 °C.

2.4. NanoLC-MS/MS analysis

NanoLC-MS/MS analyses were performed on a nanoAcquity UPLC device (Waters) coupled to a Q-Exactive Plus mass spectrometer (Thermo Scientific, Bremen, Germany). Peptides were separated on an Acquity UPLC BEH130 C18 column (250 mm × 75 µm with 1.7 µm diameter particles) and a Symmetry C18 precolumn (20 mm × 180 µm with 5 µm diameter particles; Waters). The solvent system consisted of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Samples were loaded on the enrichment column over 3 min at 5 µL/min with 1% solvent B. Peptides were eluted at 450 nL/min with the following gradient of solvent B: for the dN-TOP analysis, from 1 to 8% over 2 min, 8 to 40% over 58 min, and 40 to 90% over 1 min; for differential proteomic analysis, from 1 to 20% over 68 min, 20 to 35% over 17 min, and 35 to 90% over 1 min. The MS capillary voltage was set to 1.8 kV at 250 °C. The system operated in a data-dependent acquisition mode with automatic switching between MS (mass range 300–1800 *m/z* with a resolving power of 70,000 at 200 m/z, automatic gain control fixed at 3×10^6 ions, and a maximum injection time set at 50 ms) and MS/MS (first fixed mass 100 m/z with a resolving power of 17,500 at 200 m/z, automatic gain control fixed at 1×10^5, and the maximal injection time set to 100 ms) modes. The 10 most abundant peptides (intensity threshold 2×10^5) were selected on

each MS spectrum for further isolation and higher energy collisional dissociation fragmentation, excluding unassigned and monocharged ions. Dynamic exclusion time was set to 60 s.

Complete LC-MS/MS datasets were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD006836 for dN-TOP data and PXD006838 for differential proteomics data [23].

2.5. Bioinformatic analysis

For mass spectrometry data peptide assignment, three databases were used. The "annotated CDS" database includes all 6148 chromosomal and plasmid-born annotated CDS of *M. extorquens* DM4 reference proteome [17]. The "elongated CDS" database corresponds to the "annotated CDS" database containing elongated N-termini (N-ter) using ATG/TTG/GTG as putative translation start codons. On the other hand, the "six-frame translation" database contains a total of 69,028 proteins including all possible translated sequences between 20 and 2395 residues in the six translation frames that differ from the 6148 proteins of the "annotated CDS" database. Overall, 24 sequences from the "six-frame translation" database that gave confident identification results were finally added to the "annotated CDS" database. Common contaminants and decoys were added to each database using the database generation toolbox of the MSDA pipeline [24].

2.5.1. Differential proteomic analysis

MS raw data were processed using MaxQuant version 1.5.3.30 [25]. Peptides and proteins were searched in the "annotated CDS" database using the Andromeda search engine. Fully tryptic peptides were searched with a minimum of 7 amino acids and one missed cleavage allowed. Cysteine carbamidomethylation was set as fixed modification and methionine oxidation as variable modification. Precursor mass tolerance was set to 20 ppm for the first search and 5 ppm for the main search. Fragment ion mass tolerance was set to 0.07 Da. "Second peptides" and "match between runs" options were enabled. False discovery rate threshold was set to 1% at both the peptide and protein levels using a target decoy strategy. Peptides and proteins were quantified using their LFQ intensities. Peptides with oxidized methionines and their unmodified counterpart peptides were excluded from the quantification, as well as shared peptides. Proteins with a fold change between both conditions above 1.5 or below 0.67 and a *p*-value below 0.05 were considered as differentially abundant.

2.5.2. dN-TOP analysis

Data were treated based on previous work [7]. In brief, raw data were searched with the MASCOT search engine (Matrix science, London, UK) on a local server. A first search to identify internal peptides was performed against the "annotated CDS" database. Fully tryptic peptides were searched with one missed cleavage allowed. Cysteine carbamidomethylation was set as fixed modification, and methionine oxidation as variable modification. Precursor mass tolerance was set to 5 ppm and the fragment ion mass tolerance was set to 0.07 Da. Search results were validated using the Proline Studio software (http://proline.profiproteomics.fr/) [26] to keep only identifications with a MASCOT Ion Score above 25, a pretty rank equal to 1, and a false discovery rate below 1% at the peptide (on e-value) and protein (on Mascot modified MudPIT score) levels. Identified spectra were then removed from the original peak lists using the Recover module of MSDA to create a subset peak list composed only of unidentified and high-quality spectra (at least 4 peaks higher than 3 times the intensity of the background noise). A second search was performed using these new peak lists against the "annotated CDS" database to identify TMPP-labelled N-termini. Semi-tryptic peptides were searched, with light (+572.18 Da) and heavy (+581.21 Da) TMPP derivatization on any N-terminal amino acid or side chain of tyrosine and lysine residues set as variable modifications. Searched fragment ions were a, b and y ions. All other parameters were identical to the first search. Search results were validated using the Proline Studio software to keep only identifications with a MASCOT Ion Score above 13 and a pretty rank equal to 1. Again, identified spectra were removed from the peak lists using Recover (identical parameters), and a third search was performed for the remaining spectra against the "elongated CDS" database to identify N-termini in the extended N-ter regions. Search and validation parameters were set as for the second search, and identified spectra were removed from the peak lists using Recover (identical parameters). A fourth search was performed for the remaining spectra against the "six-frame translation" database to identify N-termini of non-annotated proteins. Once again, search and validation parameters were set identical to the second search. Finally, all N-terminal peptides were validated only if both light and heavy TMPP-labelled forms were detected within a 30 s tolerance window, if the labelling was unambiguously detected in N-ter position and not on side chains of lysine or arginine residues, and if the peptide was unique in the database. This latter validation step was performed using the dN-TOP Validation Tool of the MSDA pipeline.

2.5.3. Validation of curated, novel protein and signal peptide cleavage site annotations

TMPP-labelled peptides which matched to the reference "annotated CDS" database were further analysed according to their matching localization within the annotated proteins. We distinguished the TMPP-labelled peptides with N-ter methionine matching the predicted iMet of annotated proteins from the TMPP-labelled peptides with N-ter matching the predicted N-ter penultimate amino-acid of annotated protein. In addition, when TMPP-labelled peptides were localized within annotated CDS sequences, we determined if the protein was misannotated (shorter protein), if the protein was submitted to signal peptide cleavage, or if the TMPP-labelled peptide was a non-specific cleaved peptide (degradation product). Remaining TMPP-labelled peptides that did not match to the "annotated CDS" database were searched in the "elongated CDS" and the "six-frame translation" databases. This allowed the detection of putative misannotated N-ter (longer proteins) or of new proteins (not previously annotated). To validate new protein annotations, we used the following criteria: the presence of RBS sequence in the proximity of the newly proposed N-ter amino acid, similarity with N-ter found in bacterial homologs, as well as TMPP-labelled peptide matching within conserved domains [27].

Potential signal peptide cleavage sites within TMPP-labelled peptides starting with an amino acid other than Met and matching inside a CDS sequence were searched. SignalP 4.1 [28] and TatP 1.0 [29] software were used to detect Sec and Tat signal peptide-associated consensus sequence motifs, respectively. Among the experimentally-detected signal peptides, sequence consensus motifs were created using WebLogo [30].

The *M. extorquens* DM4 genome sequence was upgraded (chromosome at https://www.genoscope.cns.fr/agc/microscope/mage/viewer.php?, plasmid p1METDI at https://www.genoscope.cns.fr/agc/microscope/mage/viewer.php?S_id=241, and plasmid p2METDI at https://www.genoscope.cns.fr/agc/microscope/mage/viewer.php?S_id=242) to include refinements of protein annotation (new, shorter and longer proteins, experimentally validated signal peptides).

3. Results and discussion

To discover new proteins and improve the quality of the genome annotation of the reference dichloromethane-degrader *M. extorquens* DM4, cultures were grown under two methylotrophic growth conditions: one with methanol, a reference carbon source for methylotrophy [14], and the other with dichloromethane, a chlorinated methane pollutant that is degraded by a small number of bacterial strains including *M. extorquens* DM4 [13]. The workflow included two complementary MS approaches (Fig. 1).

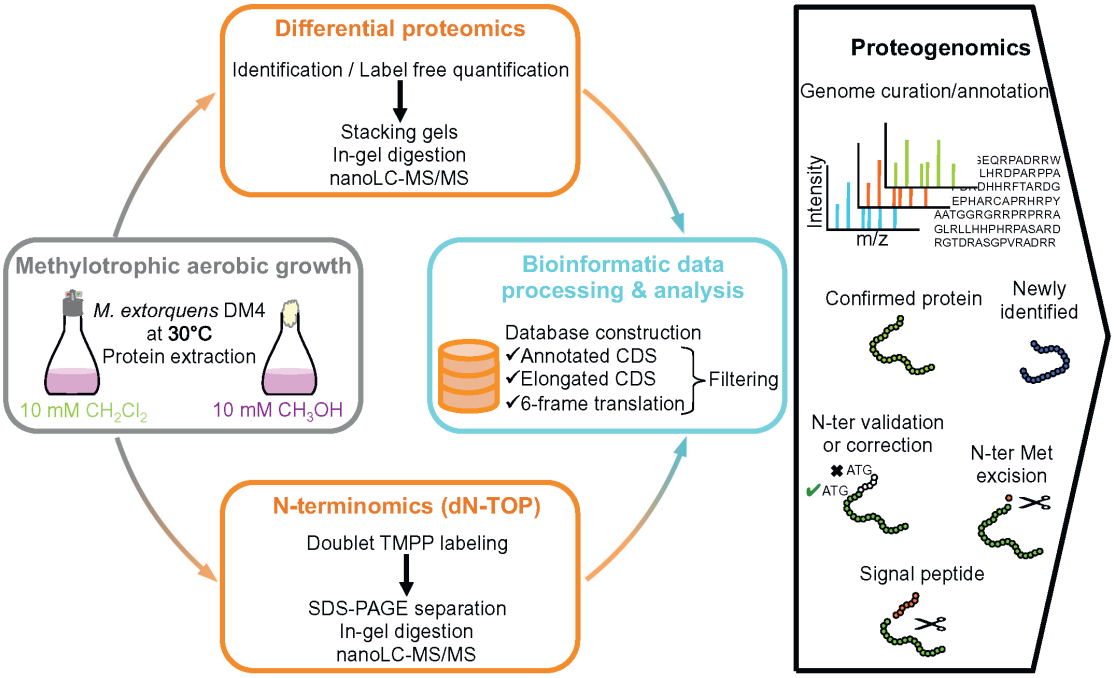


Fig. 1. Proteogenomics workflow for *M. extorquens* DM4 genome curation and annotation. *M. extorquens* DM4 wild-type was cultivated with either dichloromethane or methanol as the sole source of carbon and energy until mid-exponential phase was reached. Protein extracts were then analysed by two complementary MS approaches, namely differential proteomics and dN-TOP, allowing genome curation and annotation.

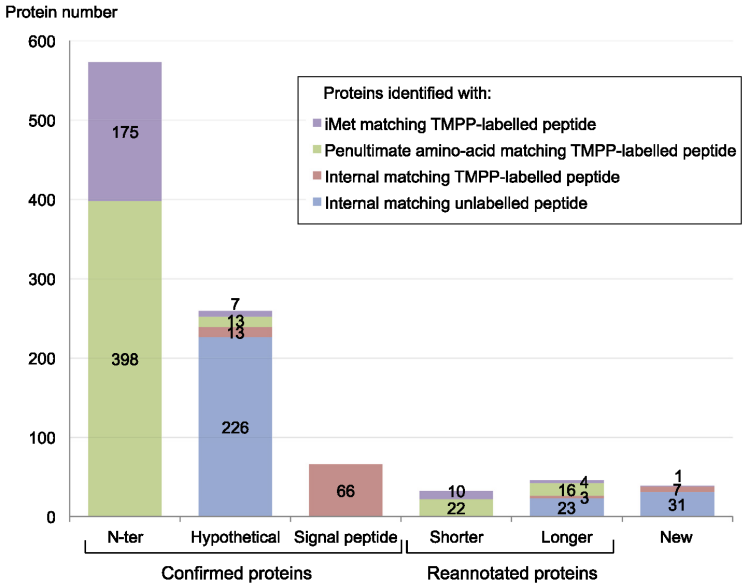


Fig. 2. Distribution of predicted/reannotated and newly identified proteins. Proteins were identified via TMPP-labelled or unlabelled peptides matching either the N-ter methionine, the penultimate amino-acid or inside the protein sequence.

3.1. Global proteome coverage

In total, 31,756 peptides were identified by the two combined MS approaches, and mapped to 2878 proteins out of 6148 CDS annotated on *M. extorquens* DM4 genome (Table 1). Among these, a total of 26,659 peptides (1090 TMPP-labelled and 25,569 unlabelled peptides) corresponding to 2737 unique proteins were identified using the dN-TOP approach. The dN-TOP strategy represents intrinsic advantages: i) TMPP-labelling offers increased hydrophobicity and ionisation efficiency to labelled peptides [6,7]; (ii) the pre-fractionation step (SDS-PAGE separation in 8 bands, see Material and Method section) included in the workflowcontributed to the high sequence coverage achieved. Conversely, no fractionation step was performed prior to LC-MS/MS analysis to perform quantification in the differential proteomics analysis. This approach allowed us to identify 22,521 peptides corresponding to 2136 unique proteins. The overlap between the two approaches is 78%. Globally, the two MS approaches covered 46.8% of the entire proteome. Proteome coverage was higher for the chromosome (47%) than for the two resident plasmids (around 42%). Proteins with a large diversity of pI (3.24 to 12.55) and molecular weight (2.4 to 375 kDa) were detected (Fig. S1). Out of 1430 hypothetical proteins with no conserved domain (COG hit) or with no homology to any previously reported sequences (on average < 25% identity over 80% of the protein size; class 5 and 6 in the MaGe platform [17]), 259 proteins (19%) were identified and validated as expressed proteins in *M. extorquens* DM4 (Table S1).

3.2. N-terminome study

A total of 1090 TMPP-labelled peptides matching 756 unique proteins and 25,569 internal peptides corresponding to 2771 unique proteins were identified by the dN-TOP approach (Table 1). Half of the N-ter of TMPP-labelled peptides matched to the N-ter of annotated CDS predicted *in silico* (Fig. 2), either exactly (Table S2) or to the penultimate amino acid (Table S3). This allowed us to validate approximately 9% of previously annotated N-ter. For the other half of the TMPP-labelled peptides, further matching analyses were performed using the "elongated CDS" and the "six-frame translation" databases to identify proteins submitted to signal peptide cleavage or misannotated N-ter, and new proteins. From the remaining TMPP-labelled peptides matching within the sequence of 283 proteins, we deduced that they corresponded to degradation products (Table S4).

3.2.1. Proteins subjected to N-terminal methionine excision

N-ter maturation of proteins by the NME process is ubiquitous [9]. *M. extorquens* DM4 harbours two methionine aminopeptidases encoding *map* genes: METDI1840 and METDI2981 that share at least 43% protein identity with one of the two characterized methionine aminopeptidases found in *Bacillus subtilis* [31]. Experimental evidence of cleavage after the iMet in *M. extorquens* DM4 was demonstrated thanks to the identification of 398 TMPP-labelled peptides. This corresponds to 40% of the detected TMPP-labelled peptides (Table S3), which is comparable to the 63% proportion of detected iMet cleaved proteins observed in other *Proteobacteria* [2,32]. NME generally occurs when small amino acids (Ala, Cys, Gly, Pro, Ser, Thr and Val) are present at position 2, whereas amino acids with higher gyration radius of their side-chain (Asn, Leu, Gln, Ile or Lys) are less susceptible to NME [33]. In agreement with this observation, the NME process in *M. extorquens* DM4 occurred predominantly for proteins with Ser, Thr or Ala on the second position, unlike proteins with no NME (Fig. 3A). Coexistence of both the precursor and the NME-cleaved form was found for 26 proteins. Of these, in agreement with published data [2], 16 proteins harbored residues containing longer lateral chains (Thr or Asn) (Table S3).

3.2.2. Signal peptides

Signal peptide cleavage was experimentally confirmed using the dN-TOP approach for 55 Sec-type and 11 Tat-type encoding precursors (Table S5). About 6% of the 1074 *in silico* predicted signal peptides containing precursors were thereby experimentally validated. A canonical "AxA" consensus cleavage site was experimentally found in *M. extorquens* DM4 (Fig. 3CD), which is not the case in some other *Alphaproteobacteria*, where Ser residues are dominant in −3 and −1 positions instead of Ala [34]. A canonical "RRxFL" Tat signal peptide conserved motif was also found for *M. extorquens* DM4 (Fig. 3B). The presence of a signal peptide which is responsible for the transport of key enzymes involved in C1 metabolism from the cytoplasm to the periplasm [35] was experimentally validated here for subunits of methanol dehydrogenases and associated components including MxaF [36], MxaD and cytochrome c_L [37], cytochrome c_H, cytochrome c oxidase, the PQQ-dependent methanol/ethanol dehydrogenase ExaF [38], as well as the signal peptide of formate dehydrogenase Fdh3A (Fig. 4B). Signal peptides were also experimentally confirmed for proteins involved in peptidoglycan synthesis and degradation, transport and stress response (Table S5; Fig. 4C).

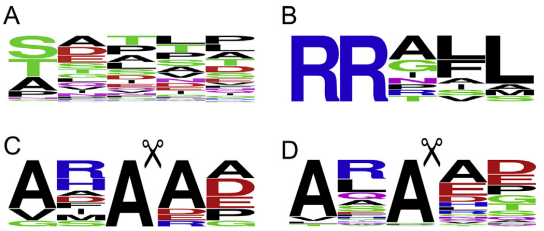


Fig. 3. Experimentally-determined N-ter sequence motifs. (A) Consensus sequence of the 5 first amino-acids after NME (N = 437). Small amino acids (Ala, Cys, Gly, Pro, Ser, Thr and Val) are predominantly found after the cleavage site; (B) Tat-type signal peptides consensus motif harbours the RRxFL sequence stretch at the interface between the positively-charged N-ter domain and the hydrophobic central domain of the signal peptide (N = 10) [48,49]; (C) Tat-type signal peptide cleavage sites (N = 10); (D) Sec-type signal peptide cleavage sites (N = 55).

3.2.3. Genome curation

A total of 78 translational start sites were corrected, based on N-ter localization of 281 TMPP-labelled and additional internal peptides. These proteins were classified in three categories (Tables S6, S7). The first category includes 32 proteins that were shortened by 2 to 32 residues with TMPP-labelled peptides matching inside their corresponding annotated CDS (Table S6, group 1). Newly proposed N-ter sequences were preceded by RBS sequences, and found in other homologous proteins within the species or in other species as checked using BlastP. Two of these proteins, DcmB and DcmC, are encoded by the *dcm* locus, which is essential for growth with dichloromethane and conserved within dichloromethane-degrading strains [20] (see Fig. 5A). In addition, the iMet was cleaved from 22 of these corrected proteins.

The second category includes 46 proteins that were elongated from 2 to 275 amino acids thanks to 79 TMPP-labelled or internal peptides matching upstream of currently annotated translation start sites (Table S6, group 2). Revised N-ter were directly confirmed by a matching TMPP-labelled peptide (20 proteins) or proposed based on the presence of a nearby RBS sequence and multi-sequence alignments with homologous proteins (26 proteins). Eight of these proteins were previously reannotated in a consistent manner [39]. N-ter reannotation included enzymes essential for growth with C1 compounds, e.g. 24 and 12 residue-long elongations for the serine-glyoxylate aminotransferase Sga and the methanol dehydrogenase large subunit MxaF [14], respectively (Table S6).

The third category contains 39 newly discovered proteins identified by one to 21 matching peptides and 168 TMPP-labelled and unlabelled

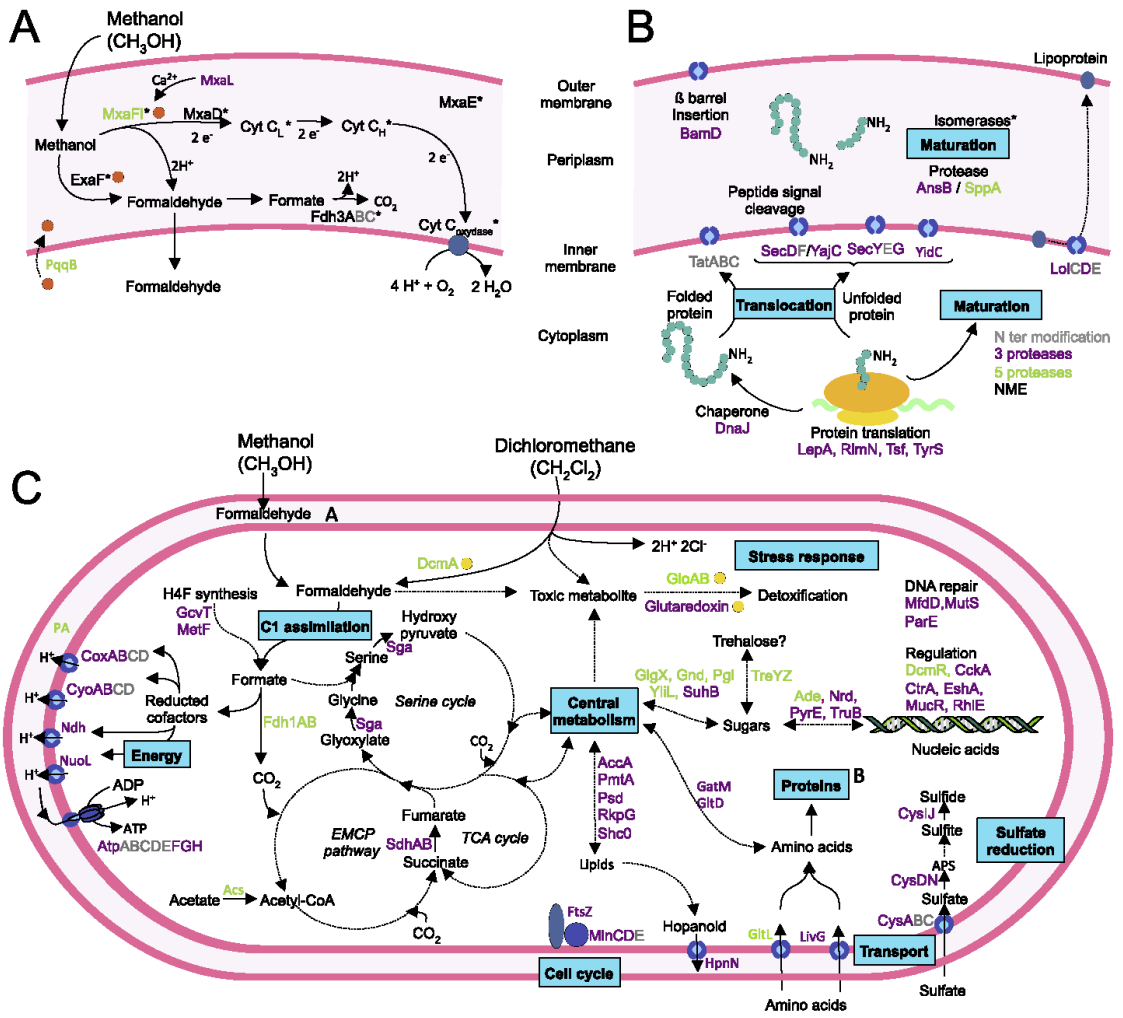


Fig. 4. Overview of the adaptive response of *M. extorquens* DM4 grown with different C1 compounds. Proteins found more abundant in dichloromethane or methanol grown cultures are indicated in green and purple, respectively. Full and dash arrows mean direct and multistep reactions, respectively. Names followed by an asterisk indicate proteins with experimentally detected signal peptides. Cofactors glutathione or PQQ (pyrroloquinoline quinone) are schematized with yellow and orange circles, respectively. APS, adenosine 5′-phosphosulfate; EMCP, ethylmalonyl-CoA pathway; H_4F, tetrahydrofolate; TCA, tricarboxylic acid; NME, N-terminal methionine excision; PA, pseudoazurin. (A) Formaldehyde is a common metabolic intermediate of both methanol and dichloromethane utilization pathways. Formaldehyde oxidation results in formate that is used for carbon dissimilation (energy) or biomass (via C1 assimilation). (B) Detailed periplasmic methanol oxidation processes. (C) Overview of protein modification and translocation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

peptides (Table S7 and Fig. 5B). METDICDS1000634D had been previously identified in another *M. extorquens* strain [39]. The newly annotated CDS METDI3692714D (23 residues) corresponds to the N-ter part of the commonly misannotated gene *prfB*, known to undergo a programmed +1 frameshift in many bacteria [40].

Fourteen discovered CDS were identified by one to 4 matching peptides (Table S7) and overlap previously annotated CDS on at least 50% of their CDS length. Among these, six newly discovered proteins have no homolog in the databases, which would suggest that these proteins may be unique to *M. extorquens* strain DM4. Evidence for overlapping CDS was obtained in only two cases as matching peptides were detected for both overlapping CDS: one peptide was detected for the newly discovered protein of unknown function METDI1071084D and p1METDI85354R, whereas two and 16 peptides matched their corresponding previously overlapping CDS METDI1162 (putative ferrichrome-iron receptor protein) and p1METDI0079 (NADPH quinone oxidoreductase), respectively (Table S7 and Fig. 5C). The overlapping CDS of 7 newly identified proteins encode proteins of unknown function conserved in other *Methylobacterium* species (METDI1650, METDI4437, p1METDI0140) or proteins with putative cellular functions: FAD-dependent pyridine nucleotide-disulphide oxidoreductase METDI0014, histidine kinase METDI1929 or transcriptional regulators METDI5083 and METDI5876. No conserved domains and no homolog in the databases were found for the previously annotated CDS

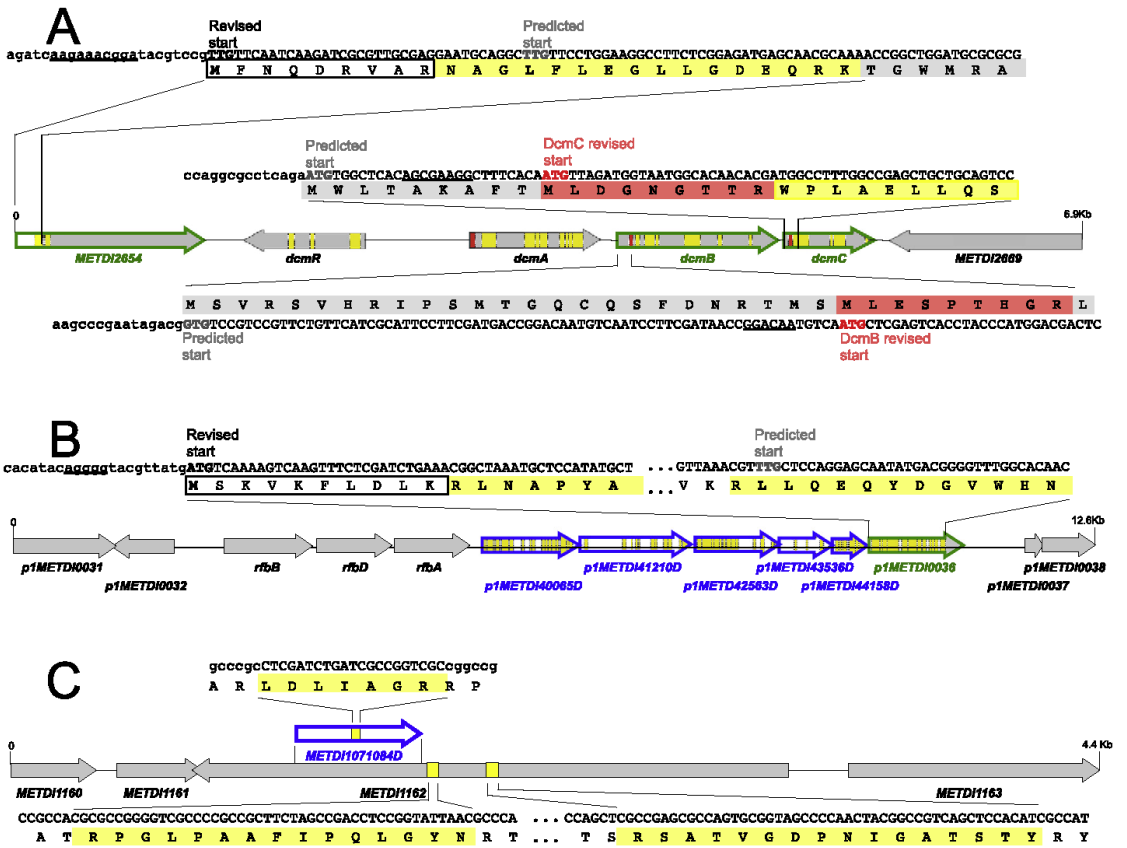


Fig. 5. Examples of genome curation. (A) Translational start site correction of the METDI2654, *dcmB* and *dcmC* CDS; (B) Identification of a new locus of 5 genes between *rfbA* and the translational start site corrected protein p1METDI0036; (C) The discovered p1METDI85354R CDS overlaps the previously annotated CDS p1METDI0079. Previously annotated genes are highlighted in grey. Corrected or newly identified coding sequences are framed in green and blue, respectively. Elongated parts of reannotated proteins are framed in black. Internal matching peptides and N-ter matching TMPP-labelled peptides are highlighted in yellow and red, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

METDI1395, METDI1971, METDI4374, METDI4680, and METDI5737, unlike their newly discovered overlapping CDS, which may suggest cases of incorrect predictions rather than overlapping genes.

3.3. *Differential proteome analysis of M. extorquens DM4 grown with dichloromethane versus methanol*

A total of 22,521 peptides was quantified with an FDR < 1%. Among the 789 proteins considered as statistically significant (p-value < 0.05), a majority (613 proteins) were considered as equally abundant in both cultures grown with dichloromethane compared to with methanol. Only 73 proteins (fold-change > 1.5) and 103 proteins (fold-change < 0.67) were respectively more and less abundant in cultures grown with dichloromethane compared to with methanol (Fig. 4 and Fig. S2; Table S8). A previous differential 2D-gel analysis testing similar conditions [20] had revealed only a small fraction (32 proteins) of the proteins detected in the present study (Fig. S1). This study provides the first experimental evidence of the translated products DcmB and DcmC encoded by the *dcmB* and *dcmC* genes of unknown function, which are conserved within the *dcm* islet of genome-sequenced dichloromethane-degrading bacteria [20]. The essential dichloromethane dehalogenase DcmA, DcmB and DcmC are among the most abundant proteins in cultures grown with dichloromethane (fold-change values of 2.44, 2.02 and 1.92, respectively). Genes *dcmA* and *dcmB* were also found among the most transcribed genes in the dichloromethane condition [15]. DcmR, the transcriptional regulator of DcmA, was detected here in both dichloromethane and methanol conditions with a fold-change value of 1.02 (p-value of 0.015), whereas in a previous 2D-gel proteomic analysis [13], it was only detected in the dichloromethane condition. Thus, this study demonstrates that DcmR is indeed present in presence of methanol, a condition where DcmR acts as a repressor [41]. Characterized C1 metabolism enzymes or cofactor biosynthesis proteins (PqqB; MxaIF; Fdh1AB) were also found to be more abundant in cultures grown with dichloromethane (fold-change values > 4) (Table S8, Fig. 4; see discussion about higher abundance of the methanol dehydrogenase MxaFI subunits in [15]).

Growth with dichloromethane is thought to involve multiple stresses [15,20,21,42]. Beside exposure to the solvent dichloromethane per se, toxic metabolites are produced during dichloromethane dehalogenation, as well as cytoplasmic pH drop and increased chloride ions. Among proteins more abundant in cultures grown with dichloromethane (Table S8), GloA-like lactoylglutathione lyase (METDI0511) and GloB hydroxyacylglutathione hydrolase (METDI3228) are predicted enzymes for the removal of toxic reactive

electrophilic species such as the alpha-oxaloadehydes glyoxal and methylglyoxal, which use glutathione (recent review see [43]), a cofactor shared with the dichloromethane dehalogenase DcmA. In other microorganisms, methylglyoxal is a side-product of different metabolic pathways, especially when the glycolysis flux is altered by increased dihydroxyacetone phosphate (DHAP), or from threonine and glycine catabolisms via aminoacetone, or from lipid peroxidation [43]. Sugar metabolism was shown recently to be essential for C1 assimilation in *M. extorquens* AM1 by controlling the metabolic flux balance between C1 assimilation for biomass and C1 dissimulation for energy ([14,44]; see Fig. 4). It is possible that carbon assimilation pathways including sugar metabolism may be disturbed during growth with dichloromethane compared to with methanol (10 proteins associated with sugar metabolism; Table S8).

In this study, growth with dichloromethane also correlated with higher abundance of enzymes potentially involved in the production of small organic molecules in response to stress in other microorganisms. In particular, trehalose reduces membrane permeability and stabilizes proteins, and was characterized in *M. extorquens* AM1 as a metabolite upregulated in solvent-stressed cultures in presence of 1-butanol [45]. In this study, a predicted trehalose biosynthesis gene cluster encoding three proteins was found to be more abundant in the dichloromethane condition (fold-change values > 1.5). Two of them are homologs of the characterized maltooligosyltrehalose trehalohydrolase TreZ and maltooligosyltrehalose synthase TreY, which in *Deinococcus radiodurans* catalyses the transformation of maltopentose into trehalose and maltotriose [46] (METDI3473 43% identity with TreZ; MEDTI3474 39% identity with TreY fused to a 4-alpha-glucanotransferase MalQ-like domain in strain DM4). Further genetic and biochemical investigations are needed to characterize the nature of compatible organic solutes in response to growth with dichloromethane.

Studies combining transcriptomic and proteomic methods have generally found a weak correlation between mRNA expression levels and protein abundance (reviewed by [47]). In this study, when differential proteomics was compared to expression profiles deduced from differentially expressed mRNA using comparative RNA-seq in similar conditions [15], the only difference found was for METDI4304. This putative stress response protein was detected as more abundant in the dichloromethane condition whereas it had higher mRNA levels in the methanol condition. This may suggest that METDI4304 could be subject to post-transcriptional regulation.

Finally, in methanol-grown cultures, proteins with increased abundance are often involved in functions such as protein translocation (Fig. 4C), electron transport chain associated with periplasmic methanol oxidation (Fig. 4B), sulfate reduction (Fig. 4A), and cell division (FtsZ and MinCD, Fig. 4C). Different proteases were more abundant in cultures grown with either dichloromethane or methanol, suggesting that different sets of proteins are targeted for protein turnover in the tested conditions.

4. Conclusion

M. extorquens DM4 is the reference bacterium for the biodegradation of dichloromethane, which is the worldwide most used chlorinated solvent found in many industrial contaminated sites. Better knowledge of the overall proteins specifically expressed during growth with dichloromethane would provide a reference proteome to help improve microbial-driven rehabilitation of polluted sites. Combined analysis of MS data obtained using N-terminomics and differential quantitative proteomics was used in this study to experimentally curate its genome. A total of 78 start codons were corrected, 39 new proteins were discovered, 66 signal peptides were annotated, and 259 hypothetical proteins were experimentally confirmed. Reannotation included C1 proteins such as DcmB, DcmC, Sga and MxaF. New research perspectives have been opened with the findings that predicted housekeeping enzymes for the synthesis of stress molecules like trehalose (TreYZ) and of metabolic detoxification (GloAB) were more abundant in dichloromethane growth condition.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2018.03.012.

Conflict of interest

Authors declare no conflicts of interest.

Acknowledgments

FB received from the Université de Strasbourg, a post-doctoral grant "Investissement d'Avenir" (Idex 2015) for SBT. BM is the recipient of a French Government PhD grant. GH is a recipient of a PhD grant from the "Association Nationale de la Recherche et de la Technologie" (CIFRE N°2014/0088). LABGeM (CEA/IG/Génoscope & CNRSUMR8030) and France Génomique National infrastructure (sponsored by the" Investissement d'Avenir " program managed by "Agence Nationale pour la Recherche"N° ANR-10-INBS-09) are acknowledged for support within the MicroScope annotation platform and for *in silico* protein library construction. This project was also supported by the French Proteomic Infrastructureand UCB Pharma (ProFI; ANR-10-INBS-08-03).

References

[1] J. Armengaud, Chapter twelve - reannotation of genomes by means of proteomics data, in: A.K. Shukla (Ed.), Methods Enzymology, Academic Press, 2017, pp. 201–216, , http://dx.doi.org/10.1016/bs.mie.2016.09.019.
[2] C. Bland, E.M. Hartmann, J.A. Christie-Oleza, B. Fernandez, J. Armengaud, N-terminal-oriented proteogenomics of the marine bacterium *Roseobacter denitrificans* Och114 using (N-succinimidyloxycarbonylmethyl) tris (2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP) labeling and diagonal chromatography, Mol. Cell. Proteomics 13 (2014) 1369–1381, http://dx.doi.org/10.1074/mcp.O113.032854.
[3] E.M. Hartmann, J. Armengaud, N-terminomics and proteogenomics, getting off to a good start, Proteomics 14 (2014) 2637–2646, http://dx.doi.org/10.1002/pmic.201400157.
[4] D. Bertaccini, S. Vaca, C. Carapito, F. Arsène-Ploetze, A. Van Dorsselaer, C. Schaeffer-Reiss, An improved stable isotope N-terminal labeling approach with light/heavy TMPP to automate proteogenomics data validation: dN-TOP, J. Proteome Res. 12 (2013) 3063–3070, http://dx.doi.org/10.1021/pr4002993.
[5] B. Westermann, A.S.V. Jacome, M. Rompais, C. Carapito, C. Schaeffer-Reiss, Doublet N-terminal oriented proteomics for N-terminomics and proteolytic processing identification, Protein Termin. Profiling, Humana Press, New York, NY, 2017, pp. 77–90, , http://dx.doi.org/10.1007/978-1-4939-6850-3_6.
[6] S. Gallien, E. Perrodou, C. Carapito, C. Deshayes, J.M. Reyrat, A.D. Van, O. Poch, C. Schaeffer, O. Lecompte, Ortho-proteogenomics: multiple proteomes investigation through orthology and a new MS-based protocol, Genome Res. 19 (2009) 128–135, http://dx.doi.org/10.1101/gr.081901.108.
[7] A.S. Vaca Jacome, T. Rabilloud, C. Schaeffer-Reiss, M. Rompais, D. Ayoub, L. Lane, A. Bairoch, A. Van Dorsselaer, C. Carapito, N-terminome analysis of the human mitochondrial proteome, Proteomics 15 (2015) 2519–2524, http://dx.doi.org/10.1002/pmic.201400617.
[8] C. Carapito, L. Kuhn, L. Karim, M. Rompais, T. Rabilloud, H. Schwenzer, M. Sissler, Two proteomic methodologies for defining N-termini of mature human mitochondrial aminoacyl-tRNA synthetases, Methods 113 (2017) 111–119, http://dx.doi.org/10.1016/j.ymeth.2016.10.012.
[9] C. Giglione, A. Boularot, T. Meinnel, Protein N-terminal methionine excision, Cell. Mol. Life Sci. CMLS 61 (2004) 1455–1474, http://dx.doi.org/10.1007/s00018-004-3466-8.
[10] I.J. Berry, J.R. Steele, M.P. Padula, S.P. Djordjevic, The application of terminomics for the identification of protein start sites and proteoforms in bacteria, Proteomics 16 (2016) 257–272, http://dx.doi.org/10.1002/pmic.201500319.
[11] I. Saraogi, S. Shan, Co-translational protein targeting to the bacterial membrane, Biochim. Biophys. Acta BBA - Mol. Cell Res. 1843 (2014) 1433–1441, http://dx.doi.org/10.1016/j.bbamcr.2013.10.013.
[12] A.M. Ochsner, F. Sonntag, M. Buchhaupt, J. Schrader, J.A. Vorholt, *Methylobacterium extorquens*: methylotrophy and biotechnological applications, Appl. Microbiol. Biotechnol. 99 (2015) 517–534, http://dx.doi.org/10.1007/s00253-014-6240-3.
[13] E.E.L. Muller, F. Bringel, S. Vuilleumier, Dichloromethane-degrading bacteria in the genomic age, Res. Microbiol. 162 (2011) 869–876, http://dx.doi.org/10.1016/j.resmic.2011.01.008.
[14] A.M. Ochsner, M. Christen, L. Hemmerle, R. Peyraud, B. Christen, J.A. Vorholt, Transposon sequencing uncovers an essential regulatory function of phosphoribulokinase for methylotrophy, Curr. Biol. 27 (2017) 2579–2588 e6 https://doi.org/10.1016/j.cub.2017.07.025.
[15] P. Chaignaud, B. Maucourt, M. Weiman, A. Alberti, S. Kolb, S. Cruveiller, S. Vuilleumier, F. Bringel, Genomic and transcriptomic analysis of growth-

supporting dehalogenation of chlorinated methanes in *Methylobacterium*, Front. Microbiol. 8 (2017, 1600), http://dx.doi.org/10.3389/fmicb.2017.01600.
[16] R. Gälli, T. Leisinger, Specialized bacterial strains for the removal of dichloromethane from industrial waste, Conserv. Recycl. (1985) 91–100, http://dx.doi.org/10.1016/0361-3658(85)90028-1.
[17] S. Vuilleumier, L. Chistoserdova, M.-C. Lee, F. Bringel, A. Lajus, Y. Zhou, B. Gourion, V. Barbe, J. Chang, S. Cruveiller, C. Dossat, W. Gillett, C. Gruffaz, E. Haugen, E. Hourcade, R. Levy, S. Mangenot, E. Muller, T. Nadalig, M. Pagni, C. Penny, R. Peyraud, D.G. Robinson, D. Roche, Z. Rouy, C. Saenampechek, G. Salvignol, D. Vallenet, Z. Wu, C.J. Marx, J.A. Vorholt, M.V. Olson, R. Kaul, J. Weissenbach, C. Médigue, M.E. Lidstrom, *Methylobacterium* genome sequences: a reference blueprint to investigate microbial metabolism of C1 compounds from natural and industrial sources, PLoS One 4 (2009) e5584, , http://dx.doi.org/10.1371/journal.pone.0005584.
[18] D. Vallenet, A. Calteau, S. Cruveiller, M. Gachet, A. Lajus, A. Josso, J. Mercier, A. Renaux, J. Rollin, Z. Rouy, D. Roche, C. Scarpelli, C. Médigue, MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes, Nucleic Acids Res. 45 (2017) D517–D528, http://dx.doi.org/10.1093/nar/gkw1101.
[19] F. Bringel, C.P. Postema, S. Mangenot, S. Bibi-Triki, P. Chaignaud, M.F.U. Haque, C. Gruffaz, L. Hermon, Y. Louhichi, B. Maucourt, E.E.L. Muller, T. Nadalig, A. Lajus, Z. Rouy, C. Médigue, V. Barbe, D.B. Janssen, S. Vuilleumier, Genome sequence of the dichloromethane-degrading bacterium *Hyphomicrobium sp.* strain GJ21, Genome Announc. 5 (2017) e00622-17, , http://dx.doi.org/10.1128/genomeA.00622-17.
[20] E.E.L. Muller, E. Hourcade, Y. Louhichi-Jelail, P. Hammann, S. Vuilleumier, F. Bringel, Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4, Environ. Microbiol. 13 (2011) 2518–2535, http://dx.doi.org/10.1111/j.1462-2920.2011.02524.x.
[21] J.K. Michener, A.A.C. Neves, S. Vuilleumier, F. Bringel, C.J. Marx, Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post--transfer refinement, ELife 3 (2014) e04279, , http://dx.doi.org/10.7554/eLife.04279.
[22] S. Roselli, T. Nadalig, S. Vuilleumier, F. Bringel, The 380 kb pCMU01 plasmid encodes chloromethane utilization genes and redundant genes for vitamin B_{12}- and tetrahydrofolate-dependent chloromethane metabolism in *Methylobacterium extorquens* CM4: a proteomic and bioinformatics study, PLoS One 8 (2013) e56598, , http://dx.doi.org/10.1371/journal.pone.0056598.
[23] J.A. Vizcaíno, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.J. Kraus, J.P. Albar, S. Martinez-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination, Nat. Biotechnol. 32 (2014) 223–226, http://dx.doi.org/10.1038/nbt.2839.
[24] C. Carapito, A. Burel, P. Guterl, A. Walter, F. Varrier, F. Bertile, A. Van Dorsselaer, MSDA, a proteomics software suite for in-depth mass spectrometry data analysis using grid computing, Proteomics 14 (2014) 1014–1019, http://dx.doi.org/10.1002/pmic.201300415.
[25] J. Cox, M.Y. Hein, C.A. Luber, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, Mol. Cell. Proteomics MCP 13 (2014) 2513–2526, http://dx.doi.org/10.1074/mcp.M113.031591.
[26] C. Carapito, L. Lane, M. Benama, A. Opsomer, E. Mouton-Barbosa, L. Garrigues, A. Gonzalez de Peredo, A. Burel, C. Bruley, A. Gateau, D. Bouyssié, M. Jaquinod, S. Cianferani, O. Burlet-Schiltz, A. Van Dorsselaer, J. Garin, Y. Vandenbrouck, Computational and mass-spectrometry-based workflow for the discovery and validation of missing human proteins: application to chromosomes 2 and 14, J. Proteome Res. 14 (2015) 3621–3634, http://dx.doi.org/10.1021/pr5010345.
[27] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (2017) D200–D203, http://dx.doi.org/10.1093/nar/gkw1129.
[28] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods 8 (2011) 785–786, http://dx.doi.org/10.1038/nmeth.1701.
[29] J.D. Bendtsen, H. Nielsen, D. Widdick, T. Palmer, S. Brunak, Prediction of twin-arginine signal peptides, BMC Bioinforma. 6 (2005) 167, http://dx.doi.org/10.1186/1471-2105-6-167.
[30] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.
[31] C. You, H. Lu, A. Sekowska, G. Fang, Y. Wang, A.-M. Gilles, A. Danchin, The two authentic methionine aminopeptidase genes are differentially expressed in *Bacillus subtilis*, BMC Microbiol. 5 (2005) 57, http://dx.doi.org/10.1186/1471-2180-5-57.
[32] T. Ouidir, F. Jarnier, P. Cosette, T. Jouenne, J. Hardouin, Characterization of N-terminal protein modifications in *Pseudomonas aeruginosa* PA14, J. Proteome. 114 (2015) 214–225, http://dx.doi.org/10.1016/j.jprot.2014.11.006.
[33] F. Frottin, A. Martinez, P. Peynot, S. Mitra, R.C. Holz, C. Giglione, T. Meinnel, The proteomics of N-terminal methionine cleavage, Mol. Cell. Proteomics 5 (2006) 2336–2349.
[34] S.H. Payne, S. Bonissone, S. Wu, R.N. Brown, D.N. Ivankov, D. Frishman, L. Paša-Tolić, R.D. Smith, P.A. Pevzner, Unexpected diversity of signal peptides in prokaryotes, MBio 3 (2012) e00339-12, , http://dx.doi.org/10.1128/mBio.00339-12.
[35] P.M. Goodwin, C. Anthony, The biosynthesis of periplasmic electron transport proteins in methylotrophic bacteria, Microbiol. Read. Engl. 141 (Pt 5) (1995) 1051–1064, http://dx.doi.org/10.1099/13500872-141-5-1051.
[36] D.J. Anderson, C.J. Morris, D.N. Nunn, C. Anthony, M.E. Lidstrom, Nucleotide sequence of the *Methylobacterium extorquens* AM1 *moxF* and *moxJ* genes involved in methanol oxidation, Gene 90 (1990) 173–176, http://dx.doi.org/10.1016/0378-1119(90)90457-3.
[37] D.N. Nunn, C. Anthony, The nucleotide sequence and deduced amino acid sequence of the cytochrome *cL* gene of *Methylobacterium extorquens* AM1, a novel class of c-type cytochrome, Biochem. J. 256 (1988) 673–676.
[38] N.M. Good, H.N. Vu, C.J. Suriano, G.A. Subuyuj, E. Skovran, N.C. Martinez-Gomez, Pyrroloquinoline quinone ethanol dehydrogenase in *Methylobacterium extorquens* AM1 extends lanthanide-dependent metabolism to multicarbon substrates, J. Bacteriol. 198 (2016) 3109–3118, http://dx.doi.org/10.1128/JB.00478-16.
[39] D. Kumar, A.K. Mondal, A.K. Yadav, D. Dash, Discovery of rare protein-coding genes in model methylotroph *Methylobacterium extorquens* AM1, Proteomics 14 (2014) 2790–2794, http://dx.doi.org/10.1002/pmic.201400153.
[40] W.J. Craigen, C.T. Caskey, Expression of peptide chain release factor 2 requires high-efficiency frameshift, Nature 322 (1986) 273–275, http://dx.doi.org/10.1038/322273a0.
[41] S.D. La Roche, T. Leisinger, Identification of *dcmR*, the regulatory gene governing expression of dichloromethane dehalogenase in *Methylobacterium sp.* strain DM4, J. Bacteriol. 173 (1991) 6714–6721.
[42] D. Gisi, T. Leisinger, S. Vuilleumier, Enzyme-mediated dichloromethane toxicity and mutagenicity of bacterial and mammalian dichloromethane-active glutathione S-transferases, Arch. Toxicol. 73 (1999) 71–79.
[43] C. Lee, C. Park, Bacterial responses to glyoxal and methylglyoxal: reactive electrophilic species, Int. J. Mol. Sci. 18 (2017), http://dx.doi.org/10.3390/ijms18010169.
[44] F. Bringel, S. Vuilleumier, Metabolic regulation: a master role for ribulose-1,5-bisphosphate in one-carbon assimilation, Curr. Biol. 27 (2017) R1127–R1129, http://dx.doi.org/10.1016/j.cub.2017.09.003.
[45] B. Hu, M.E. Lidstrom, Metabolic engineering of *Methylobacterium extorquens* AM1 for 1-butanol production, Biotechnol. Biofuels 7 (2014), http://dx.doi.org/10.1186/s13068-014-0156-0.
[46] J. Timmins, H.-K.S. Leiros, G. Leonard, I. Leiros, S. McSweeney, Crystal structure of maltooligosyltrehalose trehalohydrolase from *Deinococcus radiodurans* in complex with disaccharides, J. Mol. Biol. 347 (2005) 949–963, http://dx.doi.org/10.1016/j.jmb.2005.02.011.
[47] W. Zhang, F. Li, L. Nie, Integrating multiple 'omics' analysis for microbial biology: application and methodologies, Microbiology 156 (2010) 287–301, http://dx.doi.org/10.1099/mic.0.034793-0.
[48] P. Natale, T. Brüser, A.J.M. Driessen, Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms, Biochim. Biophys. Acta BBA - Biomembr. 1778 (2008) 1735–1756, http://dx.doi.org/10.1016/j.bbamem.2007.07.015.
[49] P.A. Lee, D. Tullman-Ercek, G. Georgiou, The bacterial twin-arginine translocation pathway, Annu. Rev. Microbiol. 60 (2006) 373–395, http://dx.doi.org/10.1146/annurev.micro.60.080805.142212.

3.3.2 Détection des extrémités NH_2 terminale par l'approche dN-TOP

La comparaison des données expérimentales de sites d'initiations de la traduction avec celles annotées dans le génome permet de valider ou corriger l'annotation du site d'initiation de la traduction. Cette méthode est appelée protéogénomique (Nesvizhskii, 2014). L'approche globale dN-TOP (Bertaccini *et al.*, 2013) permet de déterminer l'acide aminé présent au niveau de l'extrémité NH_2 terminale des protéines (Fig. 3). Celui-ci est également appelé N-ter et peut correspondre au site d'initiation de la traduction ou peut aussi provenir du clivage d'un peptide signal. Le TMPP (trimethoxyphenyl phosphonium) réagit avec le groupement NH_2 à l'extrémité des protéines. Une fois la réaction réalisée, les protéines sont clivées par une protéase (trypsine) puis séparées par chromatographie liquide à nano-débit et enfin analysées par spectrométrie de masse. Les peptides sont séparés par chromatographie pour éviter que les signaux de différents peptides se superposent. Lors de la réaction d'ajout du TMPP, une forme légère (avec du ^{12}C) et une forme lourde (avec du ^{13}C) sont utilisées. La détection d'un N-ter n'est validée que si les formes légères et lourdes sont détectées par spectrométrie de masse. Cette approche possède certaines limites et ne permet pas de déterminer le N-ter pour toutes les protéines. Une modification du groupement NH_2 terminal telle qu'une acétylation peut empêcher la fixation du TMPP (Cain *et al.*, 2014). Certains peptides marqués au TMPP ne sont pas détectés car la séparation par chromatographie liquide ne permet pas de les séparer en raison de leurs tailles ou leur hydrophobicité. La taille des peptides dépend de la localisation du site de clivage de la trypsine.

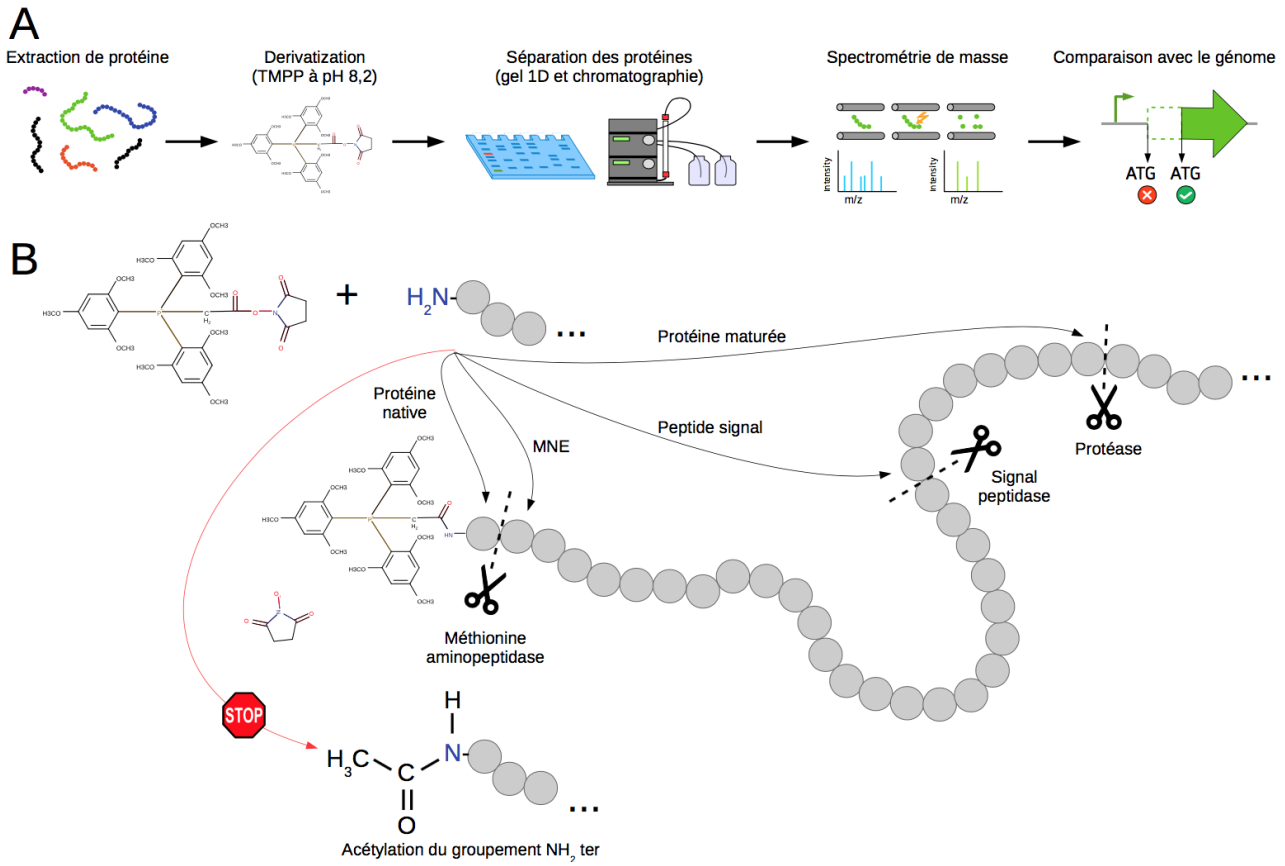


Figure 3.3: Principe de la technique dN-TOP pour déterminer l'extrémité NH_2 terminale d'une protéine

A Schéma du protocole de détection du groupement NH_2 terminale par dN-TOP. B. Réaction du TMPP avec le groupement NH_2 terminal d'une protéine. L'extrémité peut provenir d'une maturation telle qu'un clivage par une protéase. Le premier acide-aminé incorporé est toujours une méthionine, mais celle-ci peut être clivée par une méthionine aminopeptidase selon un procédé appelé NME (N-*terminal methionine excision*).

3.4 Détection des peptides signaux par protéomique

L'extrémité NH_2 terminale des protéines est sujette à des modifications *in vivo*. Des modifications post-traductionnelles peuvent intervenir telles que des acétylations ou une protéolyse (Cain *et al.*, 2014). Cette protéolyse peut notamment correspondre au peptide signal qui sert de signal d'adressage vers le périplasme par les systèmes de transport Sec et TAT (Fig. 2 ; (Saraogi and Shan, 2014)). La détection de peptide signal abordée dans l'article (Bibi-Triki *et al.*, 2018) est plus détaillée dans le cadre de cette thèse.

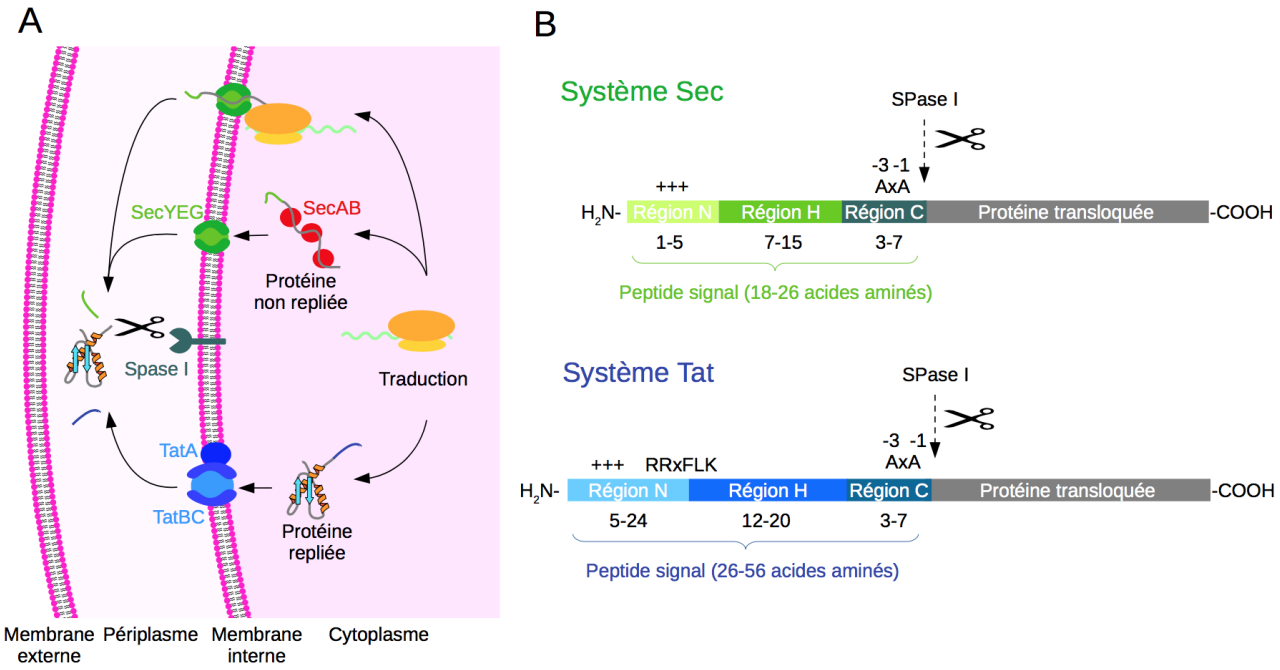


Figure 3.4: Description des peptides signaux Sec et TAT.
A. Mécanismes de translocation de protéines portant un peptide signal Sec ou TAT. Ces deux mécanismes reposent sur l'adressage des protéines vers le périplasme par des peptides signaux (Saraogi & Shan, 2014). Le système Sec permet la translocation des protéines au cours de la synthèse ou une fois qu'elles ont été synthétisées. B. Caractéristiques des peptides signaux pour les systèmes de translocation Sec et TAT. Quel que soit le système, la composition du peptide signal est conservée (une séquence chargée positivement suivie d'une séquence hydrophobe et enfin une séquence neutre mais polaire) alors que le nombre de résidus de chaque domaine varie. Le motif peptidique consensus reconnu par les Signal Peptidases (SPase I) est AxA. Les peptides signaux TAT possèdent en plus un motif de reconnaissance RRxFL du côté N terminal.

À partir des données de spectrométrie de masse obtenues, des filtres bio-informatiques sont réalisés pour valider les résultats. Pour ce faire une banque de protéines leurres correspondant à celles du génome mais lues dans le sens inverse a été réalisée. Puis les résultats expérimentaux sont comparés avec ces leurres de façon à éliminer les données artefactuelles.

Pour identifier les N-ter résultants d'un clivage par une protéase clivant un peptide signal (Saraogi & Shan, 2014), deux logiciels sont utilisés : SignalP (version 4.1 ; (Petersen *et al.*, 2011)) et TatP (version 1.0 ; (Bendtsen *et al.*, 2005).). Pour cela, SignalP calcule la probabilité de présence d'un peptide signal à partir de la séquence des 70 premiers acides aminés de la protéine. Il va notamment rechercher la présence d'un site de clivage reconnu par une protéase (signal peptidase I) ainsi que la présence d'une séquence hydrophobe, en tenant compte également de la position de ce site dans la séquence : en effet, la taille moyenne des peptides signaux chez les bactéries à paroi Gram négatif

est de 24 acides aminés. Le logiciel TatP (version 1.0) est également utilisé pour la recherche de peptides signaux, il a été développé par la même équipe que le logiciel SignalP (Bendtsen *et al.*, 2005). En plus du motif de clivage, le logiciel TatP identifie le motif reconnu par le système TAT (RRxFLK) qui se trouve entre les régions N et H du peptide signal. Ce logiciel se base sur les différences entre les peptides signaux Sec et TAT (Fig. 4). Notamment une taille plus grande pour les peptides TAT et une région H moins hydrophobe. Pour compléter la recherche de peptides signaux et caractériser les gènes correspondant d'autres logiciels ont été utilisés. Le logiciel PsortB (version 3.0) permet de prédire la localisation cellulaire par comparaison avec celle d'autres protéines, la présence de domaine transmembranaire ou de motifs (Yu *et al.*, 2010). La présence de séquence Shine-Dalgarno est recherchée avec le logiciel RBS *calculator* (version 2.0 ; (Espah Borujeni *et al.*, 2014)) et complétée par l'analyse des fonctions prédites lors de l'annotation du génome (Vuilleumier *et al.*, 2009) et par des études bibliographiques.

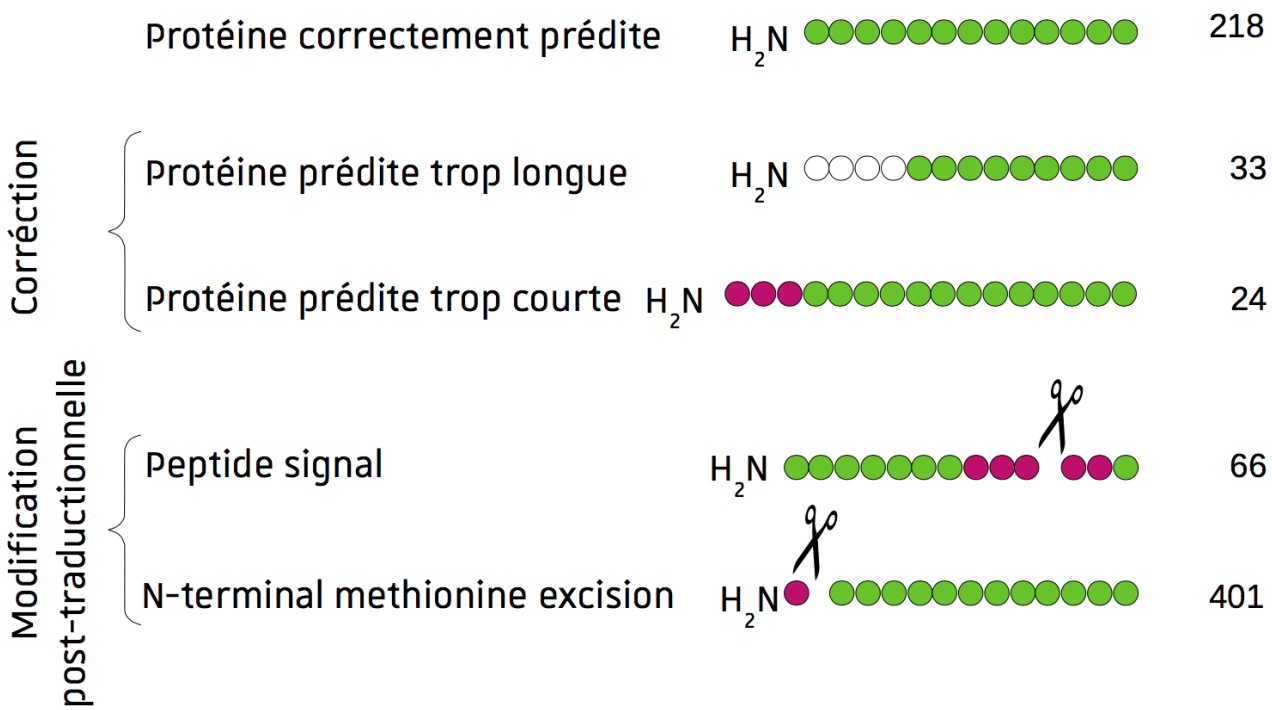


Figure 3.5: Bilan de l'approche dN-TOP.
La détection du N-ter a permis de valider le codon d'initiation de plus de 200 gènes et de modifier celui de 57 gènes trop courts ou trop longs, ainsi que de mettre en évidence des modifications post-traductionnelles.

L'approche dN-TOP a permis d'identifier 2 809 peptides présents au niveau N-ter (Fig. 5). Ils ont pu être répartis en plusieurs catégories décrites ci-dessous. Les autres correspondent à des produits de dégradation.

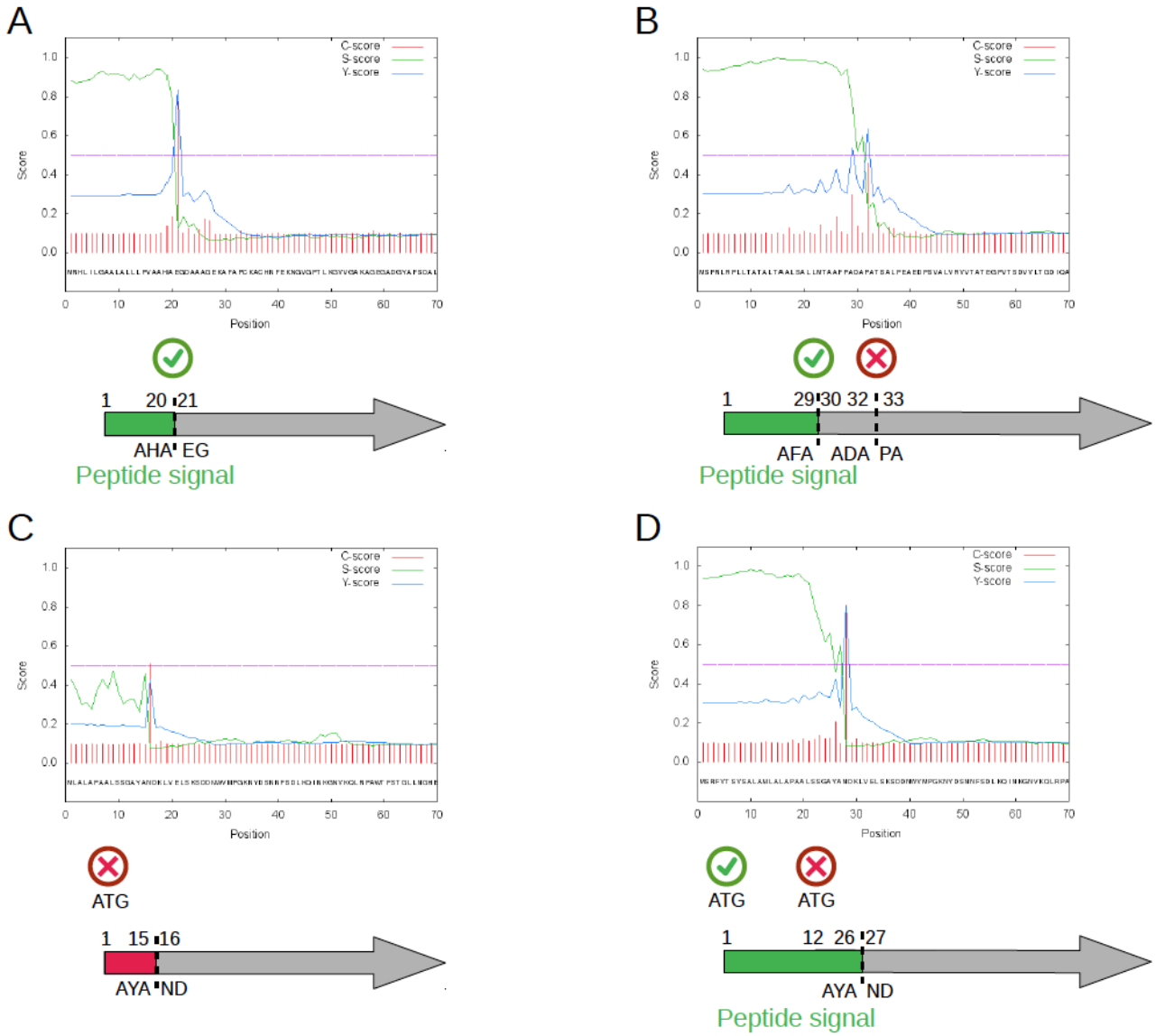


Figure 3.6: Détection de peptides signaux via SignalP et correction du codon d'initiation de la traduction.

Captures d'écrans du logiciel SignalP. L'axe des abscisses correspond à la séquence de la protéine (70 premiers acides aminés) tandis que l'axe des ordonnées indique la valeur du score obtenu pour les variables C, S et Y. Lorsque le score C (ligne verticale rouge) est supérieur à 0,5 (ligne horizontale violette), le peptide signal est validé. A. Bonne prédiction du site de clivage. B. Prédiction de deux sites de clivage dont un seul est confirmé expérimentalement. C. Mauvaise prédiction du peptide signal de *mxaF* à cause d'une mauvaise annotation D. Prédiction du peptide signal de *mxaF* lorsque l'annotation du gène a été corrigée.

L'approche dN-TOP a permis de détecter 66 peptides signaux parmi les 1 129 prédits au total pour l'ensemble du génome (logiciel SignalP 4.1). L'abondance des protéines ayant un peptide signal ne varie pas dans les conditions testées. Ces données ont permis de ré-annoter certains gènes. Par exemple dans le cas du gène *mxaF*, le logiciel SignalP a détecté un site de clivage potentiel mais l'a écarté car il se trouve trop tôt dans la séquence par rapport à ceux connus (Fig. 6D). Une analyse plus poussée a indiqué que le codon d'initiation de la traduction avait été mal annoté. Avec la nouvelle annotation, le logiciel détecte bien le peptide signal. Dans d'autres cas, le logiciel préditdeux sites de clivage mais n'en garde qu'un (Fig. 6B). Dans plusieurs cas, le site observé expérimentalement ne correspond pas à celui sélectionné par le logiciel mais à celui qui n'a pas été retenu. Les informations sur ces cas pourront servir à créer des banques de tests pour améliorer la détection de peptides signaux par de nouveaux logiciels.

Les protéines détectées, avec un peptide signal, possèdent des fonctions variées (Fig. 7) dont certaines sont impliquées dans le métabolisme du méthanol.

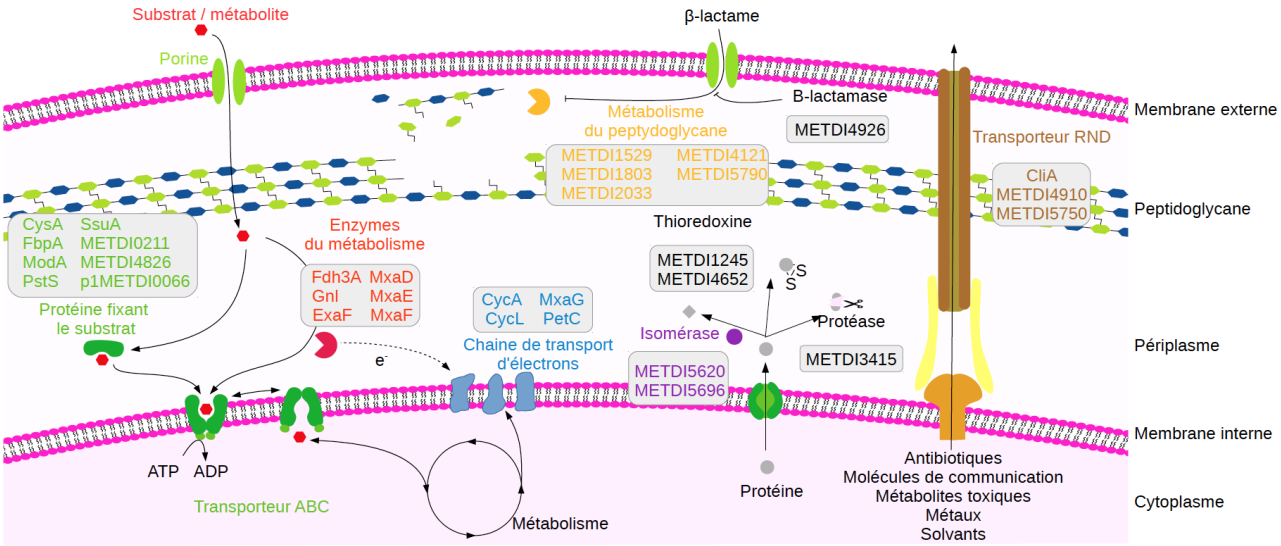


Figure 3.7: Catégories de protéines pour lesquelles un peptide signal a été déterminé expérimentalement.
Les enzymes ExaF, Fdh3A, MxaD, MxaE et MxaF font partie du métabolisme du méthylotrophe (oxydation du méthanol). Les protéines CycA, CycL et MxaG interviennent également dans le métabolisme méthylotrophe en transportant les électrons libérés par la méthanol déshydrogénase (MxaIF) lors de l'oxydation du méthanol.

3.5 Comparaison des résultats de transcriptomique et des protéomique

Les approches de transcriptomique et de protéomique ont fourni une vue d'ensemble sur la physiologie de *M. extorquens* DM4 en croissance avec du DCM ou du méthanol. L'analyse des résultats de quantification globale des ARNm et des protéines montre une corrélation positive pour six gènes en condition DCM et un gène en condition méthanol. Les protéines DcmA, DcmB et DcmC, une sous-unité de la formaldéhyde déshydrogénase 1, une protéine de fonction inconnue METDI4918 et une protéine périplasmique (METDI1959) associée à un transporteur ABC sont retrouvées comme étant plus abondantes en condition DCM comparée à la condition méthanol (Fig. 8). Ces deux approches détectent des molécules différentes, ont des contraintes et biais expérimentaux différents pouvant justifier des résultats différents. Cependant, des seuils différents ont été utilisés pour interpréter les résultats des deux techniques. Les ARN sont considérés comme plus abondants dans une condition de culture lorsque leur valeur de $\log_2$ fc est supérieure, ou égale, à 2 (soit quatre fois plus exprimé). Tandis qu'en protéomique, les protéines sont considérées comme plus abondantes lorsque leur valeur $\log_2$ fc est supérieure, ou égale, à 1,5 (soit 2,8 fois plus abondante). De plus, le ratio $\log_2$ fc entre les deux conditions de cultures ne peux pas être calculé si les protéines ne sont détectées que dans l'une des conditions, ce qui est le cas pour 138 et 266 protéines en croissance avec du DCM ou de méthanol respectivement. Par exemple, la protéine METDI3043 a été trouvée uniquement en présence de DCM mais n'a pas été détectée en présence de méthanol. En plus, des modifications post-transcriptionnelles peuvent influencer l'abondance des protéines (Zhang *et al.*, 2010). Deux gènes annotés comme *universal stress protein* (METDI0274 et METDI4304) possèdent des transcrits plus abondants en présence de méthanol mais des protéines plus abondantes lors de croissance avec du DCM. Cette différence laisse supposer un mécanisme de régulation post-transcriptionnelle telle qu'une modification de la synthèse protéique ou de la dégradation des protéines associées dans une seule condition de culture. Néanmoins le(s) mécanisme(s) reste(nt) encore inconnu(s).

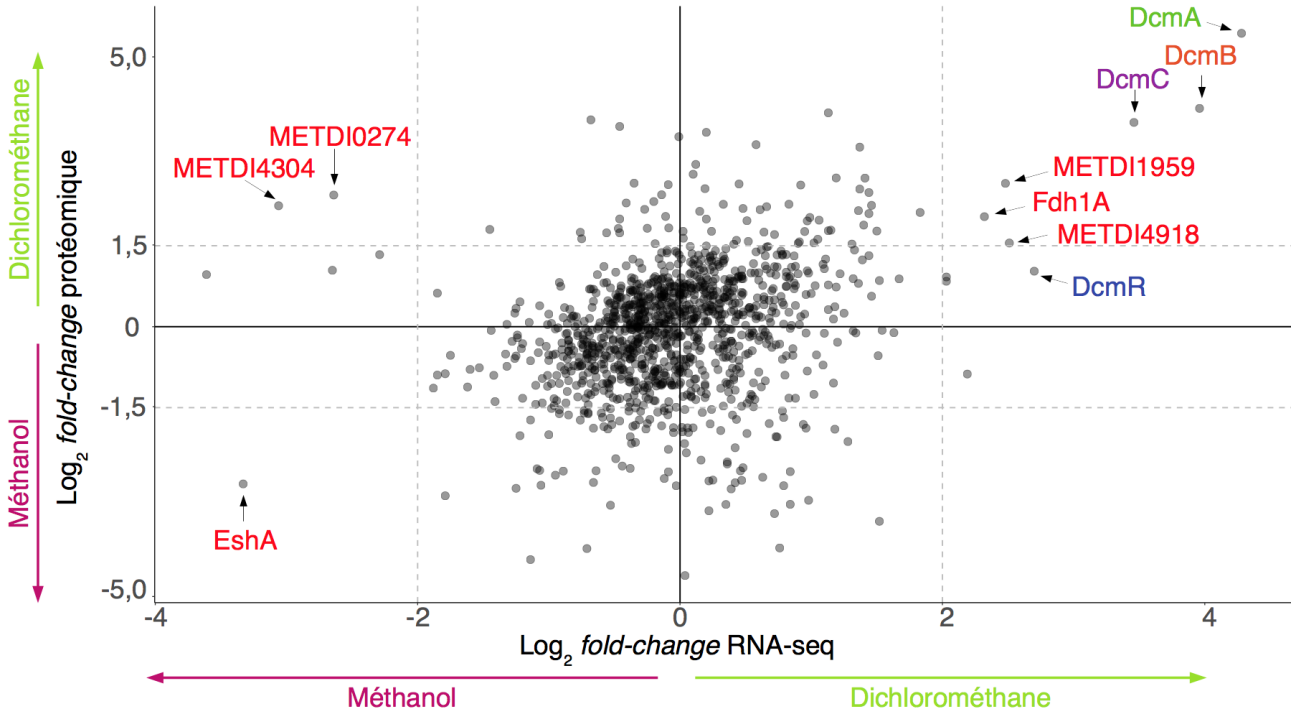


Figure 3.8: Comparaison de l'abondance relative des ARNm et protéines chez *M. extorquens* DM4 lors de culture en présence de DCM ou méthanol.
La majorité des ARNm et protéines n'ont pas de variation d'abondance entre les deux conditions testées. Dans le cas où des transcrits et protéines sont retrouvés ensemble plus abondant en même temps dans une même condition de culture, cela suggère un rôle fonctionnel dans cette condition. Le gène *dcmR* est indiqué à titre d'information même si la protéine correspondante n'a pas une valeur de $\log_2$ fc supérieure à 1,5.

L'ensemble des données obtenues au niveau transcriptionnel ou traductionnel, permet d'émettre des hypothèses quant aux mécanismes cellulaires impliqués dans la réponse au DCM. Les hypothèses décrites ici sont plus détaillées que celles proposées dans l'article (Bibi-Triki et al., 2018). Ces mécanismes sont décrits dans les paragraphes suivants :

1. Régulation de l'expression génique
2. Réparation de l'ADN
3. Métabolisme des aldéhydes
4. Métabolisme des glucides
5. Métabolisme du glutathion
6. Métabolisme énergétique

3.5.1 Régulation de l'expression génique

L'ensemble des 260 gènes annotés dans le génome de *M. extorquens* DM4 comme codant des facteurs de transcriptions ont été classifiés selon leur identité de séquence dans différentes catégories (fig. 9A). Parmi l'ensemble des facteurs intervenants dans la transcription et sa régulation, leur expression différentielle au niveau ARNm et protéine a été recherchée (fig. 9B et 9C). La modification d'abondance de certains ARNm, et donc probablement de la transcription, en présence de DCM semble impliquer des facteurs de transcription.

Les différences d'abondance des ARN ou des protéines peuvent s'expliquer par une modulation au niveau de l'expression génétique lors de la transcription. En plus des facteurs de transcription, d'autres protéines ayant un lien avec la régulation de la transcription, sont détectées comme différentiellement produites (Fig. 10). L'une des deux sous-unités de l'ADN topoisomérase IV est plus abondante en présence de méthanol tandis que l'autre sous-unité n'est pas détectée. L'ADN topoisomérase module le sur-enroulement de l'ADN ce qui a un effet sur la transcription (Dorman & Dorman, 2016).

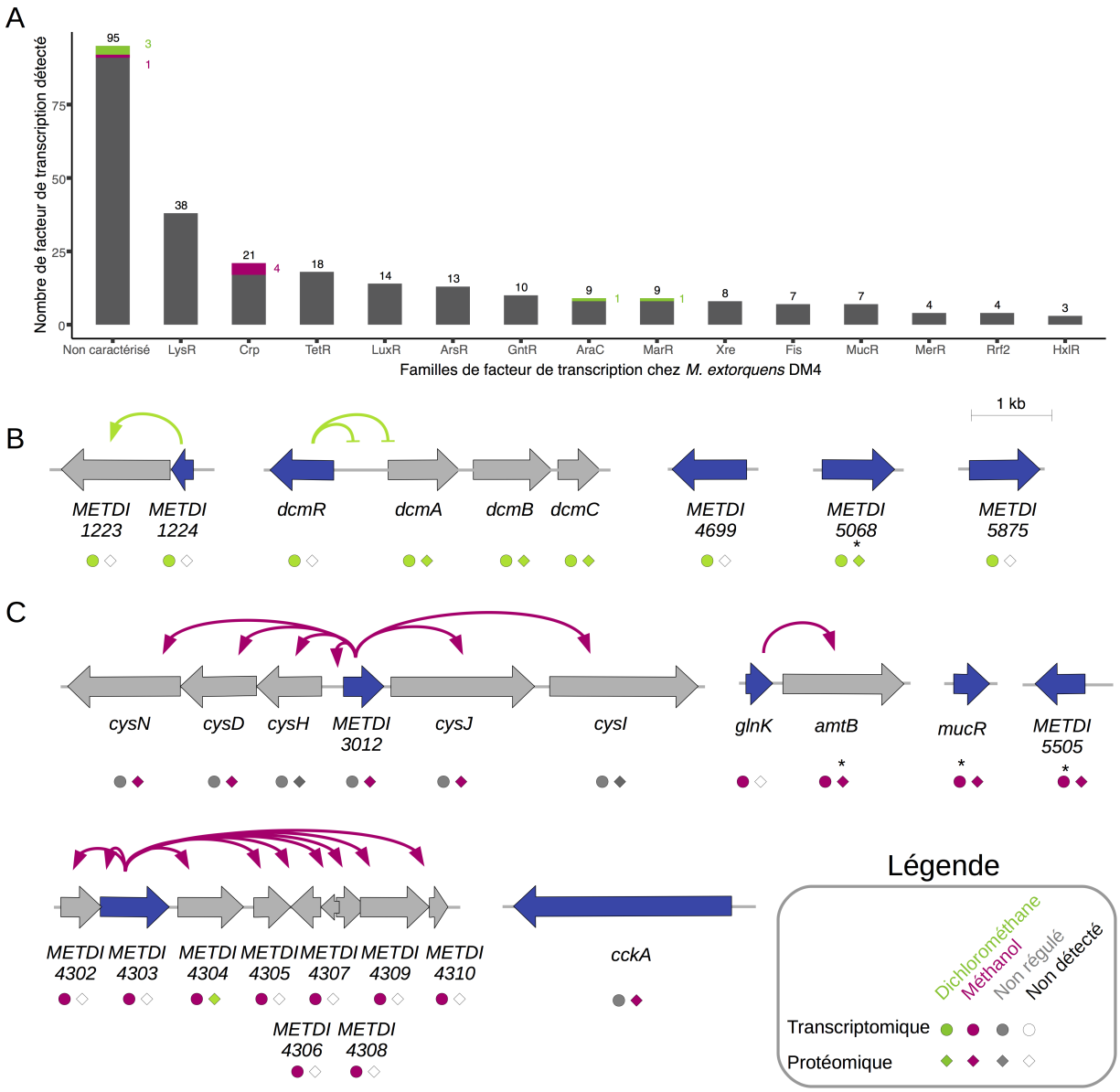


Figure 3.9: Facteur de transcription dont l'expression est régulée

A Facteurs de transcription prédits dans le génome de *M. extorquens* DM4. Les facteurs de transcriptions régulés en présence de DCM ou méthanol sont indiqués en vert ou rose, respectivement. Certains de ces facteurs se trouvent à proximité de gènes dont l'expression est aussi régulée. Ceux régulés en présence de DCM sont indiqués dans la fenêtre B tandis que ceux qui le sont en méthanol sont dans la fenêtre C. Les flèches indiquent les gènes potentiellement régulés par ces facteurs de transcription. Les étoiles indiquent que les données de quantification présentent une forte variation entre les réplicats et doivent être interprétés avec prudence.

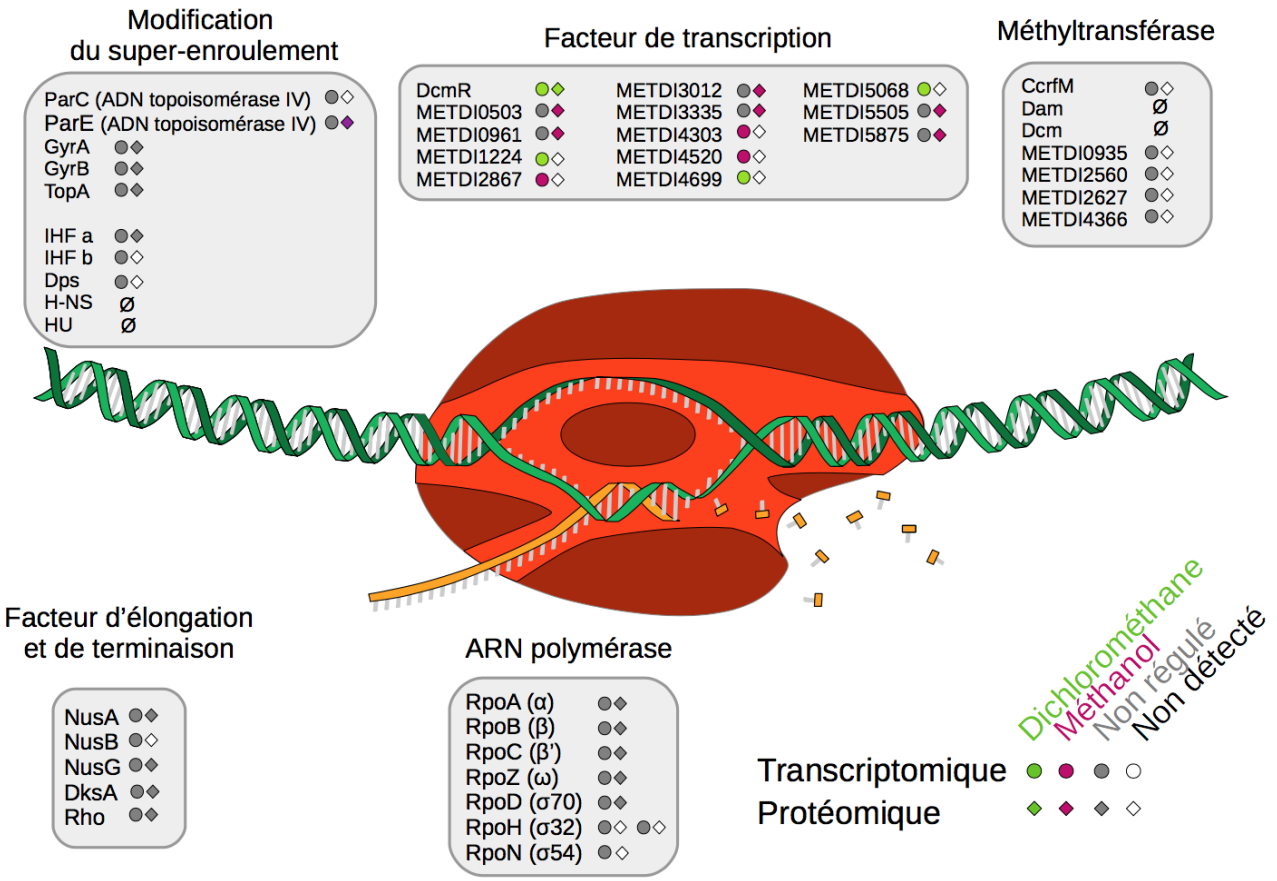


Figure 3.10: Mécanismes régulant la transcription chez *M. extorquens* DM4.
En plus des facteurs de transcription, l'expression génique peut être régulée par le sur-enroulement de l'ADN (Dorman & Dorman, 2016) et la méthylation de l'ADN (Adhikari & Curtis, 2016). Le symbole Ø indique que le gène n'a pas été trouvé.

3.5.2 Réparation de l'ADN suite à la déshalogénation

L'utilisation du DCM conduit à la formation de dommages à l'ADN lors de la déshalogénation du DCM (Kayser *et al.*, 2000). Ces dommages seraient dus à la synthèse d'un intermédiaire réactif le S-chlorométhylglutathion lors de la réaction du DCM avec l'enzyme DcmA (Kayser & Vuilleumier, 2001). Il était donc envisagé que les gènes de réparation de l'ADN soient plus exprimés durant la dégradation du DCM. L'ensemble des gènes annotés comme intervenant dans la réparation d'ADN chez *M. extorquens* DM4 a été répertorié (Fig. 11).

Chez les bactéries, la réparation de l'ADN suite à des alkylations fait intervenir généralement les protéines Ada, AlkA, AlkB, AidB ou des glycosylases (Mielecki & Grzesiuk, 2014). Un potentiel gène *ada* est présent dans le génome de *M. extorquens* DM4 (METDI5321 ; 52 % d'identité au

niveau protéique avec Ada de *E. coli*). Par contre *aidB* et l'opéron *alkAB* ne sont pas trouvés chez cette souche (Fig. 10). Néanmoins, d'autres gènes de réparation de l'ADN sont présents chez *M. extorquens* DM4 notamment des ADN polymérases dites translésionnelles. Ces ADN polymérases permettent la synthèse d'ADN malgré la présence d'adduits ou d'alkylations mais avec une fidélité plus faible que celle des ADN polymérase I et III (Goodman & Woodgate, 2013). Ces mécanismes de réparation translésionnelle interviennent chez *Pseudomonas* en réponse à l'alkylation de l'ADN (Jatsenko *et al.*, 2017). Le gène *dinB* ainsi que *umuC* et *umuD* sont trouvés dans le génome de *M. extorquens* DM4 et codent respectivement l'ADN polymérase IV et l'ADN polymérase V (Fig. 10). L'opéron *imuABC* qui code une ADN polymérase translésionnelle alternative intervenant dans la réparation de l'ADN (Alves *et al.*, 2017) a également été détecté. En addition de ce système, de nombreux gènes impliqués dans la recombinaison de l'ADN sont également retrouvés dans le génome de *M. extorquens* DM4 ainsi que des glycosylases capables de reconnaître et d'exciser des nucléotides endommagés (Lee & Wallace, 2017). La voie de réparation de l'ADN en réponse aux mésappariements est également détectée mais semble incomplète comparée à celle décrite chez *E. coli*. En effet, le gène *mutH*, codant une endonucléase recrutée par MutL au niveau du brin avec une erreur d'appariement, n'est pas trouvé chez *M. extorquens* DM4. Contrairement à ce qui est décrit chez *E. coli*, chez de nombreux procaryotes, la protéine MutL possède une activité d'endonucléase (similaire à celle de MutH) en plus de son rôle dans le recrutement de protéines de réparation (Putnam, 2016), de sorte que cette voie pourrait tout de même être opérationnelle chez *M. extorquens* DM4 malgré l'absence de *mutH*.

Les données de transcriptomique et de protéomique confirment ainsi que certains de ces gènes de réparation de l'ADN sont exprimés dans les conditions de culture testées (Fig. 10). Néanmoins, aucun ARNm ou protéine n'est détecté comme plus abondant en présence de DCM. L'expression de certains gènes de réparation de l'ADN, présents dans un régulon appelé système SOS, est régulée par le facteur de transcription LexA qui réprime leur transcription en absence de stress (Baharoglu & Mazel, 2014). L'interaction de la protéine LexA avec RecA, qui se fixent sur l'ADN simple brin, conduit à l'autolyse de ce facteur de transcription. L'abondance de la protéine LexA ne varie pas entre la condition de culture DCM et méthanol ce qui laisse supposer que le système SOS n'est pas induit lors de l'alkylation de l'ADN par le S-chlorométhylglutathion. Les seules protéines ayant une abondance qui varie entre les deux conditions sont MutS qui intervient dans la réparation des mésappariements et la protéine Mfd impliquée dans la réparation liée à la transcription (Monnet *et al.*, 2013). La protéine Mfd est capable de détecter un arrêt de l'ARN polymérase suite à un

dommage à l’ADN et de recruter la voie de réparation par excision de nucléotide (UvrABCD). Cependant, aucun gène de réparation de l’ADN n’a été identifié comme étant plus abondant en présence de DCM malgré la présence d’alkylation dans cette condition de culture laissant supposer la présence d’autres mécanismes pour faire face aux intermédiaires toxiques du DCM (Mielecki & Grzesiuk, 2014).

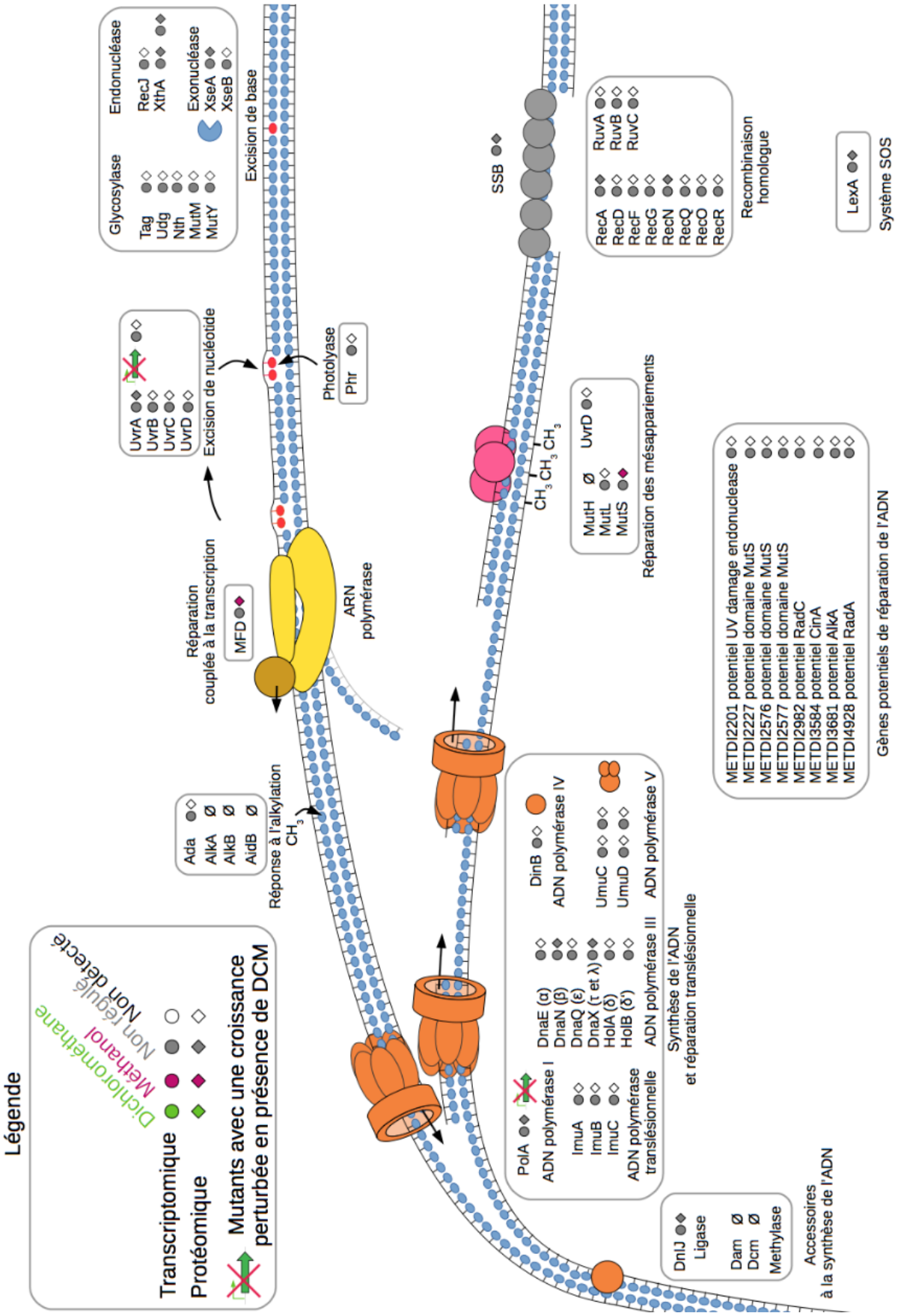

Légende
Mutants avec une croissance perturbée en présence de DCM
Protéomique
Transcriptomique
Non détecté
Non requis
Méthanol
Dichlorométhane
Glycosylase
Tag
Udg
Nth
MutM
MutY
Endonucléase
RecJ
XthA
Exonucléase
XseA
XseB
Excision de base
UvrA
UvrB
UvrC
UvrD
Excision de nucléotide
Photolyase
Phr
Réparation couplée à la transcription
MFD
ARN polymérase
Ada
AlkA
AlkB
AidB
Réponse à l'alkylation
SSB
RecA
RecD
RecF
RecG
RecN
RecQ
RecO
RecR
RuvA
RuvB
RuvC
Recombinaison homologue
MutH
MutL
MutS
UvrD
Réparation des mésappariements
METDI2201 potentiel UV damage endonuclease
METDI2227 potentiel domaine MutS
METDI2576 potentiel domaine MutS
METDI2577 potentiel domaine MutS
METDI2982 potentiel RadC
METDI3584 potentiel CinA
METDI3681 potentiel AlkA
METDI4928 potentiel RadA
Gènes potentiels de réparation de l'ADN
LexA
Système SOS
PolA
ADN polymérase I
DnaE (α)
DnaN (β)
DnaQ (ε)
DnaX (τ et λ)
HolA (δ)
HolB (δ')
ADN polymérase III
DinB
ADN polymérase IV
ImuA
ImuB
ImuC
ADN polymérase translésionnelle
UmuC
UmuD
ADN polymérase V
Synthèse de l'ADN et réparation translésionnelle
DnlJ
Ligase
Dam
Dcm
Methylase
Accessoires à la synthèse de l'ADN

Figure ci-contre

Figure 3.11: Mécanismes de réparation de l'ADN chez *M. extorquens* DM4.
Les alkylations de l’ADN sont réparées par différents systèmes ; l’excision de base (glycosylase), l’excision de nucléotide (UvrABCD), le système Ada ou encore la synthèse translésionnelle. Aucun gène intervenant dans ces voies n’est différentiellement transcrit ou traduit en présence de DCM. Le symbole Ø indique que le gène n’a pas été trouvé.

3.5.3 Métabolisme des aldéhydes et plus particulièrement du formaldéhyde

Le formaldéhyde produit lors de la dégradation du DCM peut être oxydé en formiate par la voie de la tétrahydrométhanoptérine (H_4MPT) (Ochsner *et al.*, 2015). Néanmoins, plusieurs enzymes de la famille xanthine oxydase / aldéhyde oxydoréductase sont également capables de réaliser cette oxydation. Parmi les protéines les plus abondantes en condition DCM (ou retrouvées uniquement dans cette condition), certaines appartiennent à la famille des xanthine oxydases / aldéhyde oxydoréductases : METDI0092, METDI0973-0975, METDI3582 et METDI4666 (Fig. 12). Elles présentent une forte similarité de séquence avec les enzymes PaoABC d’*E. coli* (Correia *et al.*, 2016) et Mb-AOX de *Methylobacillus* sp. KY440 (Uchida *et al.*, 2018). Ces enzymes sont composées de trois sous-unités portant respectivement des centres fer-soufre, un cofacteur avec du molybdène et un cofacteur FAD (Wang *et al.*, 2016). Elles peuvent oxyder, par l’utilisation d’oxygène, des aldéhydes aliphatiques tels que le formaldéhyde, l’acétaldéhyde ou des aldéhydes aromatiques telle que la xanthine (Uchida *et al.*, 2018). Une étude de protéomique chez *Starkeya novella* (*alpha-proteobacteria*) a montré que la production d’aldéhyde oxydase varie selon la source de carbone utilisée (Kappler & Nouwens, 2013). Il peut être supposé que ces enzymes permettent de détoxifier le formaldéhyde intracellulaire produit lors du métabolisme du DCM (tandis qu’en présence de méthanol, il est produit dans le périplasme), le formaldéhyde étant un composé toxique pour la cellule (Kawanishi *et al.*, 2014 ; Chen *et al.*, 2016). Cela pourrait expliquer aussi pourquoi l’une des sous-unités de la formaldéhyde déshydrogénase 1 est retrouvée plus abondante à la fois au niveau ARNm et protéine (Chaignaud *et al.*, 2017 ; Bibi-Triki *et al.*, 2018).

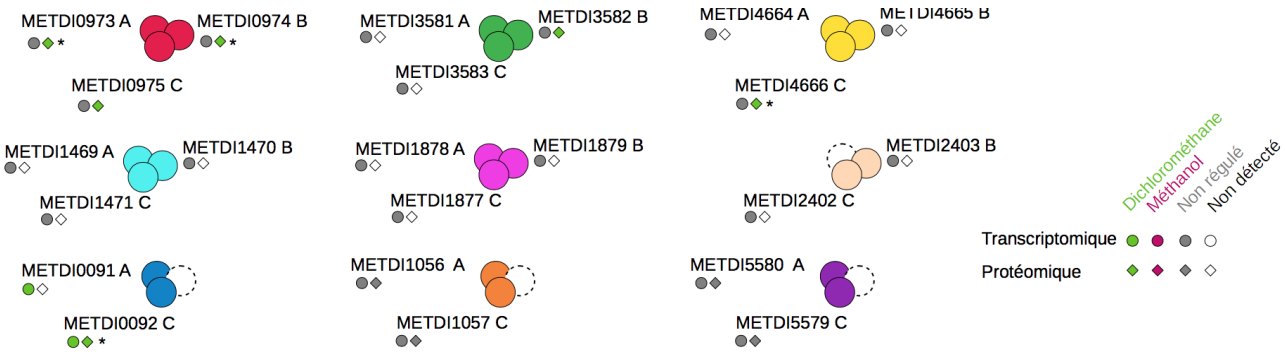


Figure 3.12: Expression différentielle des protéines de la famille des aldéhydes oxygénases chez *M. extorquens* DM4.
L'étoile (*) indique que la protéine a été seulement détectée de façon significative en présence de DCM de sorte qu'il n'est pas possible de donner une valeur de $\log_2$ fc. On admet que la protéine est plus abondante en présence de DCM. Les pointillés indiquent une sous-unité manquante.

3.5.4 Modification du métabolisme des glucides, en réponse à la croissance avec le DCM

L'utilisation du DCM comme celle du méthanol repose sur le métabolisme méthylotrophe de sorte que l'on pouvait supposer que le métabolisme central de *M. extorquens* DM4 ne soit pas modifié de façon importante entre ces deux conditions de culture. Effectivement, les données de transcriptomique et protéomique indiquent que seulement deux transcrits et 13 protéines impliquées dans le métabolisme des glucides sont plus abondants en présence de DCM (Fig. 12). Ces différences d'abondance d'enzymes pourraient servir à fournir plus d'énergie par le métabolisme glucidique en présence de DCM. Une autre hypothèse reposerait sur la production de métabolites secondaires qui auraient un rôle dans l'adaptation au DCM.

L'approche de RNA-seq a mis en évidence trois gènes impliqués dans le métabolisme glucidique comme étant davantage exprimés en présence de DCM (Chaignaud *et al.*, 2017). Des protéines ont également été détectées comme plus abondantes en présence de DCM (Bibi-Triki *et al.*, 2018), comme Gnd intervenant dans la réduction du NADP(+) en NADPH fournissant de l'énergie sous forme de pouvoir réducteur. Au contraire, les gènes correspondants n'étant pas sur-exprimés au niveau transcriptionnel, la régulation induisant une sur-abondance de ces protéines pourrait avoir lieu notamment au niveau post-transcriptionnel.

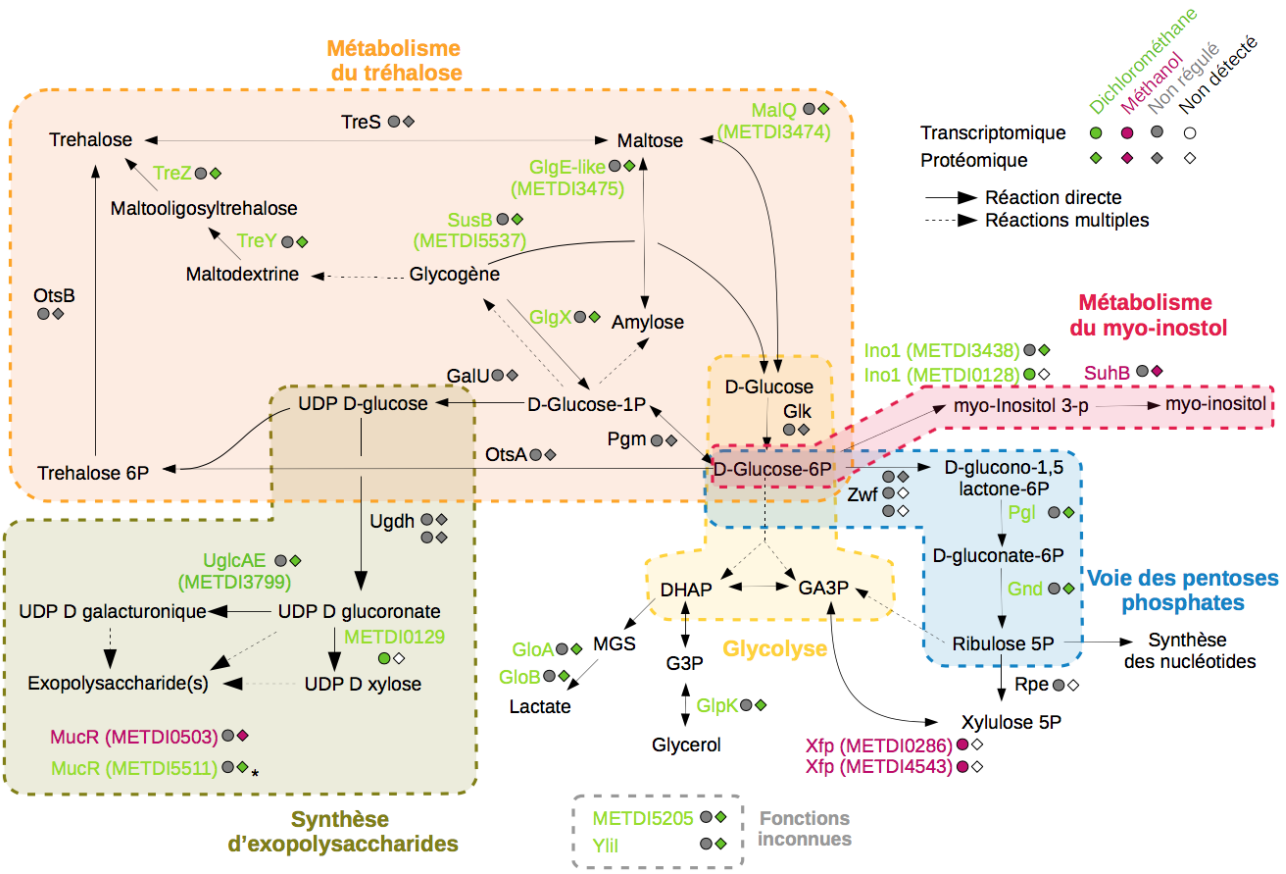


Figure 3.13: Prédiction des voies métaboliques liées aux glucides et plus exprimées en présence de DCM.

Certaines enzymes catalysant ces réactions sont retrouvées comme étant codées par plusieurs gènes dans le génome (illustrés par différents symboles sur ce schéma).

M. extorquens DM4 possède les gènes des trois voies connues de synthèse du tréhalose qui est connu pour être un osmoprotectant et pour être produit lors de différents types de stress (*ostAB*, *treS*, *treYZ* ; (Ruhal *et al.*, 2013)). Les protéines de l'une de ces voies (TreY et TreZ) ont été détectées comme plus abondantes en condition DCM (Fig 13 ; (Bibi-Triki *et al.*, 2018)). Le tréhalose est une molécule non chargée très soluble et peu toxique à forte concentration qui permet de maintenir la pression osmotique et de stabiliser les protéines. Cela pourrait être un mécanisme d'adaptation au stress osmotique dû à la libération dans le cytoplasme de chlorure lors du métabolisme du DCM. Une approche de métabolomique pourrait confirmer un enrichissement en tréhalose lors de croissance en présence de DCM.

D'autres part, des enzymes intervenant dans la dégradation du glycogène ont été détectées comme plus abondantes en présence de DCM (Bibi-Triki *et al.*, 2018). Elles interviennent respectivement dans la synthèse du tréhalose par TreYZ, du glucose 1-phosphate par GlgX ou du glucose par SusB. Ceci pourrait permettre de récupérer du carbone en condition de stress par la dégradation du glycogène. Par ailleurs, des études ont déjà établi un lien entre la dégradation du glycogène par GlgX et la formation de tréhalose par TreYZ chez *Corynebacterium glutamicum* (Seibold & Eikmanns, 2007).

Le myo-inostiol-3-phosphate semble également être un métabolite davantage produit lors de cultures en présence de DCM (Fig. 13). En effet, l'enzyme Ino1 qui produit du myo-inositol-3-phosphate est codée par deux gènes chez *M. extorquens* DM4 : METDI0128 et METDI3438. Les transcrits de l'un de ces deux gènes (METDI0128) sont plus abondants en présence de DCM tandis que la protéine codée par l'autre gène (METDI3438) est plus abondante dans la même condition de culture. L'enzyme SuhB qui dégrade cette molécule en myo-inositol est plus abondante en présence de méthanol, laissant supposer une inhibition lors de cultures avec du DCM ou une activation en présence de méthanol. Une étude de l'expression dans une troisième condition de culture, telle que le succinate permettrait de déterminer s'il s'agit d'une répression ou d'une activation. Le myo-inositol-3-phosphate est courant chez les eucaryotes où il intervient dans des mécanismes de signalisation cellulaire, tandis que chez les procaryotes, il a un rôle d'osmoprotectant et intervient dans la modification de lipides membranaires (Roberts, 2006). Ces deux cas de figure pourraient intervenir dans la réponse au DCM chez *M. extorquens* DM4.

Une étude récente semble indiquer que le gène METDI5511 coderait un facteur de transcription impliqué dans la réponse au DCM et le métabolisme des exopolysaccharides (Firsova *et al.*, 2017). Cette protéine qui a été détectée comme plus produite en présence de DCM par électrophorèse à deux dimensions (Muller *et al*., 2011) possède une forte similarité de séquence avec la protéine Ros/MucR impliquée dans la régulation de la formation d'exopolysaccharides (Schmid *et al.*, 2015). La délétion de ce gène diminue la production d'exopolysaccharides chez *M. extorquens* DM4 lors de culture avec du DCM (Firsova *et al.*, 2017). Le gène METDI0129 et la protéine METDI3799, tous deux impliqués dans la production d'exopolysaccharides ont été détectés plus abondants en présence de DCM (Chaignaud *et al*., 2017 ; Bibi-Triki *et al*., 2018). Une modification de la paroi bactérienne permettrait de limiter la diffusion du DCM dans la cellule et donc protéger la cellule de l'effet solvant de ce polluant.

3.5.5 Métabolisme du glutathion, le cofacteur de l’enzyme DCM déshalogénase

Les résultats des approches « -omiques » concordent avec des dosages biochimiques du glutathion indiquant que la concentration totale ne varie pas entre les conditions méthanol et DCM (Firsova *et al.*, 2004 ; Torgonskaya *et al.*, 2018). En effet, aucun des ARNm ou protéines de synthèse ou de dégradation du glutathion ne semble être plus exprimés (Fig. 14). De plus, la quantité de l’enzyme glutathion réductase n’est pas modifiée mais son activité augmente en présence de DCM (Torgonskaya *et al.*, 2018). Ceci laisse supposer qu’il y aurait une modification post-traductionnelle. Malgré une quantité de glutathion stable, sa forme oxydée (GS-SG) est plus abondante que sa forme réduite (GSH) et pourrait résulter d’un déficit en énergie nécessaire à sa réduction impliquant l’enzyme glutathion réductase.

Enfin, quant aux gènes codant un transporteur de type ABC, impliqué dans le transport du glutathion (Bachhawat *et al.*, 2013), les protéines correspondantes ne sont pas détectées : CydD (METDI2007) ; DppABCDF (METDI1865-1869) ; OppABC (METDI5273-5275). Par contre plusieurs enzymes utilisant le glutathion comme cofacteur voient leur quantité changer entre la culture en présence de DCM ou de méthanol. Par exemple, GloAB qui intervient dans la détoxification de composés toxiques produits lors du métabolisme des glucides. La variation d’expression de ces enzymes pourrait être un effet direct dû à la concentration en glutathion ou un effet indirect dû au stress provoqué par le DCM.

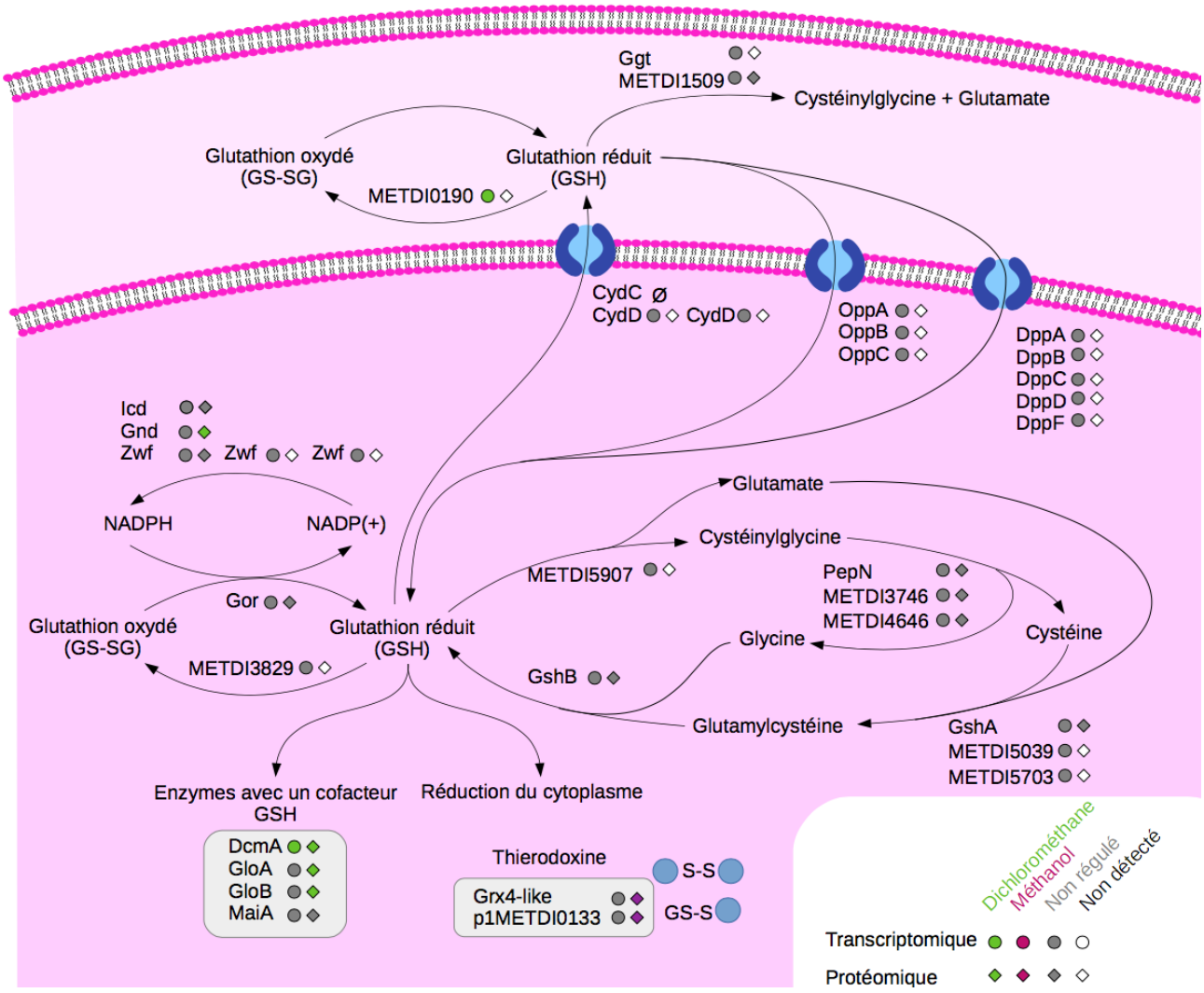


Figure 3.14: Métabolisme du glutathion chez *M. extorquens* DM4.
Le glutathion peut être trouvé sous forme oxydée ou réduite. Sous forme réduite, il maintient le potentiel d'oxydo-réduction du cytoplasme dans un état réduit et sert aussi de cofacteur à de nombreuses enzymes (Deponte, 2013). Le glutathion peut être transporté du cytoplasme vers le périplasme par des transporteurs membranaires.

3.5.6 Métabolisme énergétique en condition de déshalogénation

L'énergie dans une cellule peut être définie par trois paramètres : le gradient de proton à travers la membrane, la quantité d'ATP et celle de pouvoir réducteur (par exemple NADPH). Parmi les protéines de la chaîne de transport d'électrons (Fig. 15), 11 ont été détectées comme plus abondantes en présence de méthanol (Bibi-Triki *et al.*, 2018) tandis que cinq ARNm sont plus abondants en présence de méthanol (Chaignaud *et al.*, 2017). Des protéines membranaires qui sont

impliquées dans ces complexes n’ont pas été détectées par protéomique (Fig. 15). Cela pourrait être dû au fait que le protocole d’extraction n’a pas été optimisé pour les protéines membranaires, plus difficiles à extraire. Ces résultats laissent penser qu’une modification du métabolisme énergétique a lieu entre ces deux conditions de culture.

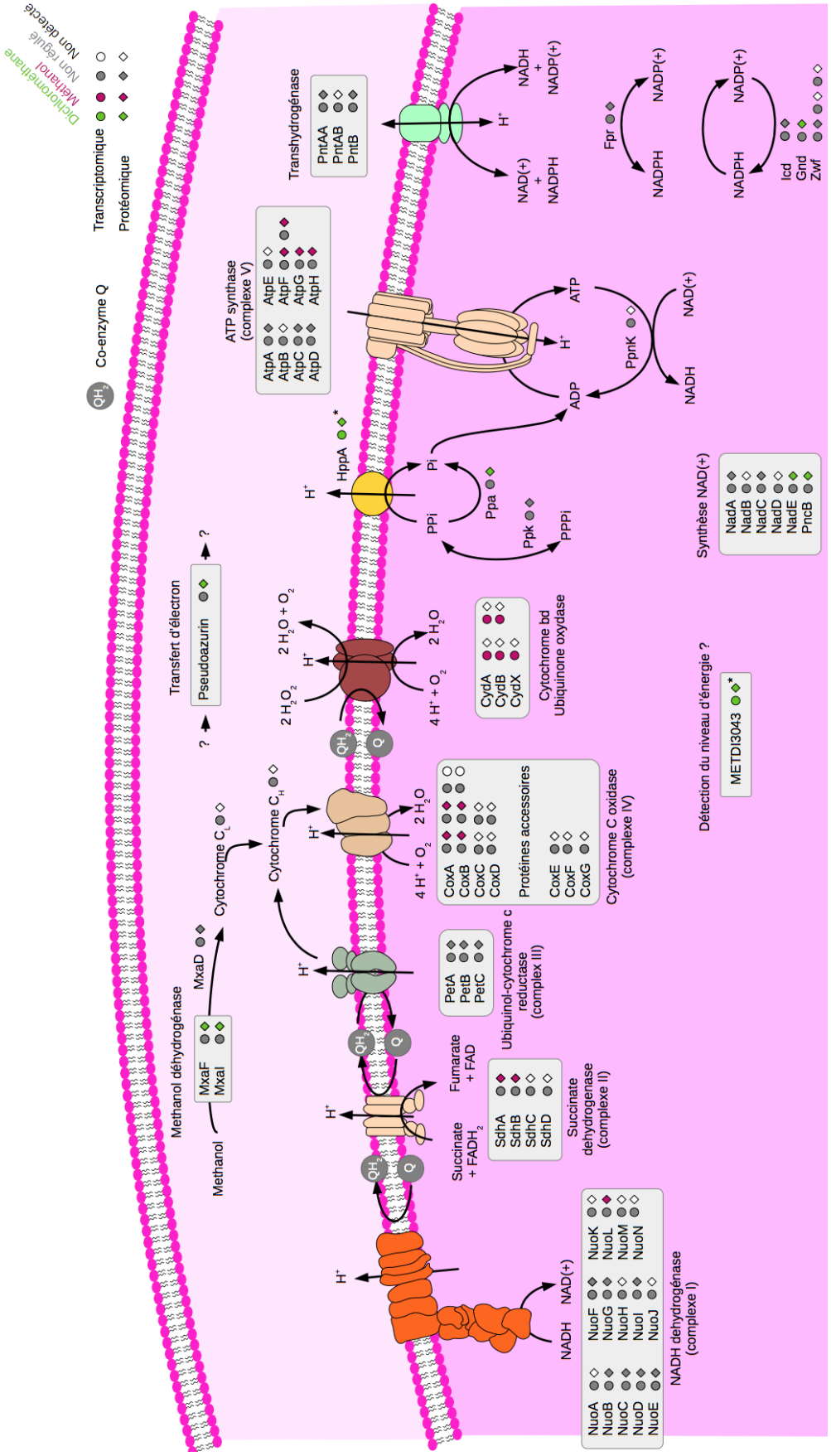

Dichloromethane
Methanol
Non regulé
Non detecté
Co-enzyme Q
Transcriptomique
Protéomique
Methanol déhydrogénase
MxaF
MxaI
MxaD
Methanol
Cytochrome C_L
Cytochrome C_H
Transfert d'électron
Pseudoazurin
?
?
Transhydrogénase
PntAA
PntAB
PntB
NADH + H+
NAD(+) + NADPH
NADP(+)
Fpr
NADPH
NADP(+)
NADPH
NADP(+)
Icd
Gnd
Zwf
ATP synthase (complexe V)
AtpA
AtpB
AtpC
AtpD
AtpE
AtpF
AtpG
AtpH
ATP
ADP
PpnK
NAD(+)
NADH
HppA
PPi
Pi
Ppa
Ppk
PPPi
Synthèse NAD(+)
NadA
NadB
NadC
NadD
NadE
PncB
Détection du niveau d'énergie ?
METDI3043
2 H2O + O2
2 H2O2
4 H+ + O2
2 H2O
CydA
CydB
CydX
Cytochrome bd Ubiquinone oxydase
CoxA
CoxB
CoxC
CoxD
Protéines accessoires
CoxE
CoxF
CoxG
Cytochrome C oxidase (complexe IV)
PetA
PetB
PetC
Ubiquinol-cytochrome c reductase (complex III)
Fumarate + FAD
Succinate + FADH2
SdhA
SdhB
SdhC
SdhD
Succinate dehydrogenase (complexe II)
NADH
NAD(+)
NuoA
NuoB
NuoC
NuoD
NuoE
NuoF
NuoG
NuoH
NuoI
NuoJ
NuoK
NuoL
NuoM
NuoN
NADH dehydrogénase (complexe I)

Figure ci-contre

Figure 3.15: Métabolisme énergétique chez *M. extorquens* DM4.
M. extorquens DM4 est chimio-organohétérotrophe, c'est-à-dire qu'elle utilise des molécules organiques comme source de carbone, d'électrons et d'énergie. La chaîne de transfert d'électron permet de produire un gradient de proton de part et d'autre de la membrane interne. L'ATP synthase utilise ce gradient pour former de l'ATP. Des transhydrogénases permettent de transformer ce gradient de proton en pouvoir réducteur (NADPH).

3.6 Détection des sites d'initiation de la transcription par TSS-seq et recherche de motifs de régulation

3.6.1 Introduction

Identifier au nucléotide près le site d'initiation de la transcription (également appelé TSS pour transcription start site) permettrait d'explorer de nouvelles voies de régulation. En effet, la régulation de l'expression génétique chez *M. extorquens* DM4 en présence de DCM pourrait en partie impliquer des processus de modifications post-transcriptionnelles de la séquence 5'UTR (*untranslated region*) des ARNm. Cette région en amont de la séquence codante contient la séquence Shine-Dalgarno qui permet la fixation du ribosome (sous-unité 30S) à l'ARNm et ainsi la traduction (Gualerzi & Pon, 2015). Des protéines telles que Hfq ou CsrA peuvent aussi se fixer à l'ARNm et en modifier la stabilité ou la traduction (Oliva *et al.*, 2015 ; Sowa *et al.*, 2017). Des ARN régulateurs interviennent aussi dans la régulation, au niveau des séquences 5'UTR en masquant, ou au contraire en rendant accessible le site de fixation du ribosome (Oliva *et al.*, 2015). L'analyse de plus de 190 ARNm chez *E. coli* a permis de montrer que les séquences 5'UTR sont très variables et possèdent des structures 3D pouvant avoir chacune des fonctions biologiques différentes (Mustoe *et al.*, 2018). Des structures secondaires telles que les *riboswitchs* peuvent changer de conformation lorsqu'ils fixent un ligand (Serganov & Patel, 2007), ou encore des tiges boucles peuvent être présentes à l'intérieur des séquences 5'UTR et ainsi réguler la stabilité des ARNm en les protégeant de la dégradation par des ARNase (Arnold *et al.*, 1998 ; Deana *et al.*, 2008). Les séquences 5'UTR sont difficiles à prédire par les outils bio-informatiques et nécessitent de connaître précisément où commence l'ARNm.

L'une des limites de la détection des TSS réside dans la discrimination entre les ARNm non maturés et appelés primaires, des ARNm maturés et dégradés par une ARNase. Or, il existe bien une spécificité dans l'extrémité 5' des ARNm primaires. Le premier nucléotide d'un ARNm possède une extrémité 5' triphosphate tandis que les suivants possèdent une extrémité 5' monophosphate à cause des liaisons phosphates entre les nucléotides. Un ARNm avec une extrémité 5' monophosphate peut être trouvé après clivage de l'ARNm par une ARNase. La présence d'une extrémité 5' triphosphate indique donc que l'ARNm n'a pas été maturé par clivage. La présence d'un nucléotide triphosphate en 5' de l'ARNm permet de protéger l'ARNm de la dégradation par des ARNase telle que la terminale exonucléase (Höfer & Jäschke, 2018). Plusieurs stratégies expérimentales se sont basées sur cette différence de nucléotide au niveau de l'extrémité 5' pour détecter les TSS : « *differential* RNA-seq » (Sharma *et al.*, 2010), « TSS-seq » (Tolba *et al.*, 2013) ou « Cappable-seq » (Ettwiller *et al.*, 2016).

Le génome de *M. extorquens* DM4 a été annoté par un *consortium* d'expert en se basant sur une analyse *in sillico* avec notamment la recherche de domaines fonctionnels caractérisés, et décrits dans la littérature scientifique (Vuilleumier *et al.*, 2009). L'annotation d'un gène nécessite de prédire la fonction supposée de la séquence codante mais également des séquences de régulation de ce gène (Eckweiler *et al.*, 2018). La comparaison des résultats expérimentaux des sites d'initiation de la transcription et de la traduction délimite les séquences 5'UTR des ARNm qui sont souvent impliquées dans les mécanismes de régulation. Ces données permettent également de confirmer expérimentalement l'annotation du génome de *M. extorquens* DM4, représentant un véritable défi à la vue de l'ensemble des séquences générées par l'évolution des technologies de séquençage (Richardson & Watson, 2013).

3.6.2 Matériel et méthode

3.6.2.A Construction des librairies pour le TSS-seq

Les banques ont été construites par la société Vertis Biotechnologie AG (Allemagne) selon un protocole décrit précédemment (Lin *et al.*, 2013) (Fig. 16) à partir des ARNs extraits et préparés à Strasbourg par une ancienne doctorante de l'équipe, Pauline Chaignaud. Les ARN ont été extraits de culture au milieu de la phase exponentielle de croissance de la souche *M. extorquens* DM4 avec comme seule source de carbone et d'énergie du méthanol ou du DCM, comme précédemment décrit

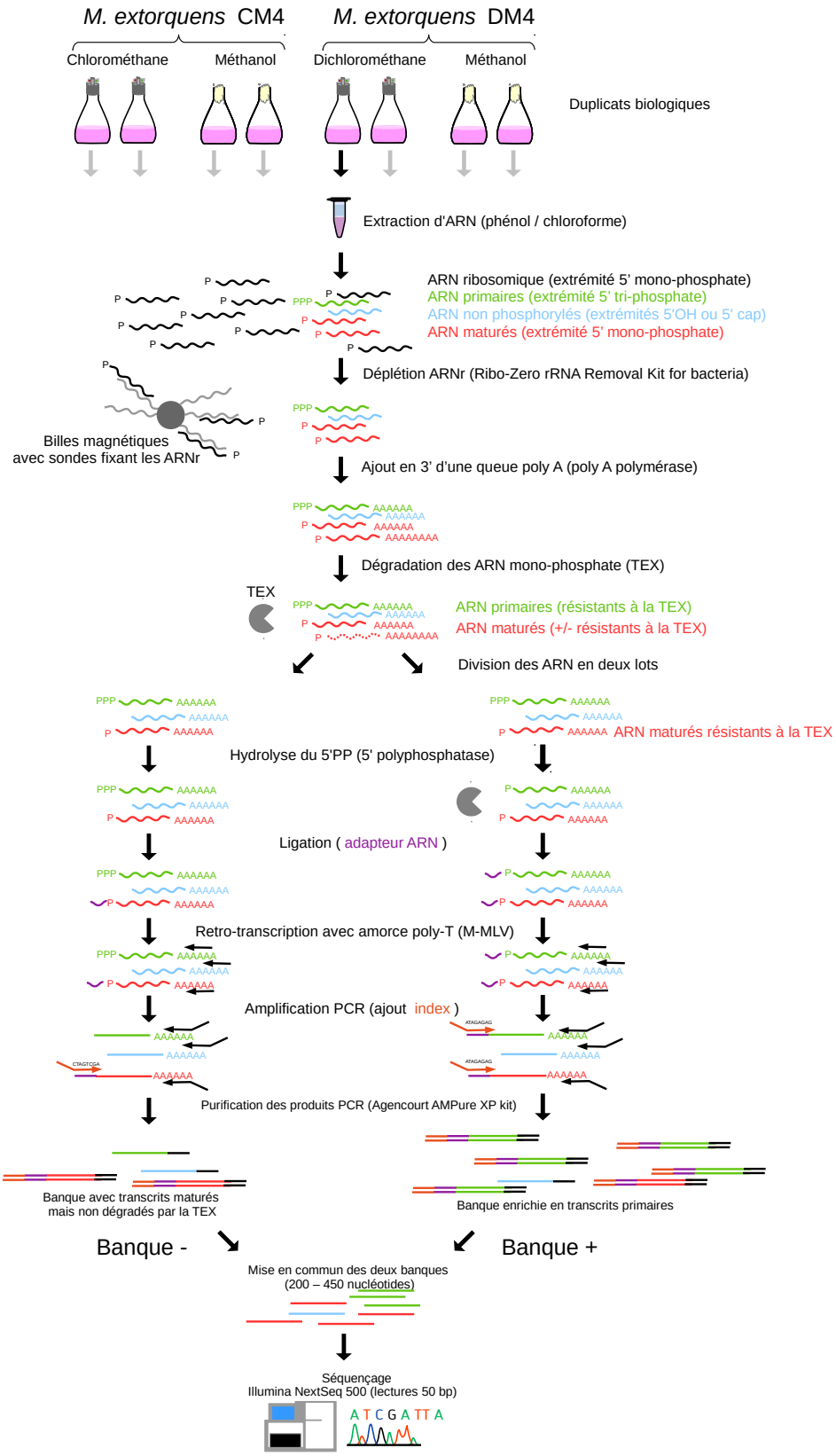
(Chaignaud *et al.*, 2017). Pour réaliser l'approche de TSS-seq, les ARN ribosomiques représentant jusqu'à 85 % des ARN cellulaires sont déplétés à l'aide du kit Ribo-Zero™ Magnetic Kit for Bacteria (Epicentre) qui consiste en des sondes nucléiques fixant les ARNr. Les ARN enrichis sont poly-adénylés en 3' par l'enzyme polyA polymérase. Dans la majorité des études, deux banques sont ensuite réalisées en parallèle : l'une traitée avec une enzyme (TEX pour Terminator EXonuclease) qui dégrade les ARN maturés (5' monophosphate) et l'autre moitié des ARN est non traitée (Sharma *et al.*, 2010). L'inconvénient de cette approche est de générer des faux positifs de par la présence d'ARN n'ayant pas une extrémité 5' triphosphate mais résistants à la TEX, par la présence de structures en tige-boucles au niveau de l'extrémité 5' de l'ARN bloquant l'action de l'enzyme (Deana *et al.*, 2008). Pour diminuer les faux positifs, la société Vertis a modifié le protocole de (Sharma *et al.*, 2010) en séparant les deux banques non pas par traitement à la TEX mais par un traitement différentiel avec une ARN 5' polyphosphate (Fig. 16 ; (Lin *et al.*, 2013)). Dans cette étude, deux banques ont été réalisées qui se différencient par un traitement de la banque « + » avec une ARN 5' polyphosphatase tandis que la banque « - » ne l'a pas été. Cette enzyme clive deux des trois phosphates qui se trouvent en 5' des ARN non maturés. Un ARN adaptateur est ajouté en 5' monophosphate par une ARN ligase (une extrémité 5 triphosphate ne permet pas la fixation). La séquence de cet adaptateur varie entre les deux banques, ce qui permet de les différencier lors du séquençage. L'étape de rétro-transcription nécessite que les amorces poly(T) se fixent aux séquences polyA pour que la rétro-transcriptase (M-MLV) puisse polymériser les ADNc. Pour amplifier ces ADNc, une PCR utilisant une ADN polymérase à haute fidélité et un couple d'amorce reconnaissant l'adaptateur et la séquence poly-T a été réalisée. Les produits PCR sont purifiés avec le kit Agencourt AMPure XP kit (Beckman Coulter Genomics) puis sont ensuite analysés par électrophorèse capillaire (Shimadzu MultiNA *microchip electrophoresis system*). Enfin, les ADN sont séquencés à l'aide de la technologie Illumina NextSeq 500 system avec une taille de lecture de 50 paires de bases. Le séquençage permet de différencier les brins « + » et « - » de l'ADN. La banque « + » a été enrichie en ARN primaires tout en ayant aussi des ARN résistants à la TEX. La banque « - » est composée des ARN résistants à la TEX mais les ARN primaires ne sont pas amplifiés par PCR car il n'y a pas eu de fixation des ARN adaptateurs.

Figure ci-contre

Figure 3.16 : Constructions et séquençages des banques « + » et « - ».
Les ARN extraits de *Methylorubrum extorquens* CM4 ou DM4 sont purifiés puis les ARN maturés sont dégradés par la TEX (terminale exonucléase). Les ARN obtenus sont ensuite divisés en deux banques qui sont traités ou non par une 5' polyphosphatase (5'PP), différenciant les ARN primaires des ARN maturés résistants à la TEX. Le même protocole a été utilisé pour le traitement des ARN extraits de *M. extorquens* CM4 et DM4.

3.6.2.B Analyse informatique des données de TSS-seq

La qualité des données brutes (.fastq) du séquençages a été analysée avec le logiciel FastQC (version 0.11.5) pour vérifier leur qualité. Ce logiciel compare plusieurs paramètres tels que la taille des séquences, le score de qualité attribué lors du séquençage pour chaque base, le ratio entre les quatre nucléotides et le pourcentage du contenu en GC. Après ce contrôle, les séquences retenues ont été alignées sur le génome de référence par le logiciel *Burrows-Wheeler Alignment tool* (Version 0.7.4). Pour localiser la position des TSS, les séquences alignées sont analysées avec le logiciel TSSAR (serveur http://rna.tbi.univie.ac.at/TSSAR ; (Amman *et al.*, 2014)), l'analyse se basant sur le postulat que la distribution des TSS sur le génome suit une loi de Poisson (Frank, 2009). Les régions génomiques ne possédant pas de lectures sont éliminées de l'échantillon pour ne pas perturber la distribution de Poisson au niveau du génome. Une distribution de Skellam est utilisée pour étudier les différences entre la distribution de Poisson pour la banque « + » et pour la banque « - ». Un TSS est défini pour chaque position génomique où la distribution de Skellam indique une différence significative entre les deux banques.


M. extorquens CM4
Chlorométhane
Méthanol
M. extorquens DM4
Dichlorométhane
Méthanol
Duplicats biologiques
Extraction d'ARN (phénol / chloroforme)
ARN ribosomique (extrémité 5' mono-phosphate)
ARN primaires (extrémité 5' tri-phosphate)
ARN non phosphorylés (extrémités 5'OH ou 5' cap)
ARN maturés (extrémité 5' mono-phosphate)
Déplétion ARNr (Ribo-Zero rRNA Removal Kit for bacteria)
Billes magnétiques
avec sondes fixant les ARNr
Ajout en 3' d'une queue poly A (poly A polymérase)
Dégradation des ARN mono-phosphate (TEX)
TEX
ARN primaires (résistants à la TEX)
ARN maturés (+/- résistants à la TEX)
Division des ARN en deux lots
ARN maturés résistants à la TEX
Hydrolyse du 5'PP (5' polyphosphatase)
Ligation (adapteur ARN)
Retro-transcription avec amorce poly-T (M-MLV)
Amplification PCR (ajout index)
Purification des produits PCR (Agencourt AMPure XP kit)
Banque avec transcrits maturés
mais non dégradés par la TEX
Banque enrichie en transcrits primaires
Banque -
Banque +
Mise en commun des deux banques
(200 – 450 nucléotides)
Séquençage
Illumina NextSeq 500 (lectures 50 bp)
A T C G A TT A

Le logiciel TSSAR effectue les analyses séparément pour chaque condition de culture et pour chaque réplicon. Ce logiciel propose plusieurs paramètres qui influent sur le nombre et la précision de TSS détectés. Le paramètre « *Noise* » correspond au nombre minimal de lectures pour détecter un TSS. Le paramètre « *Merge* » correspond à la distance maximale entre deux TSS pour qu'ils soient fusionnés et considérés comme un seul. Une p-value est calculée à partir du résultat d'un test statistique (régression de Poisson) comparant les deux banques. Le nombre de TSS détecté varie d'un facteur dix selon les paramètres choisis plus ou moins stringents (Fig. 17). Pour diminuer le nombre de faux positifs, les paramètres les plus stricts ont été retenus : p-value = 0,01 ; *Noise* = 10 et *Merge* = 1. Dans ce cas, le nombre de TSS détecté est du même ordre de grandeur que le nombre de gènes annotés (5 988) prédits à partir du génome de *M. extorquens* DM4 (Vuilleumier *et al.*, 2009).

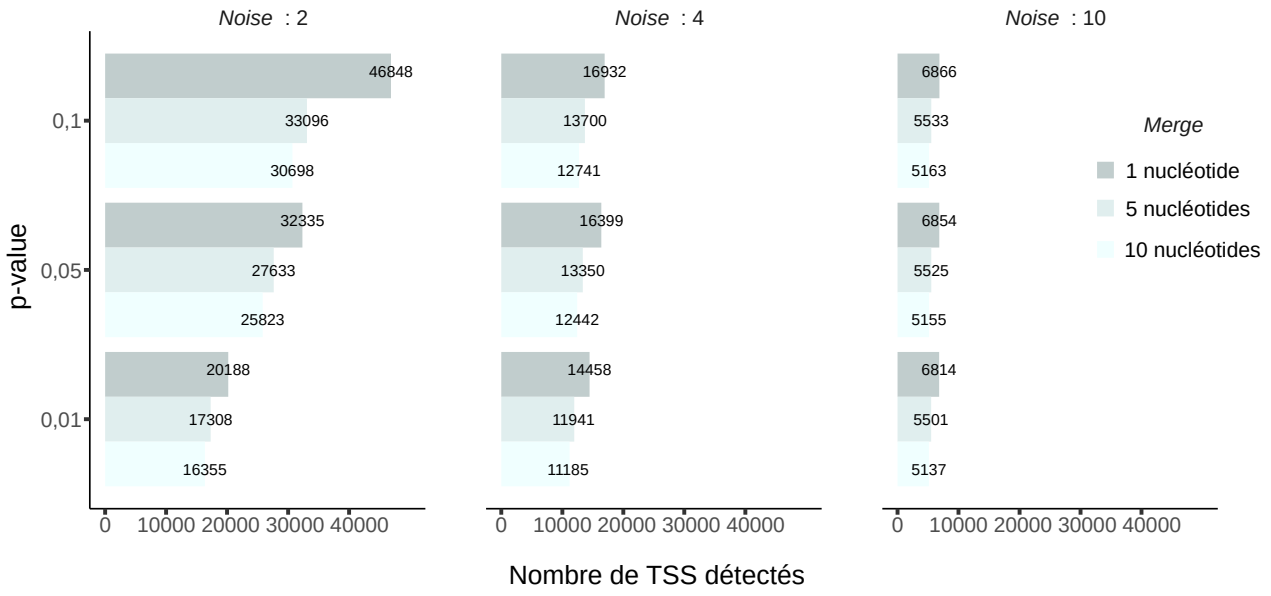


Figure 3.17: Impact des paramètres du logiciel TSSAR sur le nombre de TSS détecté pour des données obtenues pour un réplicat en condition DCM.

Une validation manuelle des TSS détectés par le logiciel TSSAR a été réalisée pour plus de 500 gènes en visualisant les lectures alignées sur le génome par le logiciel *Integrative Genomics Viewer* (Robinson *et al.*, 2011). Pour un même gène, plusieurs TSS peuvent être prédits à de très faibles distances (moins de 10 pb) les uns des autres et avec une faible différence de nombre de lectures (Fig. 4). Ces TSS seront considérés comme des faux positifs. Un filtre a été réalisé manuellement afin de les éliminer. Pour cela, le TSS considéré comme correct est celui qui aura le plus de lectures

en 5' (c'est-à-dire le plus grand ARN) et ayant un nombre de lectures supérieures au bruit de fond (fixé à 10 lectures). Le logiciel ne pouvant analyser qu'une condition de culture à la fois, la validation manuelle permet de comparer les résultats dans les différentes conditions.

Une fois qu'un TSS est détecté et validé manuellement, un script R avec les librairies *GenomicRanges* (Lawrence *et al.*, 2013) et *seqinR* (Charif *et al.*, 2005) est utilisé pour extraire la séquence nucléotidique en amont du promoteur ainsi que celle de la séquence 5'UTR. Pour rechercher des motifs à l'intérieur de ces séquences, le logiciel MEME suite (Bailey *et al.*, 2009) (Bailey *et al.*, 2009) est utilisé. Il permet la recherche de motifs *de novo via* l'outil MEME ou la recherche de séquences connues dans des bases de données telles que Swiss regulon (Pachkov *et al.*, 2007) et Prodoric *database* (version 8.9 ; (Münch *et al.*, 2003 ; Eckweiler *et al.*, 2018)) avec l'outil Centrimo (Bailey *et al.*, 2009). Pour rechercher des structures d'ARN conservées entre plusieurs séquences 5'UTR, le logiciel BEAM (version 1.6.1) est utilisé (Pietrosanto *et al.*, 2018). La recherche de structures d'ARN régulateurs a été réalisée en utilisant la base de données Rfam (Kalvari *et al.*, 2018).

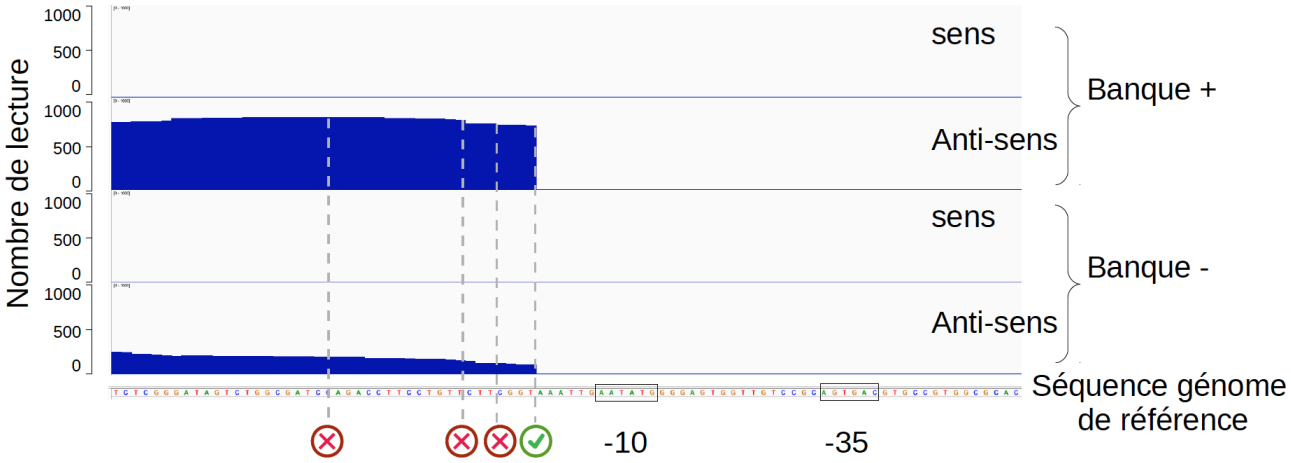


Figure 3.18: Exemple de détection de multiples TSS pour un même gène.
Le logiciel TSSAR a détecté quatre TSS pour le gène METDI2461 (plus exprimé en présence de DCM). Un seul a été retenu en tenant compte de celui qui est le plus éloigné du codon start et avec un nombre de lectures supérieur à 10. Un promoteur a pu être prédit à une distance adéquate de ce TSS. Les boîtes -10 et -35 prédites du promoteur par le logiciel PromoterHunter sont affichées. La visualisation a été obtenue avec le logiciel IGV.

3.6.3 Résultats et discussions

3.6.3.A Détection globale des TSS chez *M. extorquens* DM4

L'analyse des données de RNA-seq a révélé que seulement 5 à 14 % des lectures s'alignant sur le génome sont exploitables tandis que de nombreuses séquences polyA ont été détectées (Tableau 1). En effet, lors de la construction de la banque, la polyA polymérase semble avoir réagi avec des fragments d'ARN dégradés, aboutissant à une majorité de séquences polyA et réduisant ainsi le nombre de lectures correspondant à des ARNm. Ces artefacts de construction de banques ont été retrouvés à des taux similaires dans la banque faite en parallèle avec les ARN de *M. extorquens* CM4.

Tableau n°1 : nombre de lectures alignées sur le génome

	M. extorquens CM4				*M. extorquens* DM4			
Banque	Méthanol-1	Méthanol-2	Chlorométhane-1	Chlorométhane-2	Méthanol-1	Méthanol-2	Dichlorométhane-1	Dichlorométhane-2
+	$1{,}0\cdot10^6$ (11%)[a]	$0{,}7\cdot10^6$ (9%)	$0{,}7\cdot10^6$ (9%)	$1{,}2\cdot10^6$ (21 %)	$0{,}6\cdot10^6$ (9%)	$0{,}3\cdot10^6$ (5%)	$0{,}8\cdot10^6$ (14%)	$0{,}6\cdot10^6$ (8%)
-	$3{,}7\cdot10^6$ (77%)	$4{,}6\cdot10^6$ (75%)	$4{,}2\cdot10^6$ (78%)	$4{,}9\cdot10^6$ (81%)	$5{,}7\cdot10^6$ (76%)	$4{,}0\cdot10^6$ (74%)	$2{,}2\cdot10^6$ (51%)	$3{,}0\cdot10^6$ (64%)

[a] Le pourcentage correspond au nombre de lectures alignées comparées au nombre total de lectures obtenues lors du séquençage.

Un total de 8 273 TSS a été détecté pour l'ensemble des conditions testées (Fig. 19). Un criblage manuel a permis de filtrer 629 TSS comme faux positifs, dont la présence pourrait être liée au faible nombre de lectures alignées sur le génome. Leur élimination facilite ainsi la recherche de motifs en évitant d'avoir une répétition de séquences variant de quelques nucléotides. Les résultats ont été analysés de façon qualitative et non pas quantitative.

Les TSS sont classés en différentes catégories, selon leur position sur le génome, par le logiciel TSSAR. Le logiciel définit comme « primaires », les TSS étant à moins de 250 nucléotides du codon start d'un gène. Pour éviter la confusion entre ces TSS « primaires » en amont d'un gène et les ARNm « primaires » c'est-à-dire non maturés, ces TSS sont renommés « Gène TSS ». Ils permettent de déterminer la taille des séquences 5'UTR (Fig. 19A). La limite de 250 nucléotides

pour définir ces TSS a permis de classer un certain nombre de TSS comme de faux négatifs car se trouvant plus loin que cette limite et qui n'ont donc pas été classés comme « Gène TSS ». Toutefois, choisir une longueur trop importante peut aussi conduire à des faux positifs, les TSS primaires étant confondus avec des ARN non codants. Cependant, d'autres études ont considéré des longueurs différentes (de 300 à 500 nucléotides) pour ces TSS « Gène TSS » (Schlüter *et al.*, 2013 ; Jeong *et al.*, 2016 ; Adams *et al.*, 2017 ; Albersmeier *et al.*, 2017 ; Zhukova *et al.*, 2017).

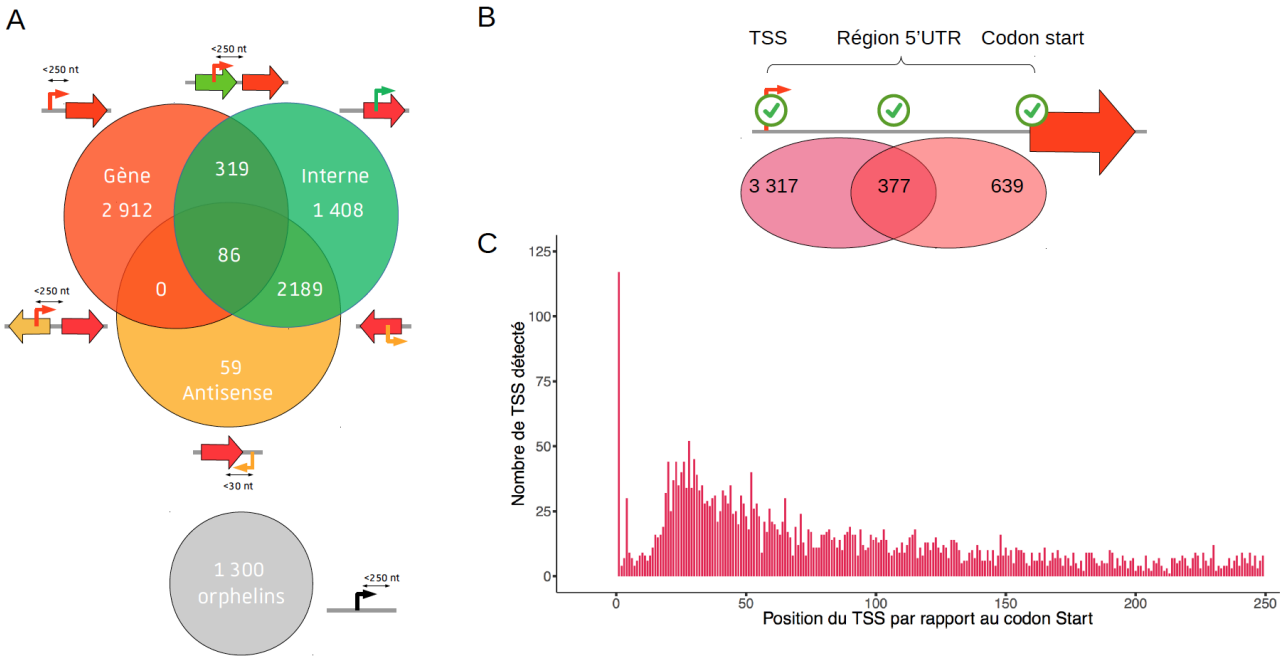


Figure 3.19: TSS détectés pour l'ensemble du génome de *M. extorquens* DM4.
A Nombre de TSS détectées pour chaque catégorie de TSS proposée par le logiciel TSSAR. B Comparaison des résultats de dN-TOP et des gènes TSS permettant de déduire les séquences 5'UTR. C Classification des gènes TSS en fonction de la taille de la région 5'UTR.

Chez *M. extorquens* DM4, la taille moyenne des 5'UTR est de 84 nucléotides, avec une valeur médiane de 64 nucléotides (Fig. 19C). Ce type de profil a été obtenu chez d'autres bactéries (Kim *et al.*, 2012 ; Kröger *et al.*, 2012 ; Schlüter *et al.*, 2013). La taille des 5'UTR peut influencer leur stabilité (Arnold *et al.*, 1998). Les gènes appelés *leaderless* ne possèdent pas de région 5'UTR car le site d'initiation de la transcription se superpose avec celui de la traduction (Beck & Moll, 2018). Cette catégorie comprend 118 gènes chez *M. extorquens* DM4 (4 % des TSS primaires). Ce pourcentage varie chez les autres souches bactériennes, de 0,7 % chez *E. coli* BW25113 à 47 % chez *Deinococcus deserti* RD19 (Beck & Moll, 2018). Toutefois les ARN *leaderless* qui résultent d'un clivage à partir d'ARN ayant une région 5'UTR ne seront alors pas détectés par cette approche

de transcriptomique (Beck & Moll, 2018).

Des TSS anti-sens ont également été détectés chez *M. extorquens* DM4, représentant 28 % de l'ensemble des TSS (Fig. 19A). Une proportion similaire a été détectée chez d'autres bactéries avec 13 % chez *Salmonella enterica* serovar Typhimurium ; 22 % chez *Sinorhizobium meliloti* et 37 % chez *E. coli* (Georg & Hess, 2018). Les ARN anti-sens peuvent avoir des rôles physiologiques variés tels que dans l'appariement avec des ARN complémentaires pour en modifier la stabilité ou la traduction (Georg & Hess, 2011, 2018). Les ARN anti-sens peuvent également provenir de la transcription pervasive correspondant à la synthèse d'ARN non-codant pour les brins sens et anti-sens (Lybecker *et al.*, 2014 ; Georg & Hess, 2018). La transcription pervasive est souvent considérée comme un « bruit de fond » mais pourrait néanmoins être liée à des fonctions biologiques tels que le couplage de la transcription et de la réparation de l'ADN (Monnet *et al.*, 2013 ; Lybecker *et al.*, 2014).

Les TSS dits « orphelins » sont situés à plus de 250 nucléotides d'un codon start. Leurs origines sont multiples. Ils peuvent correspondre à des TSS primaires dépassant la limite arbitraire de taille (250 pb), à des gènes non annotés, à des ARN non-codant, ou encore être associés à la transcription pervasive (Lin *et al.*, 2013).

3.6.3.B Analyse des TSS des gènes du transposon catabolique *dcm*

Dans le cadre de l'îlot *dcm*, un TSS a été détecté par l'approche TSS-seq pour le gène *dcmA* (P_A ; Fig. 20) à une distance de trois nucléotides de celui déterminé par une étude précédente par cartographie à la nucléase S1 (La Roche & Leisinger, 1991). Des différences de positionnement selon la méthode utilisée ont été relevées dans des études antérieures (Kim *et al.*, 2012). Pour le gène *dcmR*, un TSS a été détecté à trois nucléotides de celui déterminé pour le promoteur P_{R1}, un autre est prédit à la même position pour le promoteur P_{R2}, tandis que ceux des promoteurs P_{R3} et P_{R4} ne sont pas détectés par TSS-seq appuyant l'hypothèse qu'ils pourraient correspondre à des artefacts (La Roche & Leisinger, 1991). La séquence putative de fixation de DcmR est retrouvée dans l'un des promoteurs de *dcmR* (P_{R1}) et dans celui de *dcmA* (La Roche & Leisinger, 1991).

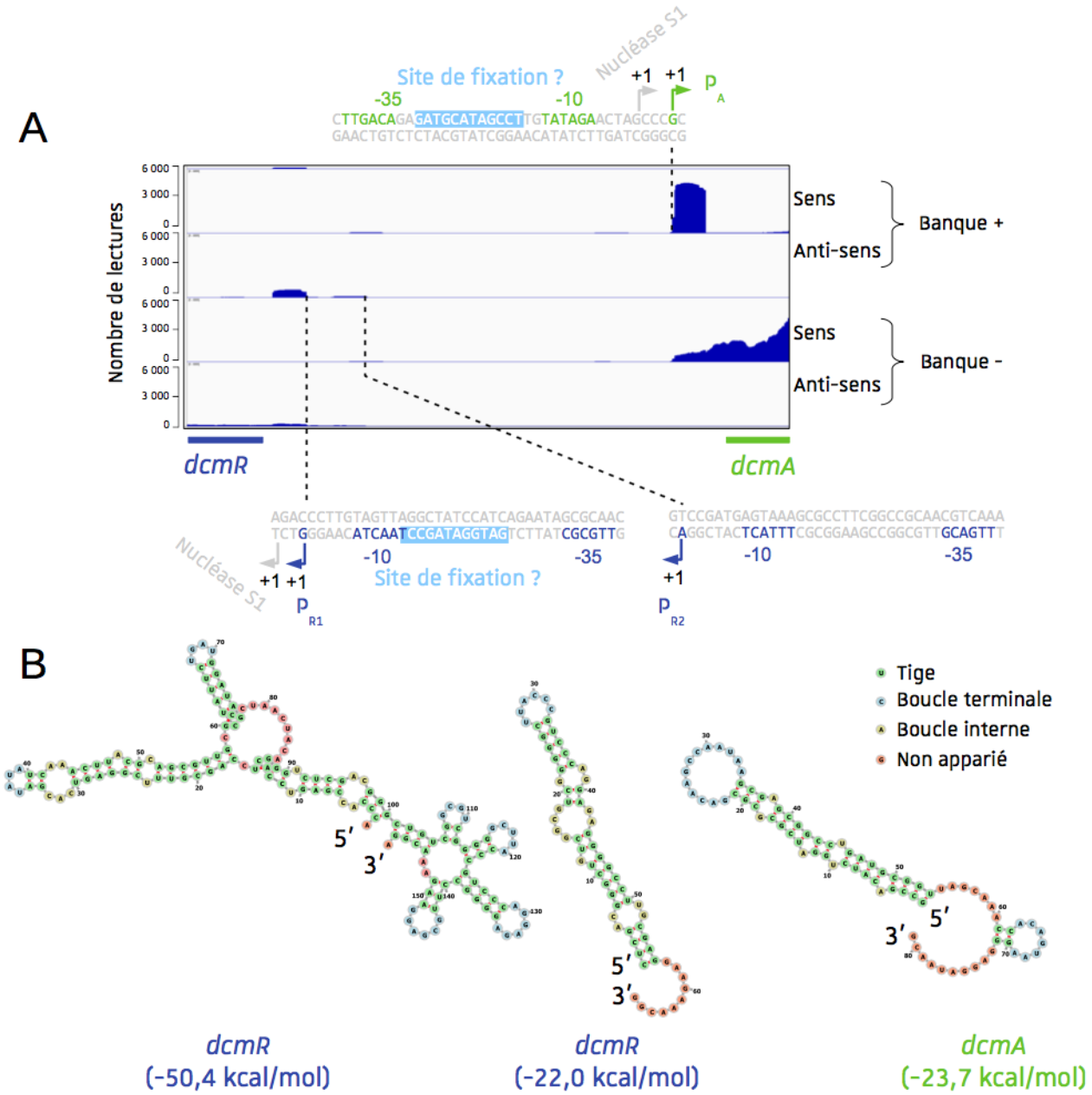


Figure 3.20: Sites d'initiation de la transcription et prédiction de la structure secondaire des séquences 5'UTR des transcrits de *dcmA* et *dcmR*.

A Représentation graphique des lectures de TSS-seq pour la région intergénique de *dcmA* et *dcmR*. B Prédiction de la structure secondaire des séquences 5'UTR des transcrits de *dcmR* et *dcmA* par RNAfold. L'énergie libre de la structure calculée par ce logiciel est indiquée en dessous des structures. Plus cette valeur est faible et plus la structure est supposée être stable.

3.6.3.C Analyse détaillée des TSS des gènes plus exprimés en présence de DCM ou méthanol

La comparaison des données transcriptomiques TSS-seq et RNA-seq a permis de déterminer les promoteurs des gènes régulés en présence de méthanol ou DCM (Fig. 21 ; Tableau 2). Un total de 33 TSS ont été identifiés pour les 69 gènes ayant des transcrits plus abondants en conditions DCM. Tandis qu'en condition méthanol, 21 TSS ont été détectés pour les 118 gènes plus exprimés. En tenant compte des opérons produits, les TSS ont été trouvés pour 37 et 33 gènes respectivement en DCM et méthanol.

Figure ci-contre

Figure 3.21 : Sites d'initiation de la transcription, identifiés pour les gènes régulés au niveau de la transcription chez *M. extorquens* DM4.
A *Loci* avec des sites d'initiation détectés en présence de DCM. B *Loci* avec des sites d'initiation détectés en présence de méthanol. C *Locus* avec des sites d'initiations détectés en présence de DCM ou de méthanol.

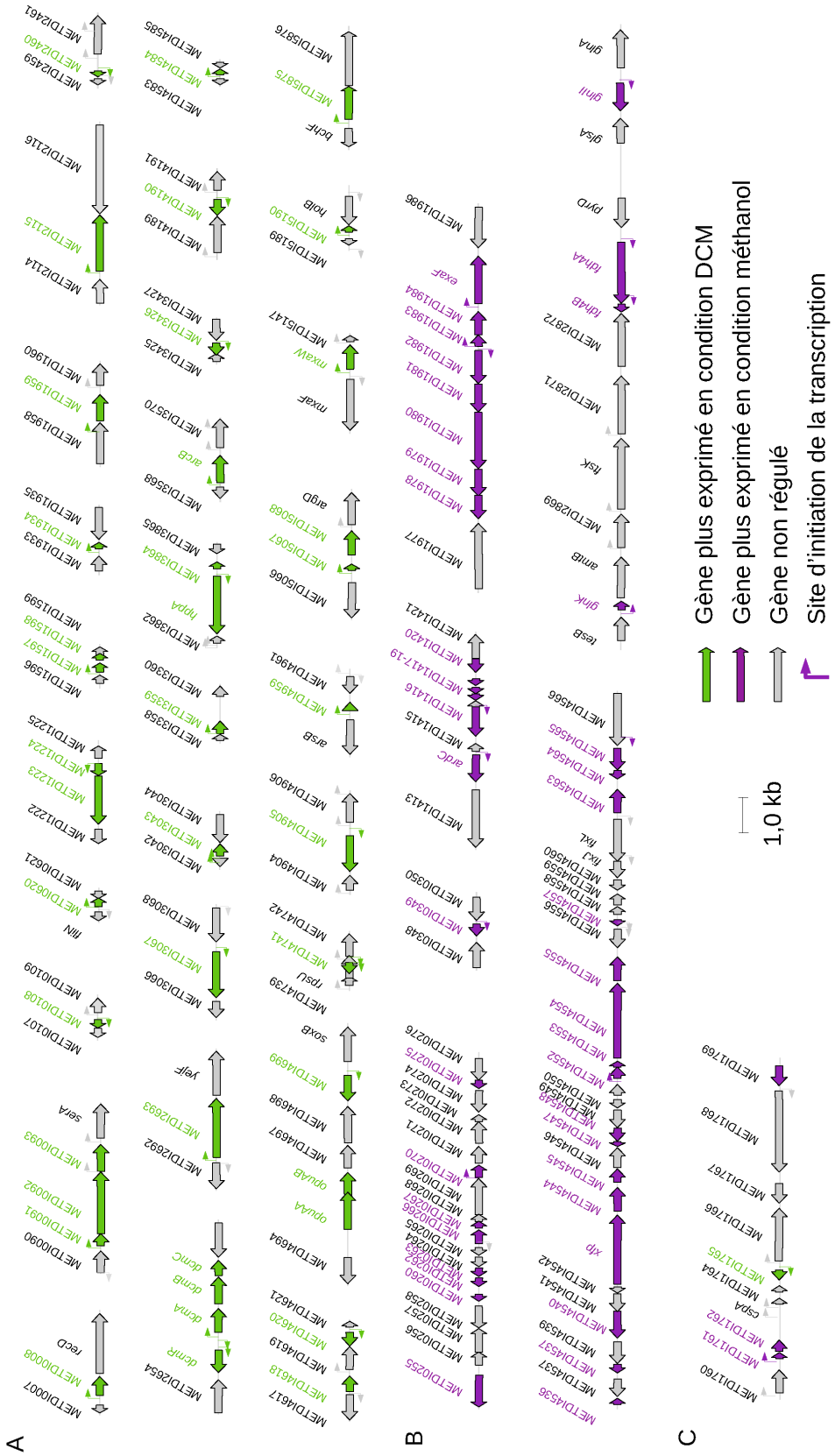

A
B
C
Gène plus exprimé en condition DCM
Gène plus exprimé en condition méthanol
Gène non régulé
Site d'initiation de la transcription
1,0 kb

Tableau n°2 : Identification des sites d'initiation de la transcription des gènes dont l'abondance des transcrits varie en présence de DCM ou méthanol

Gène	Nom	Position [a]	Brin	Nucléotide	Boite -35 [b]	Boite -10 [b]	Distance	Commentaire
Transcrits plus abondants en présence de DCM								
METDI0008	/	8097	Plus	T	ND	ND	63	P [c]
METDI0091	/	91476	Plus	A	TTGCCA	GATATT	18	P
METDI0091	/	91470	Plus	G	ND	ND	24	P
METDI0108	/	110830	Minus	G	ND	ND	48	P
METDI0620	/	586227	Plus	A	ND	ND	29	P
METDI1224	/	1123413	Minus	A	ND	ND	0	P
METDI1597	/	1464904	Plus	A	ND	ND	17	P
METDI1765	/	1634904	Minus	G	TTGACT	TACGCT	25	P
METDI1934	/	1811535	Plus	A	TTGCCG	TAAGAT	102	P
METDI1959	/	1833198	Plus	G	ND	ND	217	IP
METDI2115	/	1993686	Plus	A	TTCCCG	AAATAT	44	P
METDI2461	/	2371778	Minus	A	GTCACT	CATATT	61	P
METDI2655-1[e]	*dcmR*	2562332	Minus	C	TTGCGC	TAACTA	66	P ; P_{R1}
METDI2655-2	*dcmR*	2562423	Minus	A	TTGACG	TTTACT	157	P ; P_{R2}
METDI2656	*dcmA*	2562877	Plus	G	TTGACA	TATAGA	173	P ; P_{A}
METDI2693	/	2595011	Plus	T	TTTTTC	TAGGCT	46	P
METDI3043	/	2957385	Plus	A	TTGCCA	CACTCG	74	P
METDI3043	/	2957334	Plus	T	TTCCGA	CAATAG	125	PAi
METDI3067	/	2981559	Minus	G	ATTTCG	TAAAAT	43	P
METDI3359	/	3291364	Plus	A	TCCACA	TAACTT	0	P
METDI3426	/	3347177	Minus	A	GTGACA	TCTTCT	37	P
METDI3569	*arcB*	3503334	Plus	A	TCGATT	CATGCT	49	P
METDI3571	/	3505673	Plus	G	ND	ND	87	IP
METDI3571	/	3505666	Plus	G	ND	ND	94	IP
METDI3863	*hppa*	3799402	Minus	G	TGGACG	TAAGTT	104	P
METDI4190	/	4123281	Minus	A	TTTACG	TATCGT	151	P
METDI4584	/	4464978	Plus	G	CAGATT	TATACC	50	P
METDI4620	/	4501380	Minus	G	TAGAGC	TAAGAT	95	P
METDI4699	/	4585255	Minus	G	TTGCCT	TTAAGT	245	P
METDI4741	/	4637478	Minus	G	ND	ND	87	P
METDI4741	/	4637616	Minus	G	ND	ND	225	P
METDI4905	/	4800535	Minus	G	ND	ND	211	P
METDI4959	/	4857811	Plus	G	ND	ND	32	P
METDI5067	/	4981146	Plus	A	TTGCCT	TACCTT	87	P
METDI5068	/	4981626	Plus	G	ND	ND	154	P
METDI5068	/	4981632	Plus	A	ND	ND	148	P
METDI5146	*mxaW*	5068271	Plus	C	ND	ND	41	P
METDI5190	/	5109864	Plus	G	TGGACA	GTTAAT	80	P
METDI5875	/	5868036	Plus	A	TTGCCA	TATCGT	82	P
Transcrits plus abondant en présence de méthanol								
METDI0270	/	262250	Plus	A	TTGACG	TGATCT	61	P
METDI0349	/	331579	Minus	G	ND	ND	46	P
METDI0384	/	357355	Minus	G	ND	ND	82	P
METDI1416	/	1286085	Minus	A	ND	ND	229	P
METDI1753	*eshA*	1623245	Plus	G	TTGACC	TAGTAC	210	P
METDI1754	/	1624775	Plus	G	ND	ND	58	IP
METDI1761	/	1631616	Plus	G	ND	ND	27	P
METDI1982	/	1857504	Minus	A	ND	ND	29	P
METDI1983	/	1857618	Plus	A	ND	ND	19	P
METDI2867	*glnK*	2797153	Plus	G	TTTATG	TTATTT	188	P
METDI2873	*fdh4B*	2808782	Minus	A	ND	ND	114	IP

METDI2874	*fdh4A*	2811072	Minus	A	ND	ND	92	P
METDI3157	*glnII*	3075246	Minus	G	ND	ND	192	P
METDI4552	/	4439197	Plus	A	ND	ND	68	P
METDI4565	/	4451683	Minus	G	TCGGCA	TAGGAT	221	IP
METDI5284	/	5218043	Minus	A	ND	ND	35	P

[a] Position sur le génome de *M. extorquens* DM4 (numéro d'accession FP103042.2).
[b] Prédiction avec le logiciel PromoterHunter (Klucar *et al.*, 2010).
[c] ND : Non disponible
[d] Classification des TSS selon le logiciel TSSAR : P (primaire) ; IP (primaire interne) ; PAi (primaire anti-sens interne).
[e] Un chiffre est ajouté après le nom d'un gène lorsque deux promoteurs ont été prédits pour celui-ci.

À part le motif GATG(C/G)ATAGCCT présent au niveau des promoteurs de *dcmA* et *dcmR* (Fig. 20 et 22 ; (La Roche & Leisinger, 1991)) aucun autre motif de régulation n'est connu pour les gènes en réponse au DCM. La co-occurence de motif d'ADN a donc été recherchée au niveau des promoteurs des gènes plus exprimés en présence de DCM. De tels motifs pourraient correspondre à des séquences de régulation permettant par exemple la fixation de facteur de transcription. Néanmoins, aucun site de fixation de facteur de transcription connu (dans les banques de données Swiss regulon et Prodoric *database*) n'a été détecté pour les gènes plus exprimés en présence de DCM. Au niveau des séquences en amont des TSS, l'utilisation du logiciel MEME a permis l'identification de motifs retrouvés dans plusieurs séquences (Fig. 22). Néanmoins ces motifs paraissent peu conservés parmi les séquences et ne correspondent pas à une séquence inversée répétée, caractéristique des motifs reconnus par les facteurs de transcription. Il s'agit probablement d'artefacts prédits par le logiciel MEME et non pas de séquences ayant un intérêt biologique.

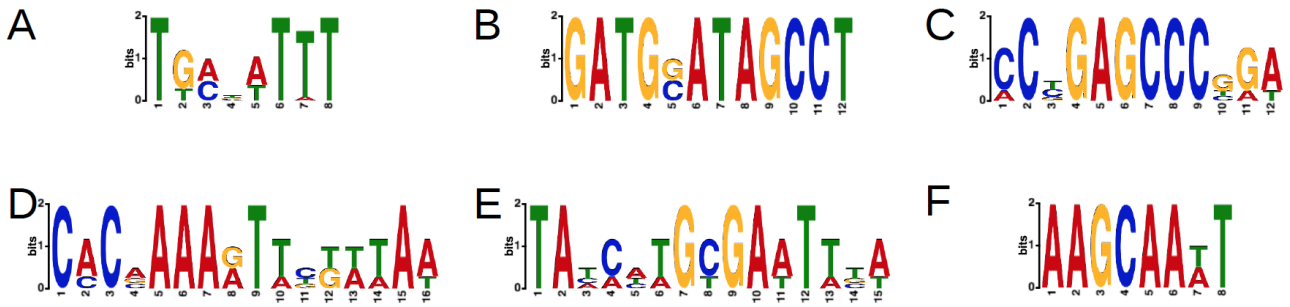


Figure 3.22: Motifs d'ADN proposés par le logiciel MEME en amont des TSS
Les motifs détectés en amont des TSS se trouvant à un maximum de 250 nt en amont des TSS détectés d'un ensemble de gènes plus exprimés en présence de DCM (A, B et C) et en présence de méthanol (D, E et F) sont représentés.

Tableau n°3 : motif trouvé au niveau des promoteurs des gènes dont l'abondance des ARNm est régulée en présence de DCM ou méthanol.

Motif	Gène	Brin	Position [a]	p-value[b]	En amont 5'	Motif détecté par MEME	En aval3'
A	METDI2115	-	227	2.96e^{-6}	gattacgcta	tgatattt	gtgaacgatc
	METDI0108	+	148	5.93e^{-6}	ggagccagcc	tgattttt	ttgggtcgta
	METDI2461	-	169	5.93e^{-6}	actacgcgtc	tgattttt	cggcgccgtc
	METDI0620	+	192	2.20e^{-5}	ccgccgcggc	tgacattt	cgctgacggc
	METDI4620	+	203	2.68e^{-5}	cggctgtgaa	tgcttttt	gcaatagagc
	METDI5190	+	203	3.46e^{-5}	cagtatcctg	ttctattt	tgcatggaca
	METDI0008	-	229	3.46e^{-5}	tcttgatatg	ttctattt	gtggataata
	METDI4190	+	185	4.41e^{-5}	gacttgaggt	tgagtttt	cgccgttgcg
	METDI1934	-	208	5.93e^{-5}	tggaacgacg	tgccattt	ggcacgccgt
	METDI3571	-	239	5.93e^{-5}	ccggg	tgccattt	gccctttatc
	METDI3067	-	197	5.93e^{-5}	gaattgttta	tgcgattt	cgcggggagg
	METDI5875	+	184	6.52e^{-5}	tcattggaga	ttacattt	tcctcagggt
	METDI5146	+	45	6.52e^{-5}	aaggatattc	ttagattt	gcgcggtctt
	METDI3043-2[c]	+	156	6.82e^{-5}	atggtggcgg	tgaatttt	tcttccgatt
	METDI4741-2	+	207	7.59e^{-5}	ttcaggctgg	tgcaattt	cgctgaaccg
	METDI1597	+	237	9.30e^{-5}	accgggtgag	tgcgtttt	cgttgca
	METDI2655-2	-	194	1.21e^{-4}	tgcgtaagtt	tgatatat	cgtgactccg
B	METDI2655-2	+	77	1.07e^{-7}	gcgctattct	gatggatagcct	aactacaagg
	METDI2656	+	73	1.07e^{-7}	gcttgacaga	gatgcatagcct	tgtatagaac
C	METDI3359	-	4	5.32e^{-7}	gccgtgtgcc	ccggagcccgga	ggg
	METDI5190	+	3	7.74e^{-7}	gg	cctgagccccga	gcattttgca
	METDI4959	+	8	7.74e^{-7}	gcatggt	cccgagccctga	tcgagcgatg
	METDI0108	+	58	1.48e^{-6}	atttttgggg	accgagcccgga	atatttcgcg
	METDI2655-1	+	5	3.75e^{-6}	ggcg	cctgagcccgat	ccgagtggtg
D	METDI2874	+	84	6.73e^{-10}	tgatgcaatg	caccaaagttctataa	gtcacgcggc
	METDI5284	+	183	7.83e^{-10}	ccgcggcgcc	cacgaaaatttgttaa	catatcttgc
	METDI1753	+	190	1.33e^{-9}	ctgcggcgtc	cacaaaaattggaaaa	tcattccgtt
	METDI2867	+	205	8.64e^{-9}	cattctgccg	cccaaaagtactttat	gccgactttt
E	METDI1753	+	122	3.68e^{-9}	atttttccat	taaaaagcgaattta	taaaacgttc
	METDI2874	+	60	9.98e^{-9}	cgcgatttct	taccatgtgaatttt	gatgcaatgc
	METDI5284	+	201	1.38e^{-8}	tttgttaaca	tatcttgcgatttca	ccggaacgag
	METDI2867	-	49	2.55e^{-8}	gcgctcgggc	tatcctgcgaataga	gcagccagga
F	METDI0349	+	113	6.07e^{-6}	ccgccggccg	aagcaaat	ggcgttcgct
	METDI4552	+	43	6.07e^{-6}	catgctcgcg	aagcaatt	ggtgcacctc

[a] La position correspond à celle sur un fragment commençant 250 nt avant le TSS.
[b] p-value calculée pour le motif par le logiciel MEME.
[c] Un chiffre est ajouté après le nom d'un gène lorsque deux promoteurs ont été prédits pour celui-ci.

Parmi les gènes plus exprimés en présence de méthanol, des motifs retrouvés chez des gènes impliqués dans le métabolisme du méthanol et souvent plus exprimés en présence de méthanol chez d'autres souches de *Methylobacterium* ou *Methylorubrum* ont été recherchés. Alors que chez *M. organophilum* XX (Xu *et al.*, 1993) le motif AAGAAATG est répété trois fois en amont des promoteurs de *mxoF* et *mxoW* tandis que chez *M. extorquens* AM1, le motif AAGAAA est aussi trouvé en amont de plusieurs gènes (*mxaF*, *mxaW* et *pqqA*), plus exprimés en présence de méthanol, (Zhang & Lidstrom, 2003). Une séquence GAAA est aussi trouvée en amont du gène *mxdB*, plus exprimé en présence de méthanol, chez *M. extorquens* AM1 (Springer *et al.*, 1997). Chez *M. extorquens* DM4, l'approche TSS-seq a permis de définir le TSS du gène *eshA* (METDI1753) plus exprimé en présence de méthanol et de trouver la séquence AGAAAA dupliquée en amont.

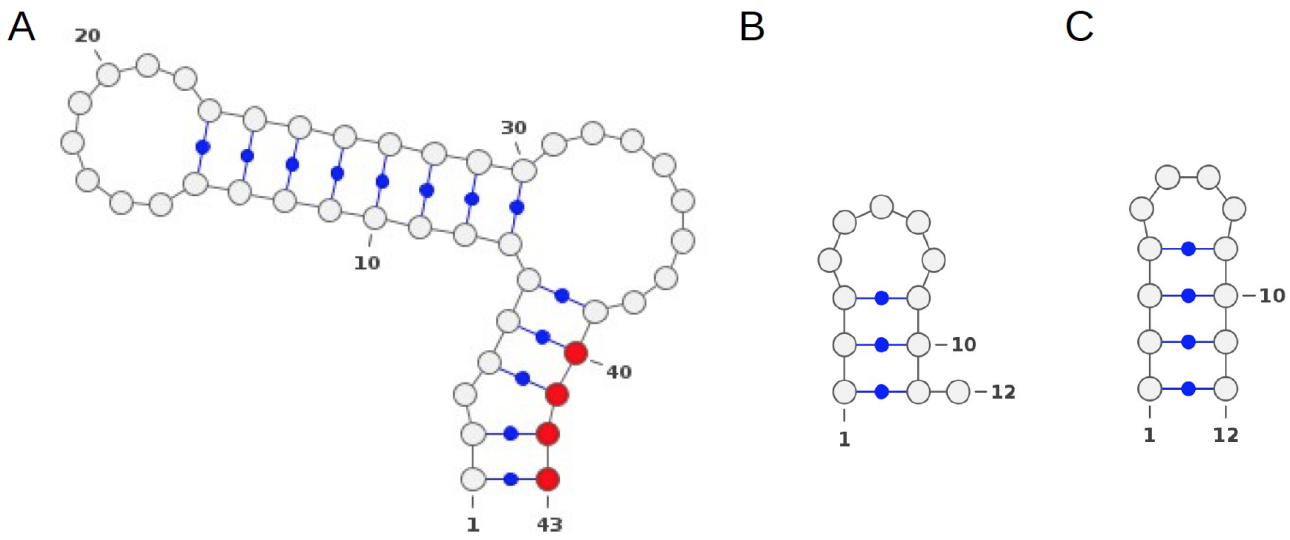


Figure 3.23: Structures prédites dans les 5'UTR des gènes plus exprimés en présence de DCM.
Les structures sont prédites par le logiciel BEAM (Pietrosanto *et al.*, 2018). Les structures A, B et C correspondent à celles décrites dans le tableau n°4.

Des structures secondaires ont également pu être prédites dans les 5'UTR des gènes plus exprimés en présence de DCM à l'aide du logiciel BEAM (Fig. 23). Certaines de ces structures ne sont pas spécifiques aux gènes plus exprimés en présence de DCM car elles sont aussi retrouvées dans des 5'UTR de gènes non régulés. Les séquences Shine-Dalgarno permettant la fixation du ribosome sur l'ARNm ne sont pas retrouvées dans ces structures, suggérant que ces structures n'auraient pas de fonction dans la régulation de la traduction.

Tableau n°4 : Structure secondaire prédite dans les ARNm plus exprimés en présence de DCM

Motif	Gène	Début [a]	Fin [a]	Score	Taille du 5'UTR	Séquence	Structure
A	METDI3043-2[b]	23	62	37,1	75	ccgcucccgcaccgcgggcugacggugcggucacccuug	((.((((((((((.....)))))))))).....)
	METDI5190	33	72	32,0	81	ucaaaagcguuggccaacccucgccgacgacuucgagga	((..((((((((((.....))))))).))).....))
B	METDI4905	12	24	8,9	212	cuucaucaaggu	(((.....))).
	METDI3863	53	65	8,6	105	gccauuccggcc	(((.....)))
	METDI4190	45	57	8,6	152	gccgguuuggcc	(((.....)))
	METDI1597	3	15	8,6	18	uuuggggguggau	(((.....)))
	METDI2115	3	15	8,5	45	gccgacguggcg	(((.....)))
	METDI3369	9	21	8,5	50	ucgggagacgga	(((.....)))
	METDI5146	22	34	8,5	42	agcgggaggcuu	(((.....)))
	METDI0008	7	19	8,3	64	cgcagagagcgg	(((.....)))
	METDI3043-2	4	16	8,3	75	ggggccaacccg	(((.....)))
	METDI4741-2	62	74	8,2	88	gcucgcacagcu	(((.....))).
	METDI1934	14	26	8,1	103	ggccagccgcca	(((.....))).
	METDI1765	2	14	8,1	26	ggcacgacgcgu	.((.....)).)
	METDI5067	6	18	8,0	88	acgcacaccggc	.((.....)).)
	METDI5875	31	43	5,2	83	gcauuaaauugc	(((.....)))
	METDI3426	8	20	5,2	38	cggccucauccg	(((.....)))
	METDI4584	12	24	4,8	51	ugugccccgaca	(((.....)))
C	METDI3067	33	44	8,4	44	ggaguccuccg	(((....)))
	METDI3863	4	15	8,4	105	cucgagggagg	(((....)))
	METDI0091	1	12	8,4	19	cucgagggagu	(((....)))
	METDI0008	26	37	8,4	64	ccgcucacggg	(((....)))
	METDI3571	43	54	8,4	88	ggccggggcuu	(((....)))
	METDI1934	64	75	8,4	103	ugaacacucgu	(((....)))
	METDI0108	35	46	8,4	49	agaggaaucug	(((....)))
	METDI4905	193	204	8,4	212	gucgggagacg	(((....)))
	METDI4741	37	48	6,9	88	gcccaugggcc	(((....)))
	METDI4699	11	22	6,9	246	cucugcggagg	(((....)))(
	METDI2655	63	74	6,3	158	uucugauggau	(((....)))
	METDI2461	19	30	6,3	62	uggaucgccag	(((....)))
	METDI5875	65	76	6,3	83	uccucagggag	(((....)))

[a] La position de la structure correspond à celle sur un fragment commençant 250 nt avant le TSS.
[b] Un chiffre est ajouté après le nom d'un gène lorsque deux promoteurs ont été prédits pour celui-ci.

L'identification de TSS a permis de réannoter certains gènes. Ceci a été le cas pour le facteur de transcription METDI1224 (NcrB-*like regulator* ; famille DUF156) plus exprimé en présence de DCM (Fig. 24). Ce TSS se trouvait à l'intérieur du gène, plus particulièrement

au niveau d'un codon ATG. L'analyse de la séquence protéique par un BlastP indique que la majorité des protéines commencent au niveau du codon ATG détecté par cette approche plutôt que le codon CTC prédit initialement lors de l'annotation. Cette protéine n'a pas été détectée par protéomique ce qui empêche de confirmer son site d'initiation de la traduction *via* un peptide.

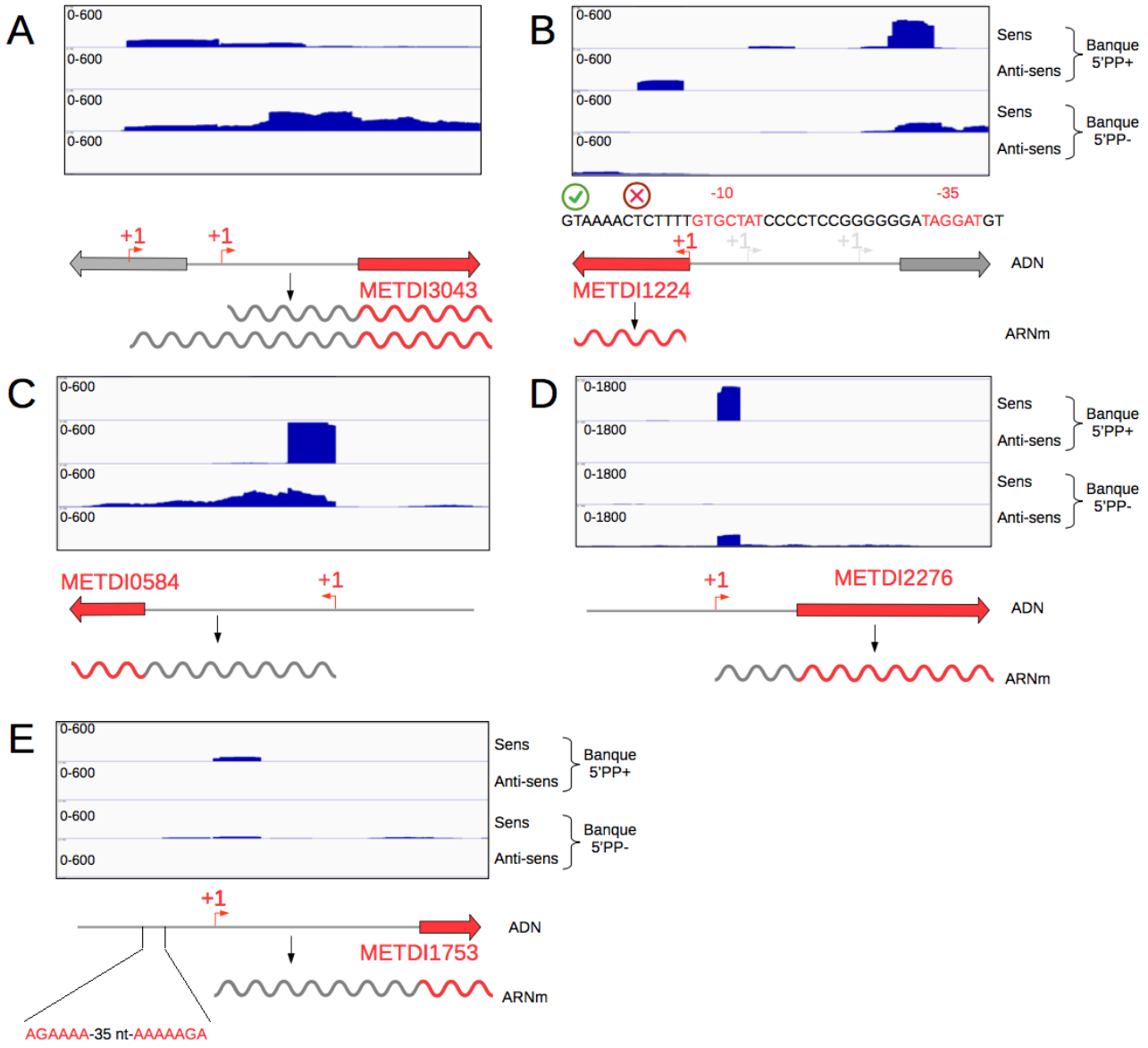


Figure 3.24: Exemples de TSS détectés.
Capture d'écran du logiciel IGV représentant les lectures alignées sur le génome. L'échelle est indiquée dans le coin gauche de chaque piste. A. METDI3043, exemple de gène ayant deux promoteurs putatifs. B. METDI1224 qui est réannoté via l'approche de TSS. Ce gène coderait un ARN *leaderless*. C. METDI0584 possédant un *riboswitch* au TPP. D. METDI2276 possédant un *riboswitch* au TPP. E. Motif AGAAA potentiellement impliqué dans la régulation en présence de méthanol retrouvé en amont de METDI1753

Les séquences 5' UTR des ARNm peuvent être utilisées pour la régulation de la transcription ou de la traduction. Des *riboswitchs* peuvent être trouvés ainsi que des sites de fixation d'ARN régulateurs. La taille des *riboswitchs* connus varie de 35 à plus de 200 nucléotides (Serganov & Patel, 2007). Cela est compatible avec la taille des séquences 5'UTR des ARN primaires proposées par le logiciel TSSAR. Parmi les 5'UTR confirmés au niveau du site d'initiation de la transcription et de la traduction, la banque de données Rfam prédit deux *riboswitchs* : un *riboswitch* potentiel détectant le thiamine pyrophosphate (TPP) trouvé pour les gènes *thiO* (METDI0584 annoté comme *glycine oxidase*) et *thiC* (METDI2276 annoté comme *thiamine biosynthesis protein*). Ces deux gènes ne sont pas différentiellement exprimés dans les conditions testées.

3.6.4 Conclusions de la partie TSS-seq

L'analyse des TSS a permis d'avoir une meilleure vue de la transcription chez *M. extorquens* DM4 ainsi que d'améliorer l'annotation du génome de *M. extorquens* DM4. Cette étude servira de socle dans l'étude de l'expression de nouveaux gènes d'intérêt dans la réponse au DCM tel que METDI3043 (Fig. 21 et 24). Ces résultats pourront être confirmés, en partie, par d'autres méthodes comme le 5'RACE ou des RT-PCR. L'analyse des sites d'initiation de la transcription n'a pas permis de trouver de nouveaux motifs ou structures impliqués dans la régulation des gènes en présence de DCM. Cependant, l'extrémité 5' des ARNm n'est pas la seule à être impliquée dans des mécanismes de régulation. En effet, des études récentes mettent en évidence des mécanismes impliquant la région 3'UTR, qui peuvent influencer la stabilité et la traduction de ces ARNm (Ren *et al.*, 2017 ; El Mouali & Balsalobre, 2018). Comme l'extrémité 5', de nouvelles stratégies expérimentales de transcriptomique permettent de cibler l'extrémité 3' des ARNm (Dar & Sorek, 2018). Une telle approche permettrait de mettre en évidence les séquences de terminaisons de la transcription, facilitant ainsi la détection d'opérons.

3.7 Conclusions générales sur les approches « -omiques »

L'ensemble combiné de ces résultats a permis de détecter plusieurs motifs au niveau nucléique ou protéique permettant d'obtenir un « portrait-robot » d'un gène chez *M. extorquens* DM4 (Fig. 25). Ce « portait » permet d'améliorer l'annotation du génome. Néanmoins, aucun motif de régulation n'a été trouvé pour les gènes plus régulés en présence de DCM.

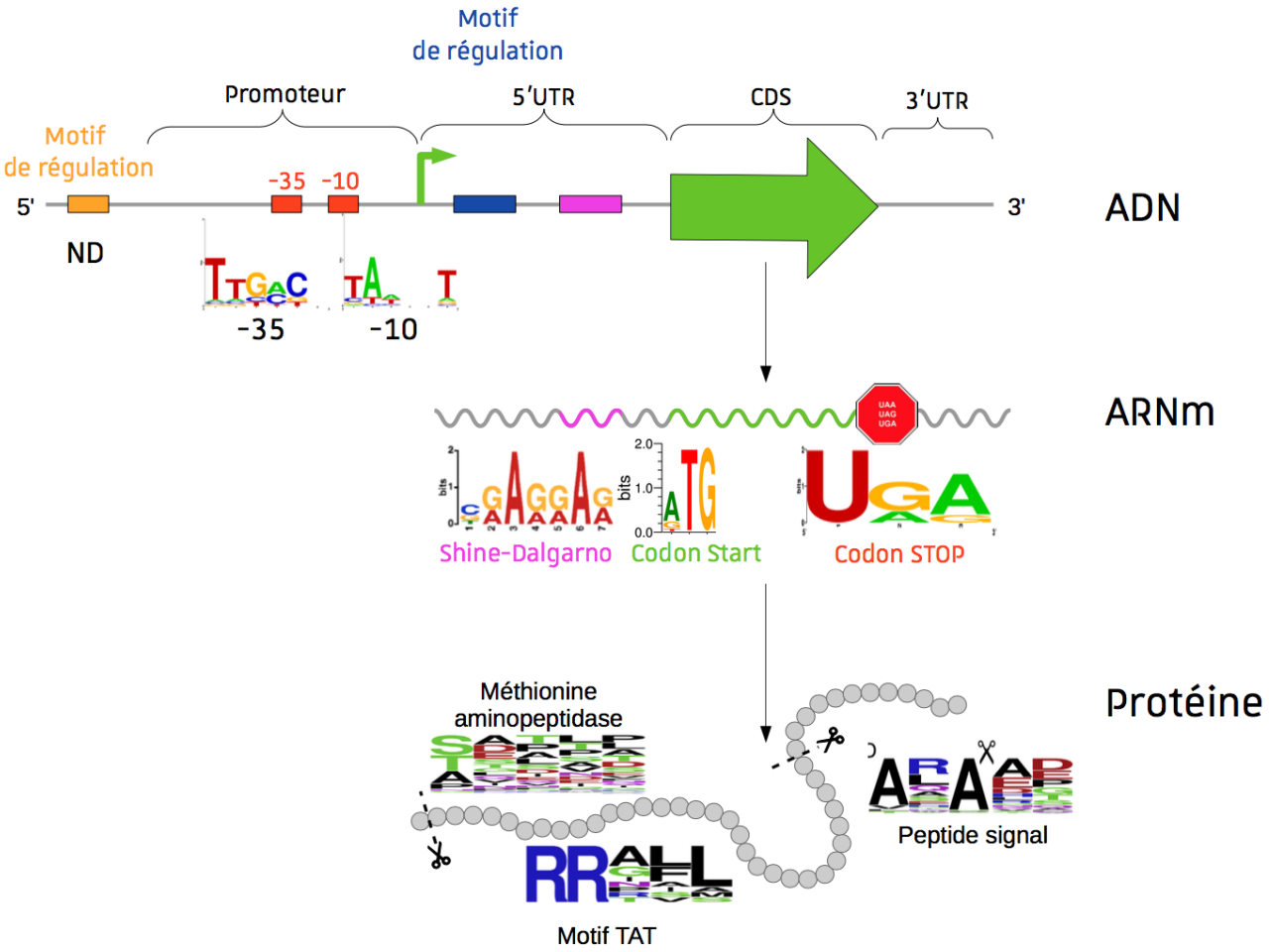


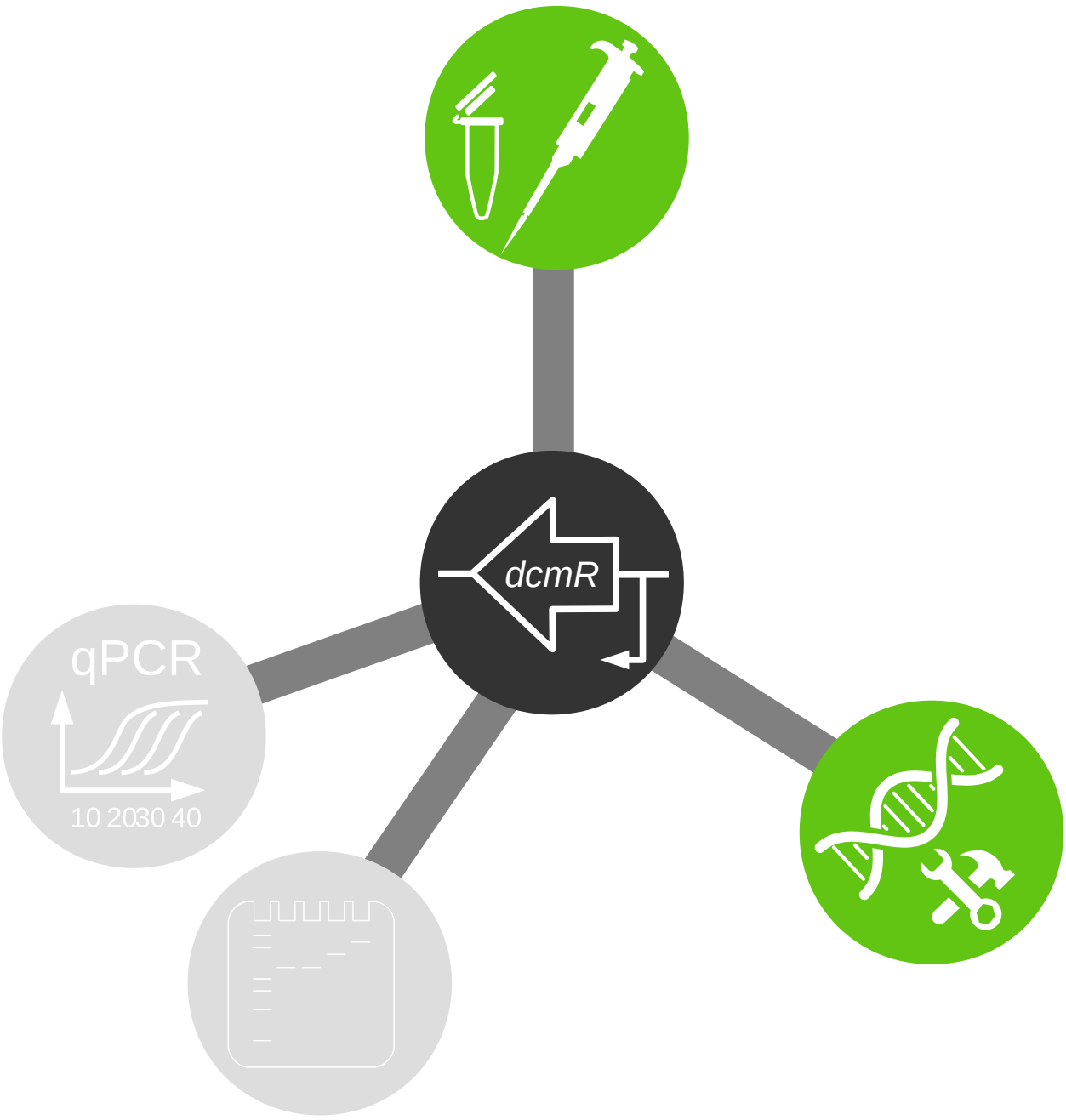
Figure 3.25: « Portrait robot » d'un gène chez *M. extorquens* DM4.
Les approches « -omiques » ainsi que les informations disponibles sur la plateforme MaGe du Genoscope permettent d'avoir une vision d'ensemble des motifs présents dans un gène chez *M. extorquens* DM4.

Ces résultats obtenu avec plusieurs approches « -omiques » indiquent que la réponse au DCM n'est pas limitée aux gènes *dcmRABC* mais repose sur l'expression différentielle de nombreux gènes. Une partie de cette réponse est transcriptionnelle, mais aussi post-transcriptionnelle ou post-

traductionnelle (Chaignaud *et al.*, 2017 ; Bibi-Triki *et al.*, 2018). L'absence de séquences de régulation commune à tous les gènes plus exprimés en présence de DCM suggère qu'ils ne sont pas régulés d'une façon similaire et pourraient reposer sur des réponses à des signaux cellulaires différents tels que l'acidification du cytoplasme, des dommages à l'ADN ou encore à des variations de cofacteurs cellulaires. Des analyses de fusions transcriptionnelles ou traductionnelles permettront de confirmer l'importance biologique des motifs trouvés. L'analyse de ces résultats permet également de mieux comprendre la physiologie de *M. extorquens* DM4 en déterminant les motifs impliqués dans la transcription et la traduction (Fig. 25). La transcription de gènes *leaderless* ou la détection de modifications post-traductionnelles telle que l'excision de méthionine permettent d'avoir une meilleure vision des mécanismes cellulaires présents.

Les expériences « -omiques » réalisées dans seulement deux conditions de culture ne suffisent pas pour déterminer si ces différences d'abondance sont liées à une activation ou une répression. Pour cela, ces expériences pourraient être réalisées dans d'autres conditions de culture, par exemple de croissance en présence de succinate ou de méthylamine comme source de carbone. La méthylamine est un substrat méthylotrophe dont le métabolisme ne produit pas de formaldéhyde (Vuilleumier *et al.*, 2009). L'expression des gènes en condition de phase stationnaire et non en phase exponentielle, pourrait fournir une meilleure compréhension de la physiologie en présence de DCM. Des cultures en présence de plusieurs substrats pourraient être aussi réalisées afin de s'approcher davantage de conditions environnementales où plusieurs substrats de croissance sont disponibles. Cela permettra de mieux comprendre comment est régulé le gène de *M. extorquens* DM4 dans des conditions se rapprochant de celles appliquées dans des stratégies de bioréhabilitation.

4. Régulation transcriptionnelle de l'îlot *dcm*



Chapitre 4 : Régulation transcriptionnelle de l'îlot *dcm*

4.1 Présentation du chapitre

Ce chapitre aborde la régulation du transposon catabolique *dcm* par le gène *dcmR*. Il est écrit sous la forme d'un article scientifique. Cette étude est basée sur l'analyse de plusieurs mutants des gènes *dcm* qui ont été construits dans le cadre de cette étude (Fig. 1). Pour déterminer le rôle de ces gènes dans la régulation de l'îlot, l'activité promotrice et l'abondance des ARNm et protéines de l'îlot *dcm* chez la souche sauvage et plusieurs mutants a été mesurée. Seul les mutants du gène *dcmR* sont abordés dans ce chapitre. Des résultats qui ne seront pas présentés dans cet articles sont inclus à la fin du chapitre.

4.2 Matériel et méthode (précisions par rapport à l'article)

4.2.1 Quantification des protéines par SRM

Une quantification des protéines DcmRABC a été réalisée par Gautier Husson et Christine Carapito de l'IPHC de Strasbourg. Pour cela une approche de SRM (*selected reaction monitoring*) a été mise en place (Fig. 2). Elle consiste à comparer par spectrométrie de masse le ratio entre un peptide de concentration inconnue et son équivalent de concentration connue et marqué avec un isotope lourd. Pour choisir les peptides marqués ceux obtenus lors de la protéolyse des protéines DcmRABC sont étudiés. Les critères de choix ont été que ces peptides soient conservés chez les autres souches dont on possède la séquence de l'îlot *dcm* (Firsova *et al.*, 2010 ; Brown *et al.*, 2011 ; Bringel *et al.*, 2017). Ils ont également été choisis ne contenant pas de méthionine (M) ou cystéines (C) pour éviter les modifications post-traductionnelles qui pourraient fausser la détection. La trypsine coupe après les lysines (K) et arginine (R) mais pas si elles sont suivies d'une proline (P). Clivage manquées. Une fois les peptides choisis ils sont synthétisé par ThermoFischer avec un isotope lourd au niveau de l'acide aminé en C terminal. Ces peptides marqués sont ajoutés en concentration connue dans un extrait protéique global.

Les protéines extraites sont purifiées par un *gel stacking* puis elles sont ensuite alkylées et réduites pour favoriser la digestion avec de la trypsine. Les peptide sont séparés par séparation par en chromatographie liquide en fonction de leur hydrophobicité. Cette séparation augmente la sensibilité des données. Les peptides sont ensuite ionisés par ESI (*Electron spray ionization*) qui est une méthode douce. Les ions obtenus sont séparés par un quadrupôle en fonction de leur rapport masse sur charge (m/z). Un second quadrupôle (cellule de collision) permet de fragmenter un peptide via une interaction avec un gaz comme l'argon puis un troisième permet de détecter un ion précurseur / fragment (appelé transition). Une optimisation des paramètres de manipulations a été établis sur une échantillon appelé matrice.

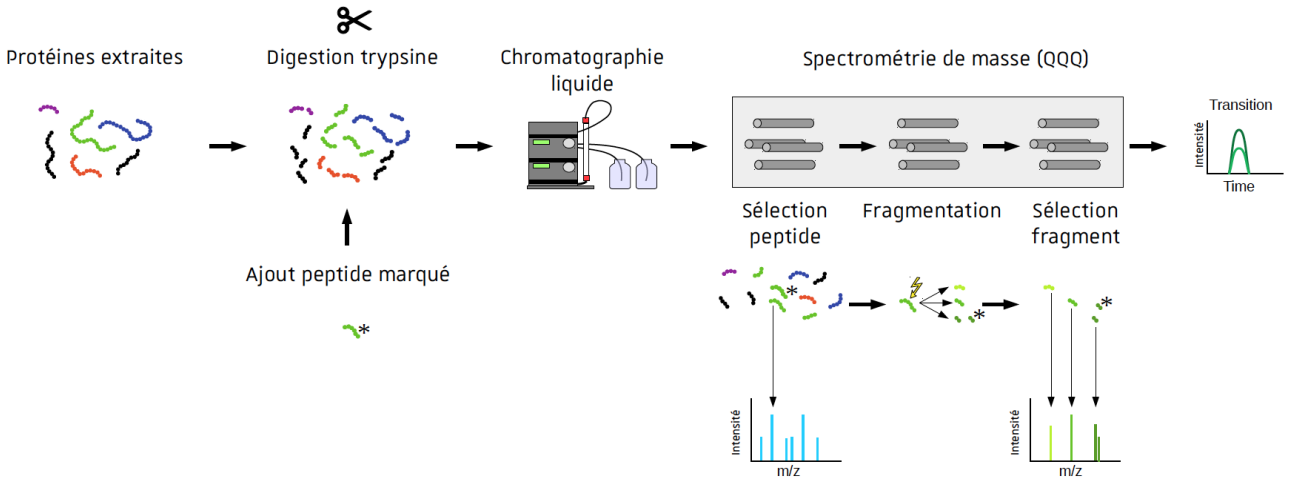


Figure 4.1: Principe de la quantification de protéine par SRM

Revisiting *dcmR*, a key regulator of bacterial dichloromethane dehalogenation

Bruno Maucourt [a], Sabrina Bibi-Triki [a], Gauthier Husson [b], Yousra Louhichi [a], Stéphane Vuilleumier [a], Christine Carapito [b] & Françoise Bringel [a*]

[a] Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France

[b] Université de Strasbourg, CNRS, IPHC UMR 7178, Laboratoire de Spectrométrie de Masse BioOrganique, Strasbourg, France

*Corresponding author: Françoise Bringel; Address: Université de Strasbourg, CNRS, GMGM UMR 7156, 28 rue Goethe, 67000 Strasbourg, France
Email: francoise.bringel@unistra.fr; phone: + 33 (0)3 68 85 18 15; fax: + 33 (0)3 68 85 20 28

Running title: Regulation of dichloromethane utilization

Key words: dehalogenation, dichloromethane, gene expression, transcriptional regulator

Abbreviations: DCM, dichloromethane; SRM, selected reaction monitoring-based proteomics; WT, wild-type

Abstract

Dichloromethane (DCM) is the most highly produced chlorinated solvent worldwide. Release of this toxic compound to the environment due to industrial activities calls for efficient monitoring and bioremediation strategies, and thus a better understanding of the expression of genes involved in bacterial degradation of DCM. In DCM-degrading bacteria growing with DCM as sole source of carbon and energy characterized so far, the essential DCM dehalogenase is encoded by the *dcmA* gene within the *dcm* catabolic transposon. The previously identified transcriptional regulator gene *dcmR* is in divergent orientation directly upstream of *dcmABC* genes, with genes *dcmB* and *dcmC* of yet unknown function. Here, *dcmR* mutants were constructed in the reference strain *Methylorubrum* (formerly *Methylobacterium*) *extorquens* DM4 in order to better understand transcriptional regulation mediated by DcmR. Transcriptional fusions with *dcmR* (P_R) and *dcmA* (P_A) promoter regions were also constructed and analyzed in wild-type and *dcmR* mutant strains. During growth of. *M. extorquens* in the absence of DCM, DcmR was shown to inhibit its own transcription through one of its two identified promoters, as well as *dcmABC* expression from the P_A promoter, shown here to be expressed as an operon, and likely also expression of other genes located outside of the *dcm* catabolic transposon, notably METDI3043 with predicted potential roles in signalling and gene regulation. For *M. extorquens* grown with DCM as the sole source of carbon, absence of *dcmR* had little effect on P_A activity, and did not affect transcript and protein abundance of *dcm* genes. Thus, an uncharacterized *dcmR*-independent process appears to induce *dcmA* expression in *M. extorquens* growing with DCM. In addition, mutational inactivation of *dcmR*, or of its expression, led to longer generation times for growth of *M. extorquens* with DCM. This suggests a distinct role of *dcmR* in addition to regulation of expression of *dcm* genes. Taken together, our results provide evidence that DCM-dependent regulation is more complex than previously suggested, and extends beyond the repressor role of DcmR on DCM dehalogenase expression in the absence of DCM.

1. Introduction

Dichloromethane (DCM, CH_2Cl_2, methylene chloride) is a chlorinated solvent massively used in industry and thus often found as a contaminant in the environment (Tsai, 2017). This molecule is

toxic for living organisms, and has been classified as a potential carcinogen in humans (Schlosser *et al.*, 2015). Its photolysis in the stratosphere is responsible for ongoing ozone layer depletion despite the 1987 ban on chlorofluorocarbon utilization (Hossaini *et al.*, 2017). To remove DCM at polluted sites, bioremediation approaches involving DCM-degrading bacteria appears both sustainable and cost-effective (Shestakova and Sillanpää, 2013). The Alphaproteobacterium *Methylorubrum extorquens* DM4 (previously *Methylobacterium extorquens* DM4; Green and Ardley, 2018) is the reference strain in investigations of DCM utilization (Muller *et al.*, 2011a), and features the only degradation pathway for DCM characterized in detail, which involves DCM dehalogenase/glutathione S-transferase encoded by the *dcmA* gene (Fig. 1A).

The *dcm* catabolic transposon is defined by four genes *dcmRABC* (the *dcm* islet) flanked by insertion elements (Fig. 1B). This transposon was discovered in strain DM4 (La Roche and Leisinger, 1990-1991) and subsequently found in most other aerobic DCM-degraders as part of a conserved gene cluster flanked with insertion elements (IS) (Schmid-Appert *et al.*, 1997). Although transposon activity has not yet been experimentally demonstrated, the presence of insertion elements and associated transposases, as well as the extremely high sequence conservation of *dcmRABC* genes in DCM-degrading strains characterised so far, strongly suggests that this transposon is active in the dissemination of the ability to degrade DCM in the bacterial domain. The gene *dcmR* is monocistronic and oriented in divergent orientation upstream of genes *dcmABC*. DCM dehalogenase (encoded by *dcmA*) is essential for growth with DCM (La Roche and Leisinger, 1990) while *dcmB* and *dcmC* are still of unknown function. DcmR was proposed to be a repressor of *dcmA* expression, on the evidence that its absence led to constitutive *dcmA* expression, whereas its presence was required to reduce *dcmA* expression in cultures of strain DM4 in the absence of DCM (La Roche and Leisinger, 1991). DcmR was also shown to regulate its own expression (La Roche and Leisinger, 1991), and promoters of *dcmA* (P_A) and *dcmR* (P_{R1} and P_{R2}) were experimentally validated by S1 nuclease mapping (La Roche and Leisinger, 1991). DcmR consists of two conserved protein domains. A helix-turn-helix (HTH) DNA-binding domain shows similarity with the HTH domain of MerR-type regulators (Couñago *et al.*, 2018). Both promoters P_{R1} and P_A share a conserved 12 bp-long sequence (Fig. 1C), predicted to represent a DcmR-binding site, although this sequence differs from usual inverted repeats found upstream of regulated genes (La Roche and Leisinger, 1991). The ability of DcmR to recognize this putative DNA-binding box was suggested to depend on DcmR binding to a ligand. Indeed, DcmR also features a second protein

domain called MEDS for MEthanogen/methylotroph, DcmR Sensory (Anantharaman and Aravind, 2005), with predicted roles in recognition of small hydrocarbon compounds. So far, no ligand of the MEDS domain has been identified experimentally. In strain DM4, promoter P_A activity was not induced by chlorinated solvents other than DCM and dihalogenated methanes (Lopes *et al.*, 2012) (reviewed in Table S1), suggesting that they may represent specific ligands of the DcmR MEDS domain.

Global differential transcriptomics and proteomics of strain DM4 grown with DCM or with methanol recently showed that *dcm* genes were induced and among the most expressed genes at both mRNA and protein levels in DCM cultures (Chaignaud *et al.*, 2017; Bibi-Triki *et al.*, 2018). Indeed, DcmA was shown to represent 25% of total protein in cell extracts of strain DM4 grown with DCM (Muller *et al.*, 2011b). However, a repressor role of DcmR in cultures grown in the absence of DCM was clearly challenged when DcmR failed to be detected in protein extracts of *M. extorquens* DM4 grown with methanol (Muller *et al*., 2011b). Thus, the goal of the present study was to clarify the role of DcmR, and especially its role in DCM-dependent regulation of *dcm* genes.

2. Materials and methods

2.1. Bacterial strains, chemicals and growth conditions

Strains and plasmids used in this study are listed in Table S2. All chemicals were purchased from Sigma-Aldrich unless otherwise noted. *Escherichia coli* strains were grown at 37°C in LB medium, and *M. extorquens* at 30°C in *Methylobacterium* M3 minimal medium (Muller *et al*., 2011b), with carbon sources supplied at 10 mM carbon (for 220 mL cultures, 89 µL neat methanol, 141 µL neat DCM, or 1.4 mL sterile-filtered 400 mM succinate). Antibiotics kanamycin and ampicillin were at final concentrations of 50 and 100 $\mu g.\mu L^{-1}$, respectively. Aerobic liquid cultures (220 mL) were performed under shaking at 120 rpm, and in 1 L Erlenmeyer flasks, hermetically sealed (Supelco Mininert caps, Sigma) for DCM cultures, as described previously (Bibi-Triki *et al.*, 2018). Liquid culture volumes were reduced to 80 mL to avoid flocculation in cultures grown with a mixture of methanol and DCM. Transcriptional fusion assays were performed with 25 mL liquid cultures of *M. extorquens* DM4 in 250 mL Erlenmeyer flasks. Growth was followed by measuring OD_{600nm}

(UVIKON spectrophotometer).

2.2. Construction of plasmids and mutant strains

Plasmid isolation, *E. coli* transformation, restriction enzyme digestion and ligation were carried out by standard protocols (Sambrook, 2001). High fidelity DNA polymerase (Phusion polymerase; Thermo Fischer) was used for PCR amplification. Plasmids for transcriptional fusions were constructed using the pLM-sYFP2 vector (Kaczmarczyk *et al.*, 2011). To test *dcmR* and *dcmA* promoter activity, inserts of complete or truncated fragments of the intergenic region between *dcmR* and *dcmA* were cloned using KpnI and SacI restriction sites in forward or reverse orientation upstream of the *yfp* gene of plasmid pLM-sYFP2 (Table S2). Cloned inserts were validated by DNA sequencing. Plasmids pME8701-pME8704 and pME8324 were constructed using PCR products as inserts, and plasmids pME8312 to pME8320 were synthesized by a commercial supplier (e-Zyvec, Lille, France). Plasmids were introduced into *Methylorubrum* strains by overnight tri-parental mating at 30°C as described previously (Fulton *et al.*, 1984), and into *E. coli* by transformation. The presence of plasmids in transconjugants and transformants was validated by PCR of plasmid-borne kan^R gene and cloned insert sequences. Mutants of *M. extorquens* DM4 were constructed by unmarked chromosomal allelic exchange as described previously (Gruffaz *et al.*, 2014).

2.3. Nucleic acid extraction

Chromosomal and plasmidic DNA was extracted from liquid cultures using the Master pure kit (Epicentre) according to the manufacturer’s protocol. RNA was extracted using the Nucleospin RNA plus kit (Macherey-Nagel), followed by treatment with Turbo DNase and Turbo DNA-free kit (Invitrogen) as previously described (Chaignaud *et al.*, 2017).

2.4. Measurements of promoter activity

Promoter activity in derivatives of the reporter plasmid pLM-sYFP*2* was measured in both homologous (*M. extorquens* DM4) and heterologous (*E. coli*) backgrounds. Liquid cultures of *M. extorquens* DM4 were sampled every hour following inoculation of a preculture (starting at a

final OD_{600nm} of 0.04), and duplicate cell suspension samples (200 µL) were transferred to 96-well microplates. Fluorescence was measured at 516 nm following excitation at 485 nm in a microplate reader (Synergy HT-1 BioTek). Cultivation of *E. coli* derivatives was directly performed in 96-well microplates, using 200 µL M9 medium per well (Gibco BRL), supplemented with 0.2% casamino acids (BD Biosciences) and appropriate antibiotic supplementation, and under agitation (Tecan microplate reader Infinite F200pro with orbital rotation at 44 rpm at 37°C). Every 30 minutes (starting at OD_{600nm} 0.01), cellular density was measured by OD_{600nm}, and fluorescence at 535 nm (485 nm excitation). To determine promoter activity, OD_{600nm} and fluorescence values were calculated by subtracting values of blank (non-inoculated medium) from the values of experimental samples. For each replicate, the stabilized, corrected, and normalized fluorescence value was obtained by dividing by the corrected OD_{600nm} value (after 7 hours in exponential phase for *M. extorquens* DM4, and after 2 hours for *E. coli*). For *M. extorquens* DM4, mean normalized values were expressed relative to corresponding values from control cultures of wild-type strain DM4 carrying plasmid pME8071 grown with methanol in parallel.

2.5. qPCR, RT-qPCR and RT-PCR

qPCR and RT-qPCR experiments were performed in qPCR 96-well plates (Agilent Technologies) sealed with Microseal B (Bio-Rad) using the Stratagene Mx3005P Thermocycler (MxPro software v4.10; Agilent Technologies). DNA-depleted RNA (2.0 µg) was used for reverse transcription with Superscript III (Invitrogen) and random hexanucleotide primers (Invitrogen) following the manufacturer's protocol, except that RNaseOUT™ was replaced by Rnasin™ (40 $U.\mu L^{-1}$; Invitrogen). Controls included RT reactions without RNA template or Superscript III. Transcript quantification by qPCR was subsequently performed using 3.75 µL template cDNA, 7.5 µL Brilliant III SYBR™ Green low ROX qPCR master mix (Agilent Technologies), 0.225 µL of each primer at 20 µM (Eurofins, primer list in Table 3), and 3.3 µL molecular biology grade water. After 3 min pre-denaturation at 95°C, 40 cycles of 20 s at 95°C and 20 s at 60°C, a final one-cycle step of 1 min at 95°C with ramping from 60 to 95°C was performed to obtain dissociation curves for quality control of PCR products. SYBR™ green fluorescence data normalized with ROX fluorescence were analyzed with the R package qpcR (v1.40) for Ct determination (Ritz and Spiess, 2008). Linearized pLM-sYFP2 plasmid was used as a standard for determination of the qPCR

reference curve.

Plasmid stability in *M. extorquens* DM4 was estimated by qPCR from DNA extracted from stationary phase cultures (biological triplicates for each sample). Dilutions of extracted DNA from 0.1 to 0.01 $ng \cdot \mu L^{-1}$ were used to quantify the plasmid-borne *kanF* gene and the chromosome-borne *gmaS* gene using standard curves with linearized plasmid DNA (10^{-1} to 10^{-5} $ng \cdot \mu L^{-1}$) of pLM-sYFP2 and pME8280 (Gruffaz *et al.*, 2014), respectively. A template no-DNA control was also performed. Plasmid copy number was estimated as previously described (Škulj *et al*., 2008).

Fold-change values were calculated by the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008), with the *rrsA* gene used as internal control, and normalisation to *dcmA* in DM4 grown with methanol as the reference condition, as previously described (Chaignaud *et al*., 2017).

Co-expression of *dcmABC* genes was assessed on RNA from WT strain cultures grown with DCM or methanol extracted essentially as described above. Retrotranscription was performed on 100 ng RNA using Superscript III (Invitrogen) and specific primers (Fig. S1, see Table S2 for primer sequences) following the manufacturer's recommendations. The resulting cDNA was PCR-amplified using primer pairs (Table S2) hybridizing at different localizations within the catabolic *dcm* transposon (Fig. S1), as described above.

2.6. Quantification of Dcm proteins

Proteins were obtained from mid-log phase cultures (45 mL). After centrifugation (10 min at 6,500 g at 4°C), cell pellets were resuspended in 800 µL pure water, lysed at maximum agitation speed by 3 consecutive 1 min cycles with glass beads (0.75 g; average size 0.1 mm; BioSpec Products) using Retsch MM2 Pulverizer Mixer Mill (Hamilton Instrument), followed by 1 min on ice. After centrifugation (10 min at 8,600 g at 4°C), proteins in the supernatant were precipitated by addition of four volumes of cold acetone, and stored at -25°C until further analysis.

The general experimental workflow of the Selected Reaction Monitoring (SRM)-based proteomic approach, as well as the selection and optimization of targeted peptide transitions, were as previously described (Hovasse *et al.*, 2016). DnaA and RpoA were selected as constitutively expressed control proteins. Selection of targeted peptides to quantify DcmRABC proteins was performed manually, based on peptides obtained after trypsin degradation of Dcm proteins and sequence analysis of conserved regions among 3 DCM-degraders of known genome sequence

M. extorquens DM4 (Vuilleumier *et al*., 2009), *Hyphomicrobium* sp. GJ21 (Bringel *et al*., 2017) and *Hyphomicrobium denitrificans* ATCC 51888 (Brown *et al*., 2011). Selected peptides (Table S4) with C-terminal heavy lysine or arginine were synthesized by a commercial provider (PEPotec SRM Custom Peptide Libraries grade 3, Thermo Fischer).

2.7. Statistical analysis

The Kruskal Wallis and Dunn post-hoc tests were performed with with the dunn.test package within R (R Core Team, 2017).

3. Results and discussion

Although the DcmR protein has been proposed to regulate the expression of its own gene through the *dcmR* promoter, as well as that of the DCM dehalogenase through the corresponding *dcmA* promoter (La Roche & Leisinger, 1991), direct evidence for this has not been obtained so far. We chose to reinvestigate the role of *dcmR* on expression of DCM-regulated genes by constructing new chromosomal mutants of strain DM4 with site-directed mutations of *dcmR*, alone or in combination with deletion of adjacent *dcm* genes (Fig. 1B). Mutant strains were compared to WT with regard to their growth abilities, mRNA and protein levels of *dcm* genes, as well as by the measurement of *dcmR* and *dcmA* promoter activities using plasmid-encoded gene fusions. Hence, transcriptional gene fusions provided information about regulation of transcription initiation activity, whereas direct quantification of transcript abundance by RT-qPCR, and of protein abundance by selected reaction monitoring-based proteomics (SRM), allowed us to explore potential regulation processes.

Accordingly, a STOP*dcmR* chromosomal mutation was designed to prevent translation of the DcmR protein, while maintaining essentially intact the DNA sequence upstream of the *dcmR* gene, as well as that of the *dcmR* gene itself. Such an approach was crucial in our view since previous investigations are considered to have been obscured by the use of plasmid-based approaches, as well as by the potential requirement of the full-length DNA sequence of the *dcm* catabolic transposon in regulation of *dcm* genes (La Roche & Leisinger, 1991; Schmid-Appert, 1996; Schmid-Appert *et al.*, 1997). To this end, the *dcmR* start codon was replaced by several stop codons

in all three translation frames (Table. S2), and the resulting STOP*dcmR* mutation was introduced by allelic gene replacement at the *dcm* chromosomal genome locus. Other investigated mutations included deletion of *dcmA* to abolish DCM degradation (Δ*dcmA*), of genes *dcmB* and *dcmC* (Δ*dcmBC*), and deletion of the entire *dcm* catabolic transposon (Δ*dcm*), in order to assess whether DCM or its degradation products may be involved in DcmR-mediated regulation. The phenotypes of mutant strains were assessed by growth in M3 mineral medium with different carbon and energy sources DCM, methanol as alternative C_1 substrate, succinate as a multi-carbon substrate, and an equimolar mixture of DCM and methanol, a condition also known to be conducive to expression of DCM dehalogenase.

3.1. Impaired growth of strain DM4 lacking DcmR with DCM

According to the prevailing model of the role of *dcmR* as a central regulator of DCM-dependent gene expression, inactivation of *dcmR* will alter growth of strain DM4 with DCM by impairing expression of DcmA, the only essential protein for DCM degradation (La Roche & Leisinger, 1991), or alternatively, by interfering with other functions required for adaptive growth with DCM (Bibi-Triki *et al*., 2018; Muller *et al*., 2011b). All constructed mutants of strain DM4 grew with methanol or succinate with no significant difference compared to the WT strain. As expected, the absence of *dcmA* prevented growth with DCM (mutants Δ*dcmA* and Δ*dcm*) (Fig. 2). In addition, similar generation times were observed for mutant Δ*dcmA* compared to the WT strain in cultures provided with both DCM and methanol, suggesting that DCM *per se* does not inhibit growth under the tested experimental conditions. Strikingly, however, the STOP*dcmR* mutant showed a significantly longer generation time for growth with DCM (12.1 ± 1.2 h) than WT, suggesting that *dcmR* is required for optimal growth during DCM utilization. Lack of *dcmBC* genes had no additional detrimental effect on growth (viz. STOP*dcmR*Δ*dcmBC* as compared to mutant STOP*dcmR*, Fig. 2). When DCM was provided together with methanol as the source of carbon and energy for growth, moreover, only the STOP*dcmR* mutant displayed a growth phenotype (5.1 ± 0.4 h) compared to the WT strain (3.6 ± 0.6 h) and other mutants (STOP*dcmR*Δ*dcmBC*, Δ*dcmA* and Δ*dcm* mutants). Thus, the absence of DcmR specifically alters growth of strain DM4 with DCM irrespectively of the presence of methanol, suggesting that regulation of *dcm* gene expression by *dcmR* is required for optimal growth in the presence of DCM.

3.2. New insights on the repression of *dcm* genes expression in absence of DCM

The regulatory role of *dcmR*, and of its expressed transcript and protein product DcmR (also see Fig. 6 below), were assessed in both homologous (*M. extorquens* DM4) and heterologous (*E. coli*) backgrounds by using newly constructed plasmid-borne transcriptional fusions of P_A, P_{R1} and P_{R2} promoter sequences to a promoterless fluorescent reporter *yfp* gene (Table S2). Plasmid constructs were stable in the tested conditions in *M. extorquens* DM4, with one plasmid copy per chromosome (qPCR validation, Fig. S2).

The effect of *dcmR* expression on promoter activity of the *dcmR-dcmA* intergenic region was first explored in the heterologous genetic context of *E. coli*, which lacks the genes for DCM utilisation. Activity was detected for the *dcmA* (P_A) promoter (pME8313) but not for the *dcmR* (P_R) promoter (pME8316) (Fig. 3A), so only the P_A promoter was further investigated. When *dcmR* was also expressed, a reduced fluorescence signal was obtained (Fig. 3A, compare pME8313 with pME8320). Thus, DcmR represses the activity of P_A under the tested condition. Moreover, mutations of either or both previously identified conserved 12 bp-boxes (Fig. 3B, mutated boxes denoted as P'_A and P'_R), which were designed to prevent the proposed binding of DcmR, had no effect on P_A promoter activity, irrespectively of whether DcmR was present or not (Fig. 3A). Thus, in the heterologous background of *E. coli*, and in the absence of DCM, repression of P_A-dependent gene expression by DcmR is independent of the 12-bp box sequence in the *dcmR-dcmA* intergenic region.

We then performed similar experiments with closely related plasmids in the homologous genetic context of *M. extorquens* DM4 (Fig. 4). Basing on the deduced role of DcmR on gene expression from the P_A promoter, detected here in *E. coli* (Fig. 3), as well as on previous investigations (La Roche & Leisinger, 1991), one would expect increased *dcmA* and *dcmR* promoter activity in the STOP*dcmR* mutant grown with methanol or succinate in the absence of DCM compared to the WT strain. For both WT and STOP*dcmR* mutant, P_A and P_R showed similar activity for growth with methanol and succinate (Fig. 4A and 4B). In absence of DcmR (Fig 4B), however, activity of both P_A and P_R was significantly increased in the absence of DCM.

Taken together, these new results on the role of DcmR in the absence of DCM, obtained in both heterologous *E.coli* and homologous strain DM4 backgrounds, suggest that DcmR acts as a repressor of the *dcmA* promoter P_A, and in a way that is independent of the conserved 12 bp box previously identified as a potential DcmR-binding sequence. This renews the unresolved question of the target of DcmR-mediated repression of DcmA expression in the absence of DCM.

In this context, and in order to test if transcriptional repression of gene expression from the P_A promoter involves upstream sequences of the P_A promoter, transcriptional fusions were constructed using the entire *dcmR-dcmA* intergenic region or a short fragment of 427 bp bearing the *dcmA* promoter P_A only, cloned upstream of the promoterless *yfp* reporter gene (Fig. 4A). In wild-type DM4 grown in absence of DCM, the long and short reporter constructs resulted in similar fluorescence production (Fig. 4C, P_A). Thus, the short fragment contains the information required for repression of *dcmA* gene expression in the WT background. In *dcmR* mutants, in contrast, the larger fragment yields significantly higher promoter activity than the short fragment (long/short ratio of 4.0, Fig. 4C, P_A, and Fig. S4B). This reveals a conditional requirement (such as e.g. absence of DcmR) for the upstream region of P_A in repression of *dcmA* expression. Testing this and other alternative hypotheses will require further experiments, starting with different combinations of the plasmids and strains newly obtained in this work, such as testing pME8701 and pME8703 plasmids in Δ*dcmA* and Δ*dcm* mutant backgrounds.

Transcriptional fusion experiments in the background of strain DM4 in both WT and *dcmR* mutant contexts (Fig. 4) also allowed to define *dcmR*-related targets of DcmR in the absence of DCM. For this, transcriptional activity of the 883 bp-long entire *dcmA-dcmR* intergenic region was compared to that of a shorter 427 bp fragment which still featured both *dcmR* promoters P_{R1} and P_{R2} (Fig. 4A, top right). Fluorescence production was comparable for the two constructs (ratio ≈1), irrespectively of growth substrate or genetic context (Fig. 4C). Then, using nested primers designed to target specifically mRNA from P_{R2} but not from P_{R1}, we demonstrated that P_{R2}-dependent transcript abundance was similar under all tested conditions and in all DM4 strain derivatives, unlike transcript abundance produced from P_{R1}, which was regulated as described previously (La Roche and Leisinger, 1991) (Fig. 5). To further validate the role of P_{R1} in regulation of *dcmR* expression, a P_{R1} deletion mutation was introduced in the STOP*dcmR* mutant (Fig. S3A), with no detected

phenotypic difference except for the amount of *dcmR* transcripts (Fig. 2B and C, Fig. S3B and S3C). Thus, the P_{R2} promoter likely ensures basal constitutive *dcmR* expression, whereas P_{R1} affords *dcmR*-regulated transcription. In other words, P_{R1} but not P_{R2} is the target of DcmR-mediated auto-repression in strain DM4.

Finally, our transcriptional fusion experiments on strain DM4 in the absence of DCM also provided a new and unexpected finding. In particular, an increased level of fluorescence of P_A- or P_R-reporter fusions similar to that of the STOP*dcmR* mutant was observed in the Δ*dcmA* mutant compared to WT (Fig. 4B). This *dcmA*-dependent phenotype was unexpected, since DcmA production is not required in the absence of DCM. It is possible that DcmA may play an additional direct or indirect regulatory function besides its role in DCM degradation. An alternative, perhaps more likely hypothesis would involve the polar effects of a *dcmA* deletion on the expression of downstream *dcmBC* genes of unknown function. Further experiments are needed to test hypotheses of "co-repressor like" roles of *dcmABC* genes in *dcm* gene expression.

3.3. New insights on activation of *dcm* genes expression in presence of DCM

Up to now, activation of the expression of *dcm* genes in *M. extorquens* DM4 during growth with DCM was thought to result solely from DcmR-mediated derepression (La Roche & Leisinger, 1991). This hypothesis was revisited here by assessing the impact of the STOP*dcmR* mutation on the activity of P_R and P_A promoter reporter fusions in *M. extorquens* DM4 grown with DCM.

We showed here that increased gene expression from the P_R promoter was higher with DCM than with methanol or succinate, irrespectively of the presence of DcmR (Fig. 4B, P_R). Moreover, gene expression from P_A in the presence of DCM was even higher in mutants carrying the STOP*dcmR* mutation than in the WT strain (Fig. 4B, 22.6 ± 3.6 vs. 14.4 ± 2.4, respectively). This last result could be explained by incomplete derepression due to residual repressor activity of DcmR in the WT strain, so that P_A activity increases upon removal of DcmR. Alternatively, an as yet uncharacterized factor may also be involved in increased expression of *dcmA* in the presence of DCM. However, this DCM-dependent P_A "activation" process seems ineffective in the presence of

methanol (Fig. 4B, compare WT and mutant cultures with DCM, and DCM supplied with methanol). In addition, significantly higher transcription from *dcmA* promoter P_A was observed with reporter plasmid pME8703 harbouring the complete *dcmR-dcmA* intergenic region, compared to plasmid pME8701 with the short promoter fragment (Fig. 4C, P_A, and Fig. S4). On the other hand, long and short intergenic regions show similar promoter activity in WT and STOP*dcmR* genetic backgrounds. Thus, activation of *dcmA* transcription appears to also occur in the absence of DcmR, and involves a sequence located near the P_R *dcmR* promoter which has an effect on the activity of the P_A promoter of *dcmA*.

3.4. New lessons from quantification of dcm catabolic transposon transcripts and proteins

We wished to explore the influence of DcmR presence or absence on the expression of *dcmA*, *dcmB* or *dcmC* in a quantitative way, and how this could correlate with the observed regulation of DCM dehalogenase and growth phenotypes in strain DM4. Thus, abundance of transcripts and proteins of the *dcm* catabolic transposon were compared for WT and mutant strains grown with DCM and other carbon sources (Fig. 6). Reference genes noted in previous studies to display basal constitutive expression in *M. extorquens* grown with either DCM or methanol (*dnaA*, DNA replication initiator protein; *rpoA*, DNA-directed RNA polymerase, alpha subunit) (Bibi-Triki *et al.*, 2018; Chaignaud *et al.*, 2017) were selected as controls. For the broader range of conditions in this study (mutant genetic background and substrates), however, only gene *dnaA* proved suitable (Fig. 6), since expression of *rpoA* was associated with high background noise at the transcriptional level, and its protein product RpoA showed higher abundance in the WT strain grown with succinate (Fig. S5). Several peptides were tested for SRM quantification of DcmRABC proteins (Fig. 6; Table. S4). Some peptides were eliminated after correlation studies for all peptides targeting the same protein (Fig. S6 and Table S4). Incidentally, the designed phenotype of STOP*dcmR* mutants was fully validated, with undetectable DcmR protein despite production of *dcmR* transcripts (Fig 6). Similarly, both *dcmB* and *dcmC* mRNA and proteins remained undetected in the STOP*dcmR*Δ*dcmBC* mutant as expected from its Δ*dcmBC* genotype (Fig. 5). We also checked that no Dcm protein were detectable in the Δ*dcm* mutant (Fig. 6B). It is worth noting in this context that similar expression patterns were observed for all genes of the *dcmABC* operon in the WT strain,

with less transcript abundance of *dcmBC* genes more distant from the P_A promoter than *dcmA*. In particular, both *dcmA* transcript and DcmA protein abundance were about 10-fold higher than those of *dcmBC* transcripts and DcmB and DcmC proteins (Fig. 6A, Fig. 6B) in the WT strain grown with DCM.

Our analysis also confirmed that the DcmR protein was indeed detectable in the absence of DCM, as shown previously (Bibi-Triki *et al.*, 2018), and unlike found initially by two-dimension gel proteomics (Muller *et al.*, 2011b). Worthy of note, the present study as well as the two studies just mentioned all show that DcmR protein is present in larger abundance in the presence of DCM than in its absence, in clear contradiction to the originally proposed model for regulation of DCM dehalogenation by DcmR (La Roche and Leisinger, 1991). Further, all four *dcm* transcripts were detected with higher relative abundance in STOP*dcmR* mutants than in the WT strain in the absence of DCM, whereas the detected abundance of *dnaA* remained unchanged (Fig. 6A). This last result confirms that DcmR represses transcription of all four *dcm* genes in the WT strain growing with other substrates than DCM. In contrast, no significant differences in the abundance of transcript and protein gene products were detected for *dcmA*, *dcmB*, and *dcmC* between WT and *dcmR* mutants growing with DCM (Fig. 6). In other words, the longer generation time of *dcmR* mutants (Fig. 2) does not correlate with lower DcmA abundance. It cannot be excluded that as yet unknown post-translational modifications of DcmA may differ between WT and *dcmR* mutants. Alternative hypotheses would include that DcmR has an additional function besides transcriptional regulation, or that DcmR targets other genes than those of the *dcm* catabolic transposon, which would directly or indirectly be required for growth with DCM.

3.5. Towards the identification of new DcmR targets outside of the dcm catabolic transposon? METDI3043 as a test case

Gene of unknown function METDI3043 was one of the CDS shown in previous studies to be significantly up-regulated during growth with DCM compared to with methanol, both at the RNA and protein levels (Chaignaud *et al.*, 2017; Bibi-Triki *et al.*, 2018). In contrast, a close paralog (60% ID at the protein level, METDI1604) was not. METDI3043 encodes a protein of unknown function comprising a pair of conserved CBS (cystathione β-synthase) domains. Such proteins are

widespread in prokaryotes and eukaryotes, and have potential roles in signaling and gene regulation (King *et al.*, 2008; Anashkin *et al.*, 2017). Several studies have suggested that proteins with paired CBS domains can bind nucleotides such as ATP and AMP or DNA, and may be active in regulation of transcription (McLLean *et al.*, 2004; Aguado-Llera *et al.*, 2010; Kozhevnikova *et al.*, 2012).

To evaluated whether METDI3043 could be a target of DcmR-mediated regulation, the effect of DcmR on expression of METDI3043 was investigated using transcriptional fusions and RT-qPCR (Fig. 7). RT-qPCR data confirmed differential expression of METDI3043 at the transcriptional level under DCM and methanol growth conditions (Fig. 7B). Two transcription start sites (TSS) and corresponding promoters were found for this monocistronic gene (unpublished RNA-seq analysis of TSS identification, data not shown), which is strongly conserved in phylogenetically closely related strains unable to use DCM for growth. Strikingly, METDI3043 transcript abundance was significantly higher in the STOP*dcmR*Δ*dcmBC* strain than in the WT strain with DCM but not with methanol as the growth substrate, suggesting that genes *dcmR* (or *dcmBC*) may be involved in the observed differential expression. Transcriptional fusion data, however, indicated that the predicted P1 promoter of METDI3043 was not regulated during growth with DCM, and that *dcmR* mutants had no impact on transcription of the gene (Fig. 7C). How expression of *dcm* genes and METDI3043 interconnect in response to growth with DCM remains to be investigated. As a first step, a plasmid construct of the transcriptional fusion of the P2 promoter will be used in future experiments to check whether differential expression of METDI3043 occurs from this second promoter.

4. Conclusions and perspectives

In this study, we have confirmed by new experiments that DcmR is a transcriptional repressor of *dcmR* and *dcmABC* expression during growth with methanol, and plays an as yet uncharacterized role during growth with DCM, independently of its regulatory role on expression of *dcm* genes. Several hypotheses may contribute to explain why lack of DcmR impacts the growth rate of *M. extorquens* DM4 with DCM, but not abundance of DcmA (and DcmB and DcmC) proteins. First, DcmR may alter or modify the activity of another gene or protein that contributes to a higher growth rate with DCM. Possibilities include production of a protecting factor against detrimental

toxic products produced by dehalogenation of DCM such as formaldehyde, or conversely, post-translational modifications of the DcmA protein leading to higher activity of the DCM dehalogenase enzyme. Secondly, the yet uncharacterised MEDS domain of unknown function of DcmR (Anantharaman and Aravind, 2005) may be involved in other functions beyond regulation of expression of *dcm* genes, such as DNA repair as shown for the Ada transcription factor (Mielecki and Grzesiuk, 2014). In the case of DcmR, this may occur through e.g. recognition and deletion of alkylation products derived from DCM or its reactive glutathionyl adduct S-chloromethylglutathione generated by DCM dehalogenase catalysed transformation of DCM (Kayser and Vuilleumier, 2001). Certainly, future experiments will need to assess whether MEDS plays a role in DcmR function.

In conclusion, this study suggests that regulation of the catabolic *dcm* transposon, while clearly involving the *dcmR* gene and its protein product, is even more complex than anticipated. With new omics tools now at hand, the roles of transcripts, proteins, and protein-protein interactions in this process now need to be investigated in detail, with the potential additional involvement of still unknown, *dcmR*-independent contributions for activation of *dcmA* expression in the presence of DCM.

Acknowledgments

BM is the recipient of a PhD grant from the French Ministry of Research and Higher Education. SBT is the recipient of a postdoctoral grant "Investissement d'Avenir" (Idex 2015) from the University of Strsbourg, awarded to FB. Internship students A. Boyon and E. Chevalier are thanked for technical assistance.

Figures and legends

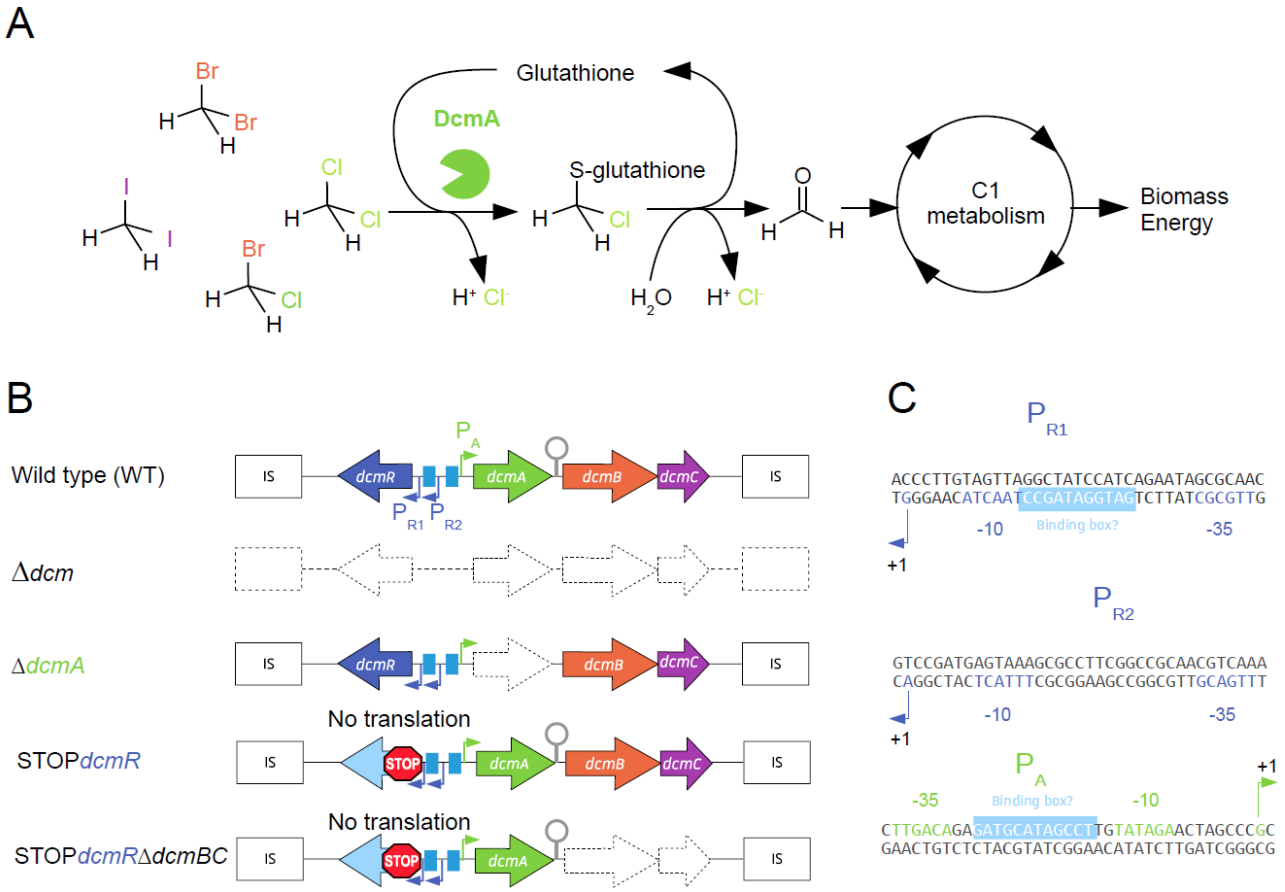


Fig. 1. The *dcm* catabolic transposon. **A)** DCM degradation involving DcmA, a DCM dehalogenase of the glutathione-S-transferase family. DcmA activity has deleterious side effects including intracellular production of HCl and of genotoxic metabolites such as S-chloromethylglutathione and formaldehyde, a central intermediate of methylotrophic metabolism. **B)** Genetic structure of the *dcm* catabolic transposon in *M. extorquens* DM4. Previously proposed (La Roche and Leisinger, 1991) promotors for expression of *dcmA* (P_A) and d*cmR* (P_{R1} and P_{R2}) promoters, putative transcription terminator found downstream of *dcmA* (grey) and DcmR-binding boxes (blue rectangles) are shown. **C)** Sequence features of the three detected promoters within the *dcmR-dcmA* intergenic region. The 12 bp conserved sequence of putative DcmR-binding boxes differ for P_{R1} et P_A differ by 1 nucleotide only (la Roche and Leisinger, 1990). No canonical -10 and -35 promoter sequences were found upstream of *dcmB* and *dcmC* genes, suggesting that the promoter of *dcmA* (P_A) is probably responsible for *dcmBC* downstream gene expression, as experimentally confirmed here by RT-PCR (Fig. S1).

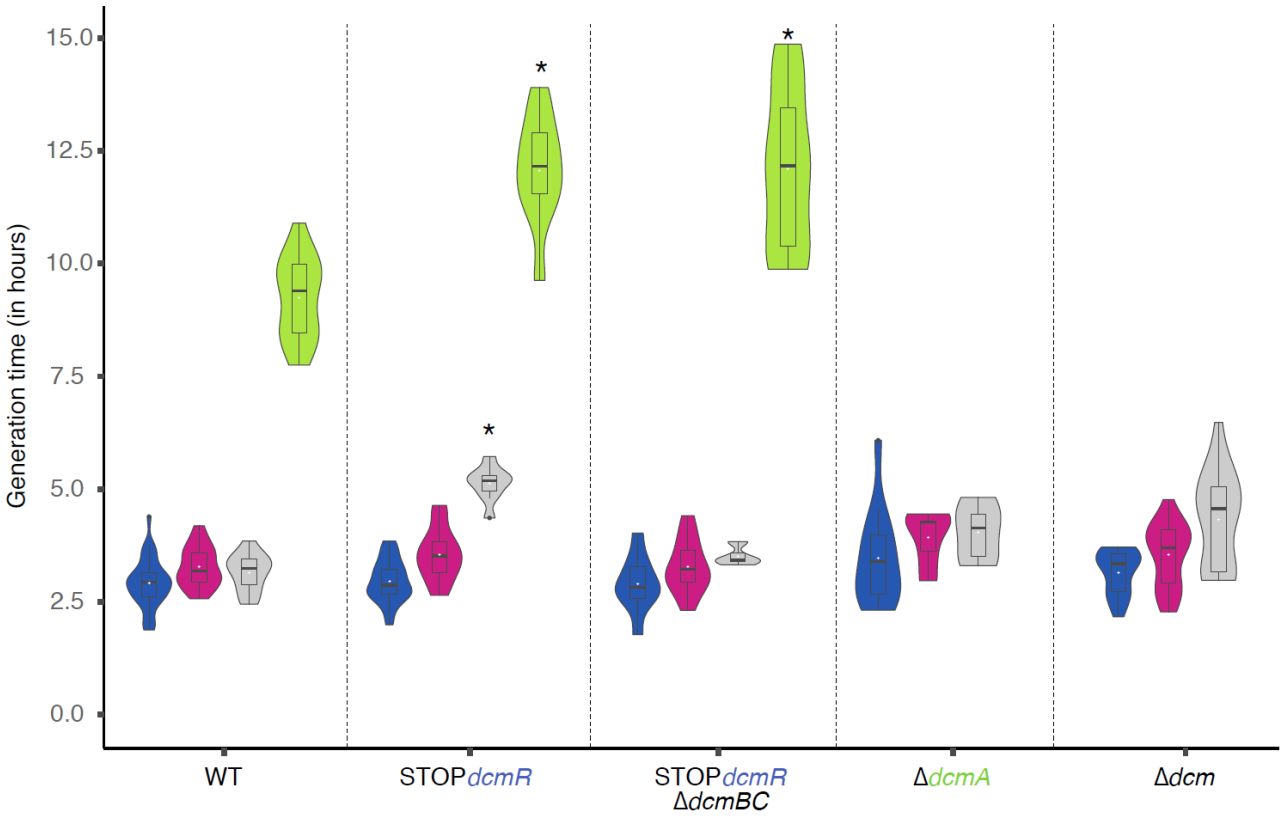


Fig. 2. Growth of *M. extorquens* DM4 WT and *dcm* mutants. Generation times with 10 mM succinate (blue), methanol (pink), methanol and DCM (grey), and DCM alone (green) were determined in growth experiments (5-38 biological replicates). Generation time of the WT strain DM4 with DCM (9.2 ± 0.9 h) was longer than with methanol (3.3 ±0.4 h), and similar to previously published values (Muller, *et al.*, 2011b; Chaignaud *et al.*, 2017). No growth with DCM was observed for Δ*dcmA* and Δ*dcm* mutants, since gene *dcmA* is essential for growth with DCM (La Roche and Leisinger, 1990). Stars indicate statistically significant differences between mutant and WT growth under the same condition.

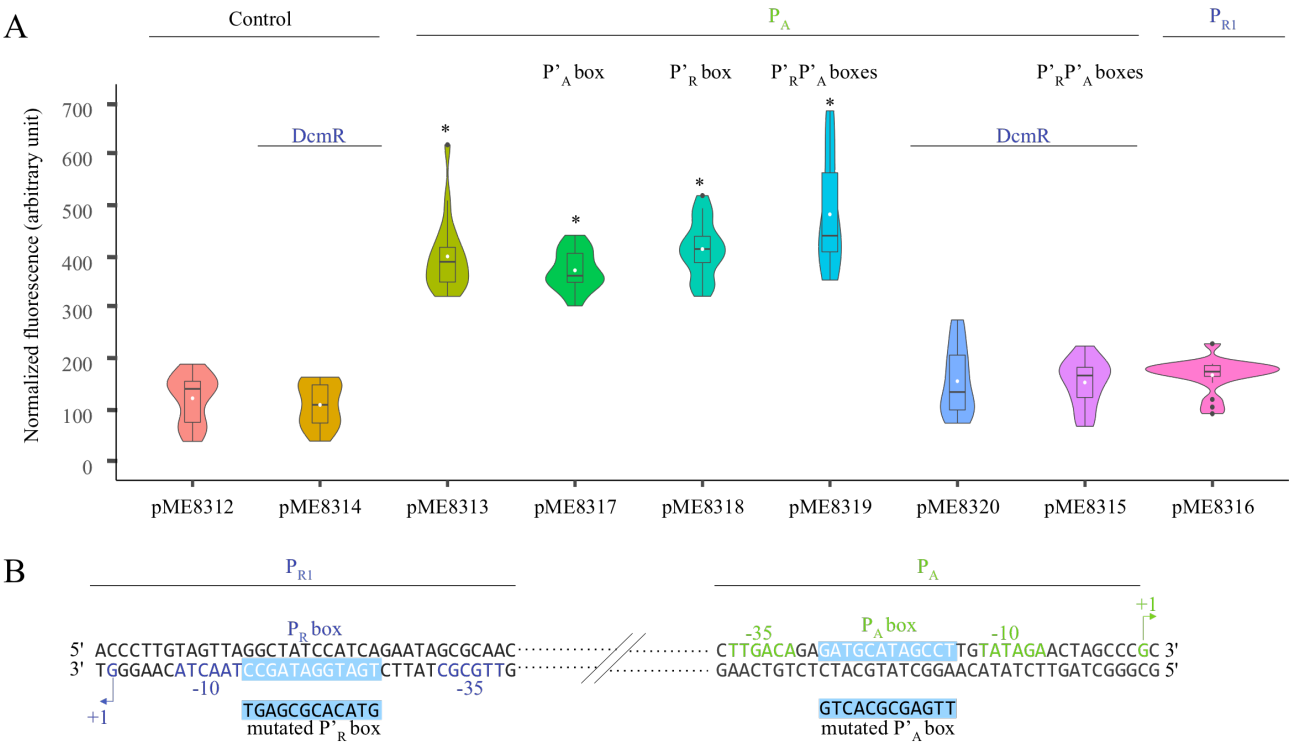


Fig. 3. *dcmR*-dependent repression of *dcmR-dcmA* intergenic region promotor activity in *E. coli*. **A)** Promoter activity was assessed by transcriptional fusion of the 0.7 kb-long intergenic *dcmR-dcmA* region cloned in both orientations relative to the promoterless *yfp* reporter gene (18 biological replicates). The activity of P_A (plasmid pME8313) or of P_R when cloned in the reverse orientation (pME8316) were compared to control plasmids with no insert present upstream of *yfp* (pME8212 and pME8314) (Table S2). The intergenic region was also mutated in the P_A box (plasmid pME8317), the P_R box (pME8318) and in both boxes (pME8315, pME8319). The *dcmR* gene was expressed from the P*cmtO* promoter (Chubiz *et al*., 2013) in the control plasmid pME8314 as well as in plasmids pME8315 and pME8320, but was absent in the other plasmids. The P_R promoter (plasmid pME8316) showed only weak activity, similar to background *yfp* expression. **B)** Putative DcmR-binding sequence between -35 and -10 boxes of *dcmA* and *dcmR* promoters. This sequence was disrupted in plasmids pME8315, pME8317, pME8318, pME8319 in order to prevent potential binding of DcmR, with no detectable effect compared to the non-mutated sequence present in plasmids pME8313 and pME8320. Stars indicate statistically significant differences compared to control plasmids.

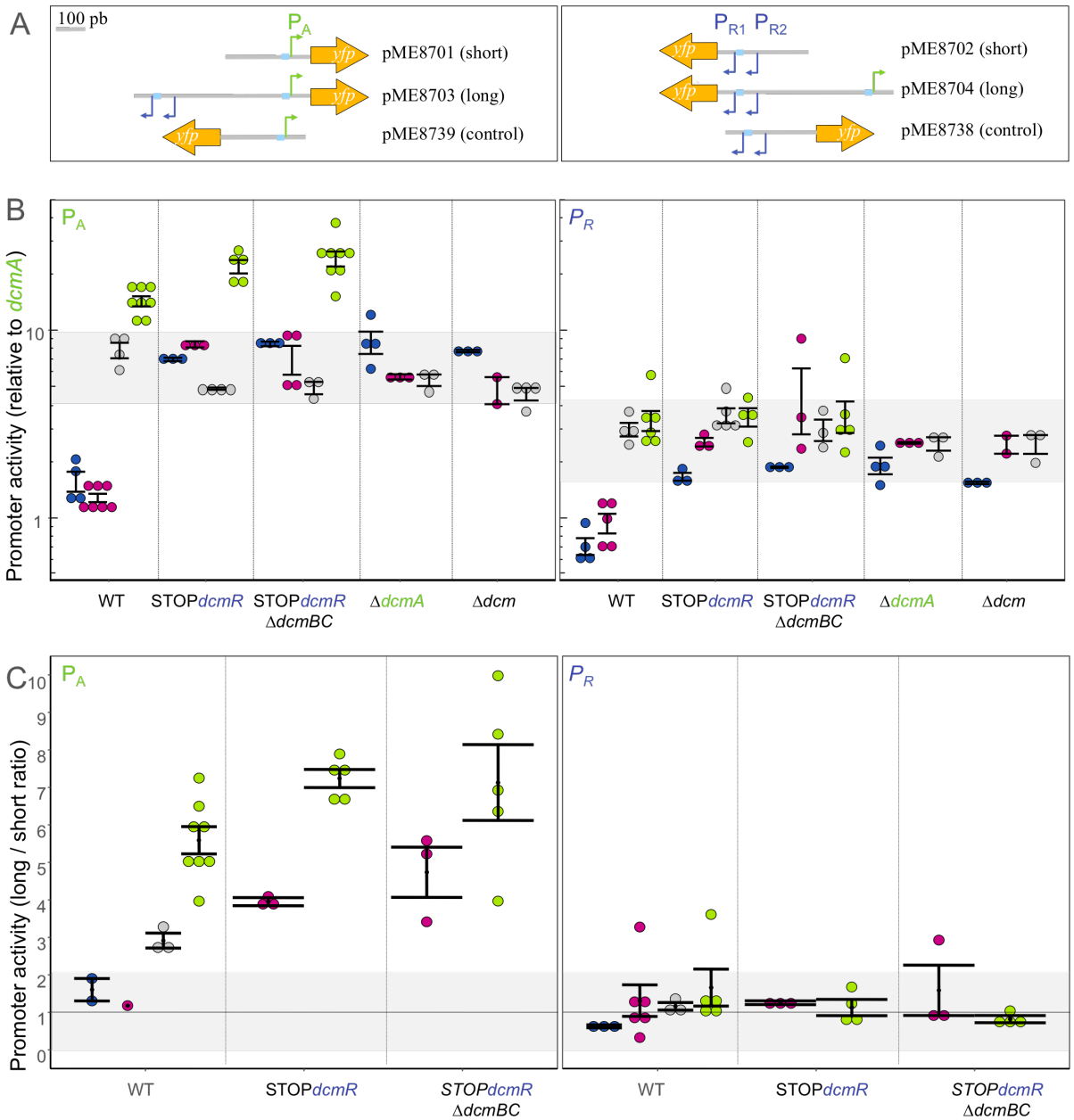


Fig. 4 Activity of *dcmR* and *dcmA* promoters in *M. extorquens* DM4. Promoter activity was assessed from cultures grown with succinate (blue), methanol (pink), an equimolar mixture of DCM and methanol (grey), or DCM alone (green). Promotor activity of the *dcmR-dcmA* intergenic region cloned in promoter probe plasmid pLM-sYFP2 for expression of *dcmA* (left panels) and *dcmR* (right panels) was tested. **A)** Length, features and orientation of tested *dcmR-dcmA* intergenic region segments. **B)** Promoter activity of the entire *dcmR-dcmA* intergenic region in WT and *dcmR* mutants. Basal expression considered in absence of repression or activation is highlighted in grey. N.S., non-significant. **C)** The entire *dcmR-dcmA* intergenic region is required for *dcmA* expression but not for *dcmR* expression. The “long” to “short” ratio corresponds to the ratio of values obtained

for pME8703 and pME8701 in the case of P_A, and for pME8704 and pME8702 in the case of P_R, respectively. Ratios below 2 were considered not be to significant. Control plasmids (pME8738 and pME8739) displayed background levels of fluorescence similar to those of the pLM-sYFP2 vector (data not shown).

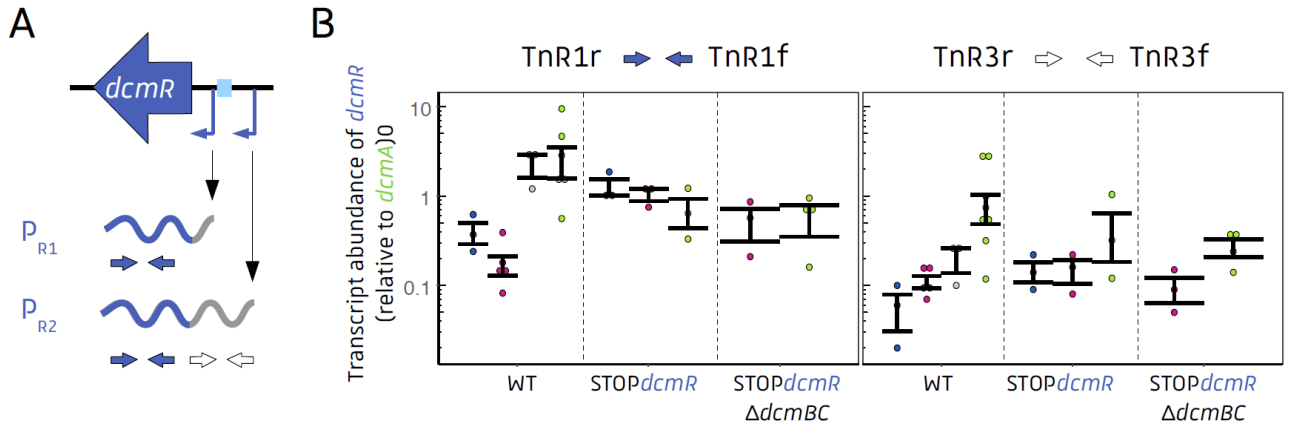


Fig. 5 Expression of *dcmR* from DCM-regulated (P_{R1}) and constitutive (P_{R2}) promoters. **A)** RT-qPCR using nested primers was used to demonstrate that only P_{R1} is regulated bu DCM. Primer positions on *dcmR* mRNA are shown. Primer set TnR1r/TnR1f detects transcripts from both promoters. Primer set TnR3r/TnR3f does not detect transcripts from P_{R1}. The predicted 12bp-long DcmR-binding site in promoter P_{R1} (La Roche and Leisinger, 1991) is indicated by a blue box. **B)** Relative abundance of *dcmR* mRNA transcripts. Dot colors correspond to growth with succinate (blue), methanol (pink), an equimolar mixture of DCM and methanol (grey), and DCM alone (green).

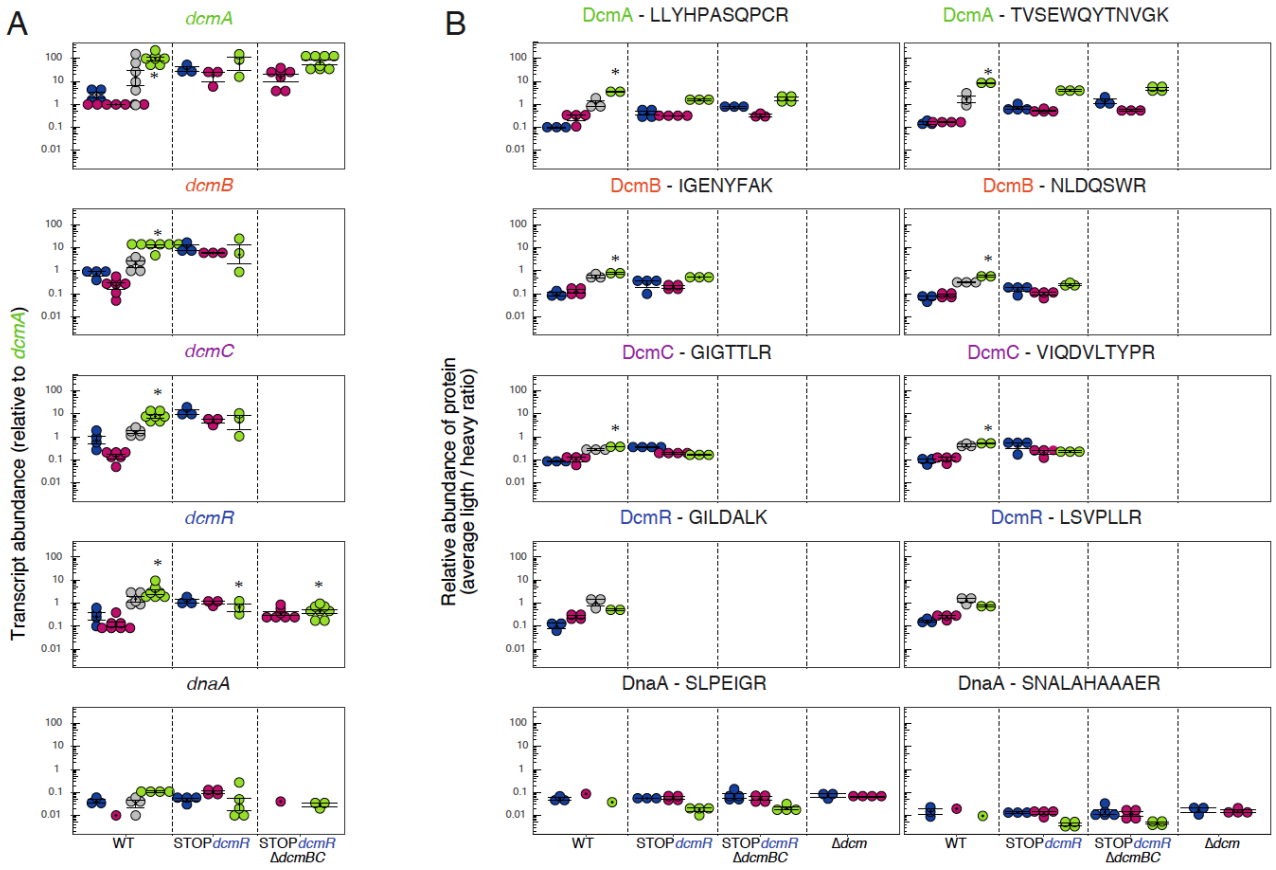


Fig. 6 Relative abundance of *dcm* transcripts and proteins in strain DM4 WT and *dcmR* mutants. **A)** Quantification of transcripts using RT-qPCR. Fold change differences were calculated by the $2^{-\Delta\Delta Ct}$ method, and normalized to the *rrsA* gene. Abundance is given relative to *dcmA* expression in the WT strain growing with methanol. **B)** Protein quantification using SRM. Reference peptides are indicated for each protein. Dot colors correspond to growth with succinate (blue), methanol (pink), an equimolar mixture of DCM and methanol (grey), and DCM alone (green).

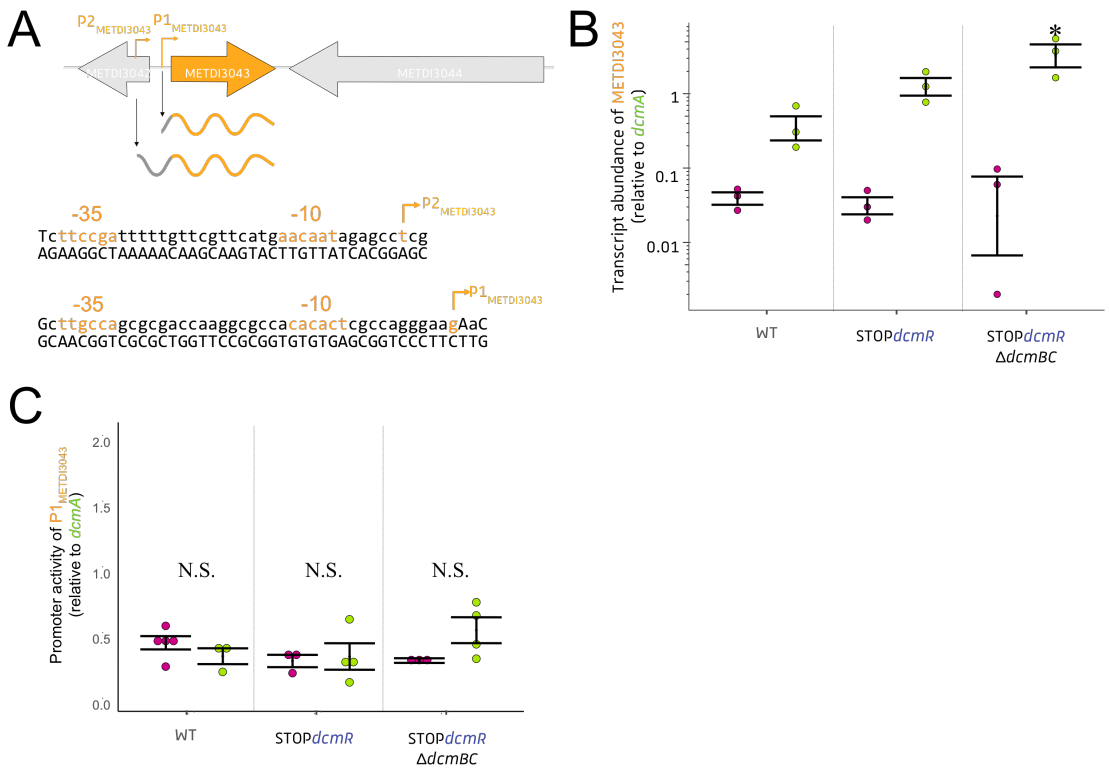


Fig. 7 DCM-dependent transcriptional regulation of METDI3043. **A)** Chromosomal localization of METDI3043 and predicted promoter sequences. The monocistronic MEDTI3043 gene encodes a conserved protein of unknown function with a tandem repeat of CBS domains. Structural studies have shown intramolecular association of such CBS domain pairs (see text for explanation). The interface between two CBS domains forms a cleft that is a potential ligand binding site (NCBI CBS_pair_10; cd04623). The downstream gene METDI3044 codes for a conserved protein with a domain found in proteins associated with biosynthesis of membranes, the cell envelope or the extracellular cell wall. The upstream gene MEDTI3042 encodes a hypothetical protein of unknown function. Predicted P_1 and P_2 promoters of METDI3043 are located at positions 2957298-2957336 and 2957343-2957387 on the *M. extorquens* DM4 chromosome (Vuilleumier *et al*., 2009). **B)** RT-qPCR quantification of METDI3043 transcripts using primer sets hybridizing within the gene. **C)** Promoter activity of P_1 cloned to the promoterless *yfp* fluorescent reporter gene in plasmid pME8324.

Supplementary figures and tables

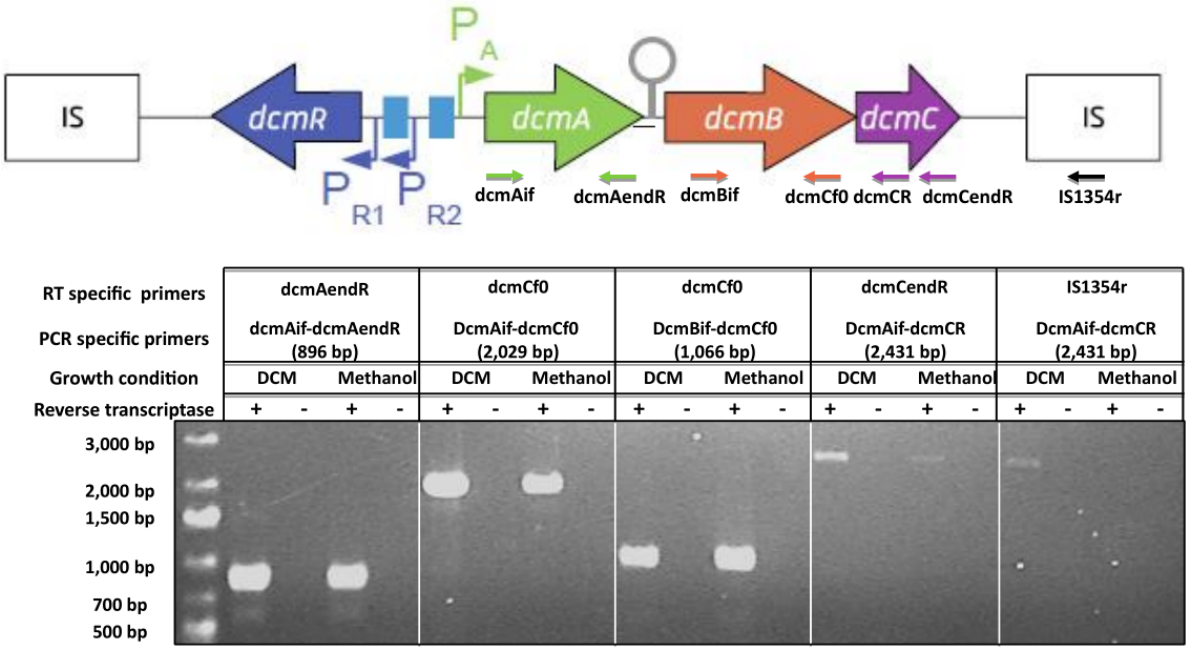


Fig. S1. Operon organisation of *dcmABC* genes. **A)** Position of primers used for reverse transcription and amplification of cDNA. **B)** Size-separated RT-PCR products in 1 % agarose gel electrophoresis stained with ethidium bromide. Control reactions with no added reverse transcriptase confirmed the absence of DNA contamination in RNA samples. Genes *dcmA* and *dcmB*, and genes *dcmB* and *dcmC*, are separated by short intergenic regions of 177 and 61 bp, respectively. A terminator sequence was previously identified 41 bp downstream of *dcmA* (La Roche and Leisinger, 1991). Results showed that *dcmABC* are co-transcribed under both DCM and methanol growth conditions. The IS*1354* downstream insertion sequence is co-transcribed with *dcmABC* only in the DCM growth condition.

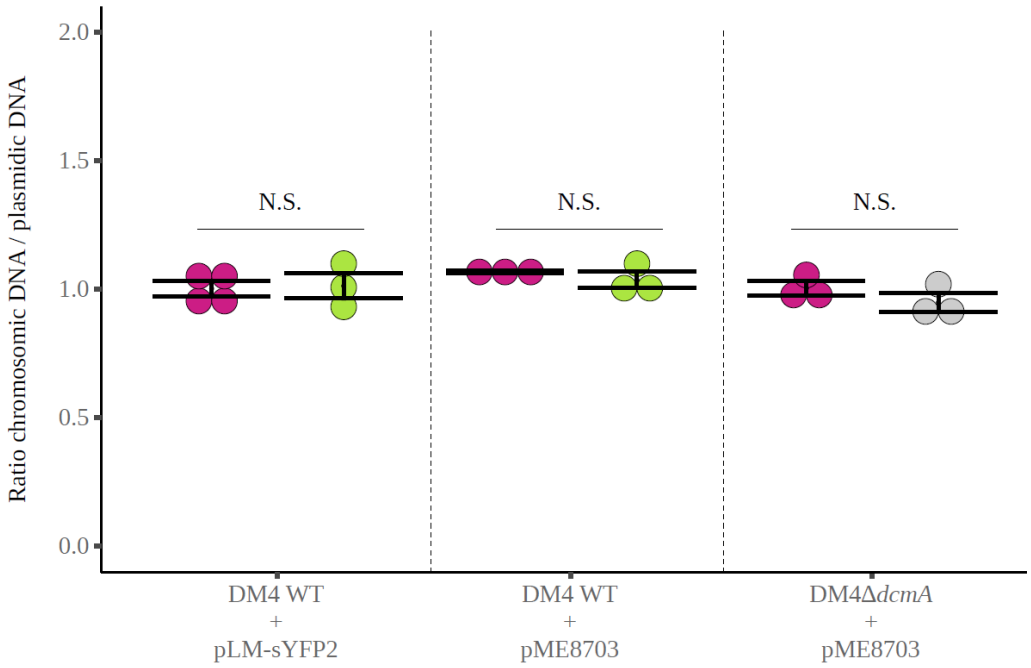


Fig. S2. Plasmid stability in transcriptional fusion assays. Three examples are shown in the genetic context of the *M. extorquens* DM4 wild-type and *dcmA* mutant strains for plasmids carrying different inserts. Copy number ratio between the chromosomal *gmaS* gene and the plasmid *kanR* gene was calculated. Statistical significance was evaluated by Kruskal-Wallis and Dunn post-hoc (Bonferonni correction) tests ; N.S. non-significant.

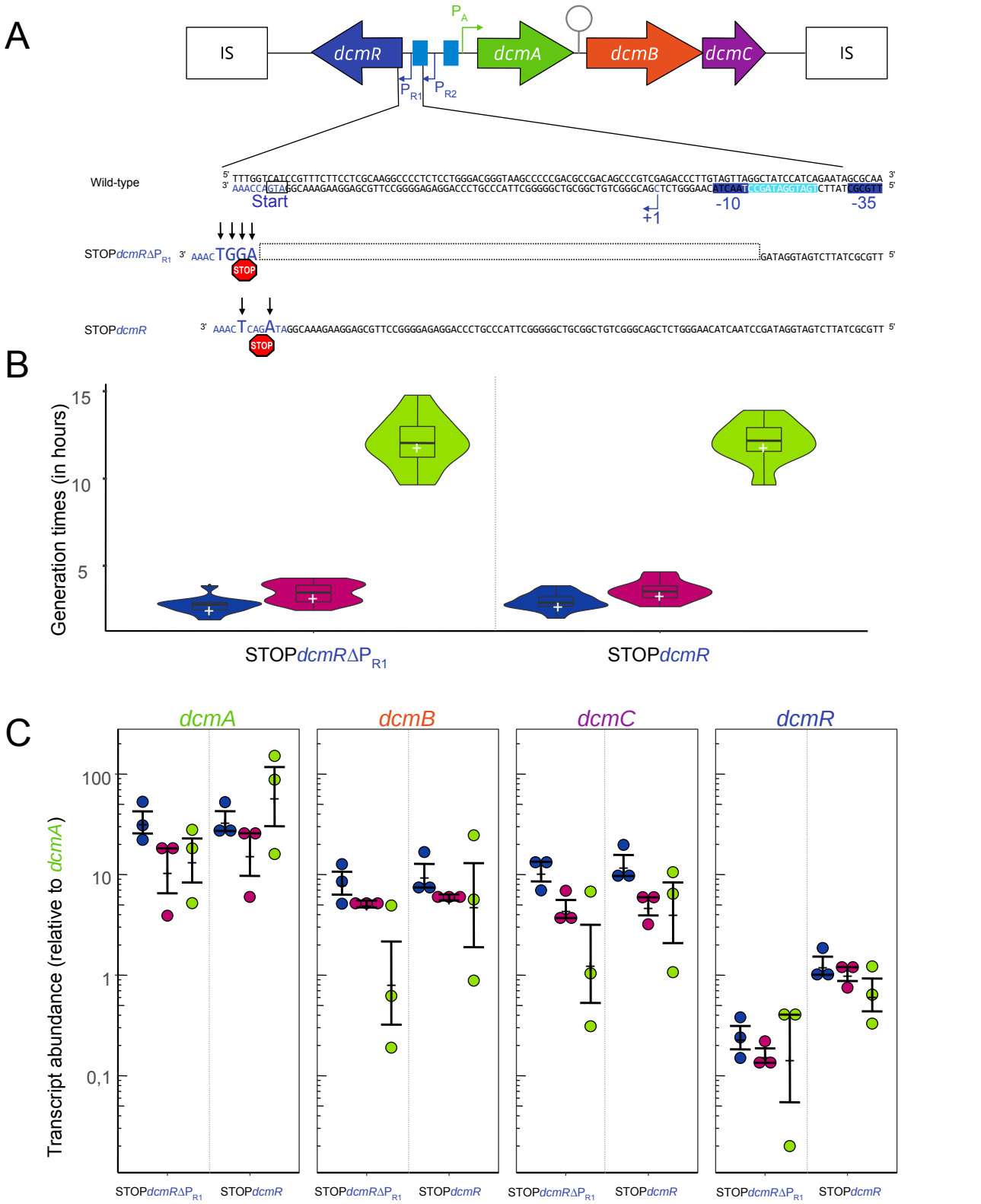


Fig. S3. Comparison of *dcmR* mutants STOP*dcmR*ΔP$_{R1}$ and STOP*dcmR*. **A)** Genetic map and mutation positions in *dcmR* mutants. Black arrows indicate nucleotide insertions. Deletion of P$_{R1}$ is indicated by a dotted rectangle. In mutant STOP*dcmR*ΔP$_{R1}$, *dcmR* transcription proceeds from

promoter P_{R2} only. In both mutants, translation initiation of DcmR is blocked by stop codons in all three reading frames. **B)** Generation times. Similar growth phenotypes were observed for both mutants. Data were obtained from 13 to 21 biological replicates. **C)** Quantification of *dcm* transcripts. The only observed difference was lower mRNA abundance of *dcmR* in *dcmR*ΔP_{R1}. Dot colors correspond to growth with succinate (blue), methanol (pink) and DCM (green).

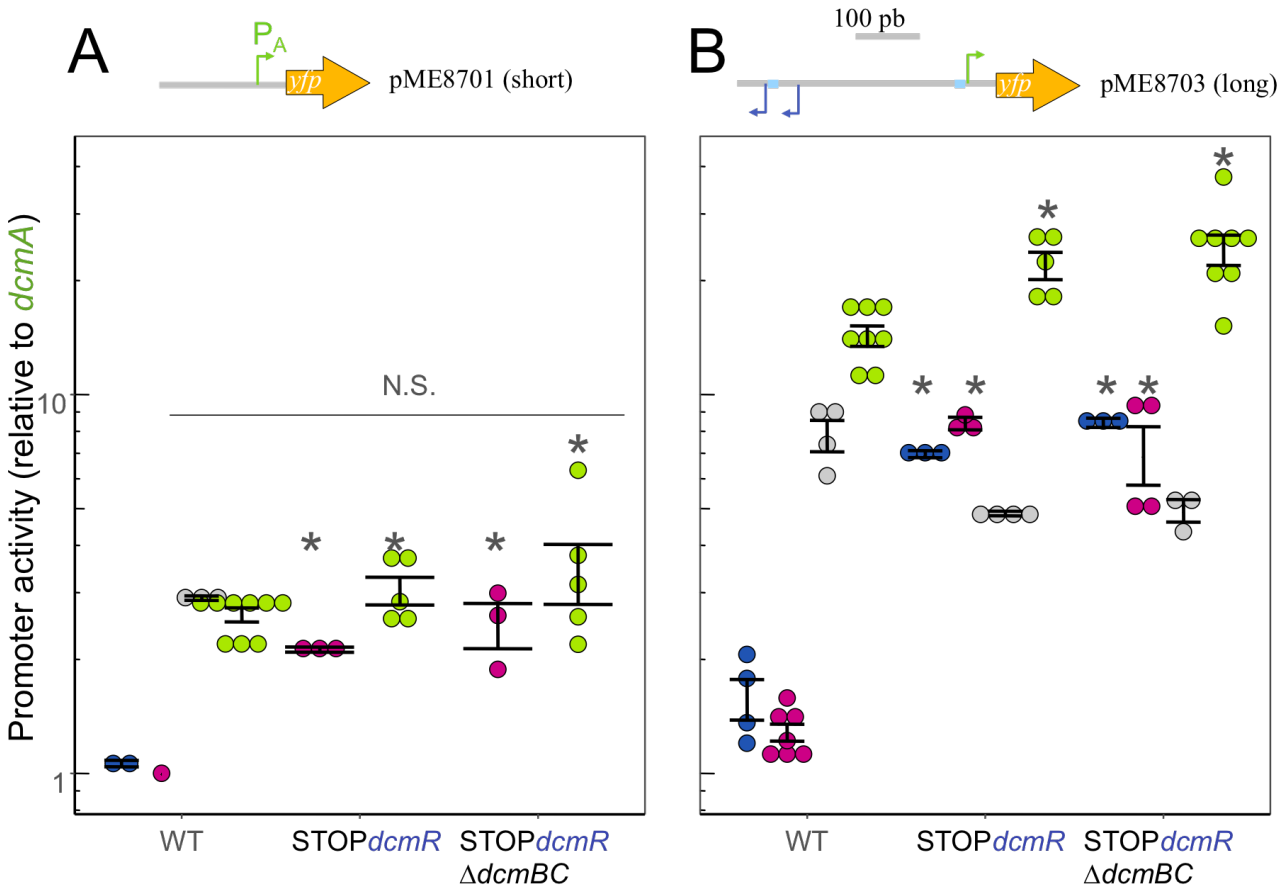


Fig. S4. P_A promoter activity assessed by plasmidic gene fusions. **A)** Promoter activity of P_A with short 5' usptream sequence. **B)** Promoter activity of P_A with long 5' upstream sequence. Dot colors correspond to growth with succinate (blue), methanol (pink) and DCM (green). Statistical significance was assessed by the Kruskal Wallis test with Dunn post-hoc comparison (Benjamini-Hochberg correction). Stars indicate a significant difference in mutants compared to WT strains under the same condition. N.S., non-significant.

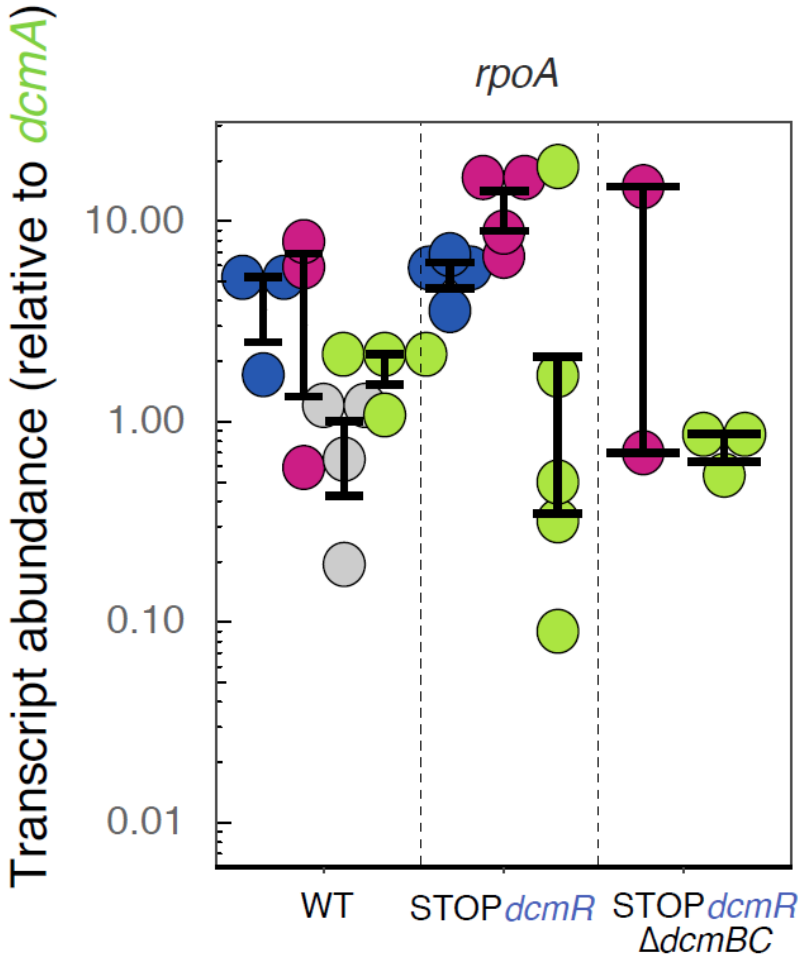


Fig. S5. Relative expression of *rpoA* in wild-type and *dcmR* mutants. Fold change after RT-qPCR quantification was calculated by the $2^{-\Delta\Delta Ct}$ method and normalized to *rrsA* gene expression. Abundance is given relative to *dcmA* expression in the wild-type strain grown with methanol. Dot colors correspond to growth with succinate (blue), methanol (pink), an equimolar mixture of DCM and methanol (grey), and DCM (green).

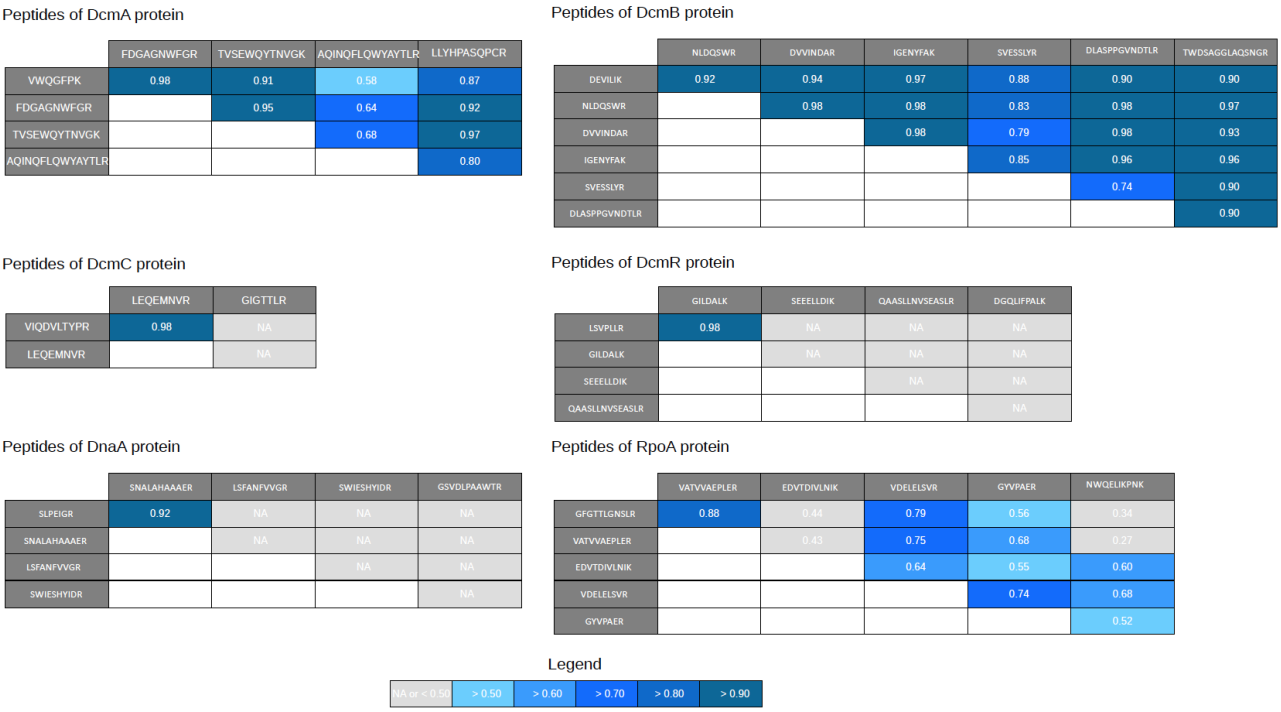


Fig. S6. Correlation of SRM results for each peptide by the Pearson test. Values of R^2 are indicated.

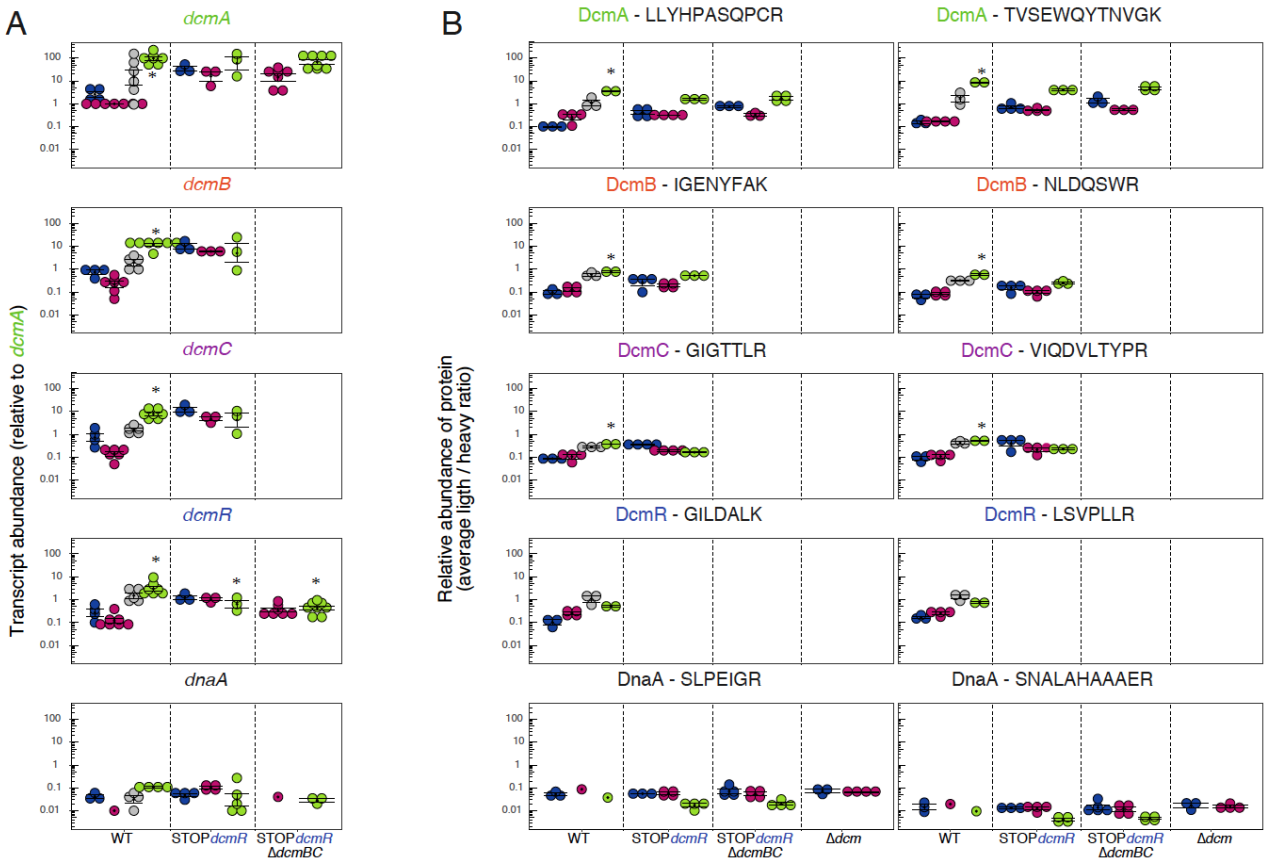


Fig. S7. Relative abundance of mRNA and proteins encoded by the *dcm* catabolic transposon in the wild-type strain and the STOP*dcmR* mutant. **A)** Relative abundance of mRNA transcript by RT-qPCR. **B)** SRM quantification of DcmA, DcmB, DcmC, DcmR, DnaA and RpoA proteins. Dot colors correspond to growth with succinate (blue), methanol (pink), an equimolar mixture of DCM and methanol (grey), and DCM alone (green).

Table S1. Ability of DCM-degrading strains containing the *dcm* catabolic transposon to degrade DCM and to grow with halogenated compounds.

Substrate	Growth [a]	DcmA enzymatic activity (km in µM) in strain			Bioreporter induction in strain		Comment
		DM2 [b]	DM4 [c]	DM11 [c, d]	DM2 [e]	DM4 [f]	
Chlorinated methane							
Chloromethane CH_3Cl	No	ND [g]	ND	ND	ND	No	/
Dichloromethane CH_2Cl_2	All strains	30	9	59±6 [c] 31± 6 [d]	Yes	Yes	/
Bromochloromethane CH_2ClBr	ND	15	ND	1.62±0.01 [d]	Yes	ND	/
Fluorochloromethane CH_2ClF	ND	ND	ND	53±18 [d]	ND	ND	/
Iodochloromethane CH_2ClI	ND	ND	ND	2.0±0.3 [d]	ND	ND	/
Trichloromethane $CHCl_3$	No	ND	ND	ND	No	No	/
Tetrachloromethane CCl_4	ND	ND	ND	ND	ND	ND	/
Non-chlorinated halogenated methane							
Dibromomethane CH_2Br_2	Only DM11	13	ND	3.5±0.4 [c] 3.1±0.5 [d]	Yes	ND	Tested in DM2 and DM4
Diidomethane CH_2I_2	ND	5	ND	1.7±0.4 [c] 49±11 [d]	ND	ND	/
Multi-carbon chlorinated solvent							
Chloroethane C_2H_5Cl	ND	ND	ND	330±46 [d]	ND	ND	/
Bromoethane C_2H_5Br	ND	ND	ND	330±40 [d]	ND	ND	/
Iodoethane C_2H_5I	ND	ND	ND	39±7 [d]	ND	ND	/
Dichloroethane $C_2H_4Cl_2$	No	ND	ND	ND	No	ND	Inhibits DCM metabolism
1,1,1-trichloroethane $C_2H_3Cl_3$	ND	ND	ND	ND	ND	No	/
Trichloroethylene C_2HCl_3	ND	ND	ND	ND	ND	No	/
Tetrachloroethylene C_2Cl_4	ND	ND	ND	ND	ND	No	/

[a] Scholtz *et al.,* 1988; [b] Kohler-Staub & Leisinger, 1985; [c] Vuilleumier & Leisinger, 1996; [d] Stourman et al., 2003; [e] Henrysson & Mattiasson, 1993; [f] Lopes *et al*., 2012.
[g] ND Not determined

Name	Description	Reference
DH5α	*Escherichia coli* strain. Used for cloning and plasmid extraction Used for tri-parental conjugation; carrying pRK2013	Grant *et al.*, 1990
DM4	*M. extorquens* wild-type DCM-degrading strain	Galli and Leisinger, 1985
DM4Δ*dcm*	*M. extorquens* mutant of strain DM4. Deletion of *dcm* islet (2559013-2567076)[a] including flanking IS*1354*; spontaneous mutant following insertion of plasmid pCM433Δ*dcmB*	This study
DM4Δ*dcmA*	*M. extorquens* mutant of strain DM4. Deletion of *dcmA* gene (2562957-2563820)	This study
DM4 STOP*dcmR*ΔP_{R1}	*M. extorquens* mutant of strain DM4. Deletion of *dcmR* promoter (2562261-2562347) and AGGT (2562261) insertion to disrupt *dcmR* start codon. Spontaneous mutant following insertion of plasmid pCM433*dcmR2*	This study
DM4 STOP*dcmR*	*M. extorquens* mutant of strain DM4. Insertions of A (2562264) and T (2562261) to disrupt *dcmR* start codon	This study
DM4 STOP*dcmR* Δ*dcmBC*	*M. extorquens* mutant with *dcmBC* (2563926-2566982) deletion in strain DM4 *dcmR2*	This study
Plasmids		
pCM433	Broad-host-range *sacB*-based allelic exchange vector for *Methylobacterium*	Marx, 2008
pCM433Δ*dcmA*	XbaI-restricted PCR-amplified fragment of 1.07 kb with Δ*dcmA* mutation cloned into XbaI-linearized pCM433	This study
pCM433Δ*dcmB*	XbaI-restricted PCR-amplified fragment of 1.12 kb Δ*dcmB* mutation cloned into XbaI-linearized pCM433	This study
pCM433*dcmR2*	XbaI-restricted PCR-amplified fragment of 0.62 kb STOP*dcmR* mutation cloned into XbaI-linearized pCM433	This study
pCM433Δ*dcmBC*	XbaI-restricted PCR-amplified fragment of 1.07 kb Δ*dcmBC* mutation and deletion of adjacent IS*1354* cloned into XbaI-linearized pCM433	This study
pLM-sYFP2	Promoterless *yfp* gene; *kanR* (kanamycin resistance gene) used for qPCR standard curve	Kaczmarczyk *et al.*, 2011
pME8280	pCM80::*gmaS*; used for qPCR standard curve	Gruffaz *et al.*, 2014
pME8701	pLM-sYFP2::P_A (2562623-2563049)	This study
pME8702	pLM-sYFP2::P_R (2562631-2562165)	This study
pME8703	pLM-sYFP2::P_R-P_A (2562165-2563049)	This study
pME8704	pLM-sYFP2:: P_A-P_R (2563049-2562165)	This study
pME8312	Bi-reporter promoter probe with divergent mCherry and *yfp* reporter genes	This study
pME8313	Intergenic *dcmR-dcmA* region (2562165-2563049) in pME8312, P_A oriented towards *yfp*	This study

pME8315	Intergenic *dcmR-dcmA* region (2562165-2563049) in pME8314, P_A oriented towards *yfp*, *dcmR* cloned in pME8312 for DcmR protein expression	This study
pME8316	Intergenic *dcmR-dcmA* region (2562165-2563049) in pME8314, P_R oriented towards *yfp*	This study
pME8317	Intergenic *dcmR-dcmA* region (2562165-2563049) in pME8314, mutated P_R P_A oriented towards *yfp*,	This study
pME8318	Intergenic *dcmR-dcmA* region (2562165-2563049) in pME8314, mutated P_A, P_A oriented towards *yfp*	This study
pME8319	Intergenic *dcmR-dcmA* region (2562165-2563049) in pME8314, mutated P_A and P_R P_A oriented towards *yfp*	This study
pME8320	Intergenic *dcmR-dcmA* region (2562165-2563049) in pME8314, mutated P_A and P_R P_A oriented towards *yfp*, *dcmR* cloned in pME8312 for DcmR protein expression	This study
pME8324	pLM-sYFP2::$P1_{METDI3043}$(2957333-2957470)	This study
pME8326	pLM-sYFP2::$P2_{METDI3043}$(2956778-2957470)	This study
pRK2013	Plasmid for mobilization of noneself-transmissible plasmid	Figurski and Helinski, 1979

[a]Nucleotide position in the chromosome of *M. extorquens* DM4 (accession number FP103042.2)

Use	Targeted sequence, gene or mutant (comment)	Primer set and 5'->3' sequence [a]	Tm (°C)[b]	PCR prod. (bp)	Genome localization (nt) [c]	References
For PCR validation of DNAse-treated RNAs						
	rrnA	27f	GTTTGATCCTGGCTCAG	53.1	859	5 copies (100% Id)	Chaignaud *et al.*, 2017
		927r	CCGTCAATTCCTTTRAGTTT	54.1			
	dcmA	dcmA1F	CGATCTAGATAGCGCGGTCTAAGCGACT	65.1	879	2562957-2562974 2563815-2563834	Chaignaud *et al.*, 2017
		dcmAendR	TCGATCTAGATAGCGCGGTCTAAGCGAC	67.1			
For RT-qPCR							
	dcmA	TnA1f	TTCGGCTGTTTGATCTATGG	59.5	159	2563344-2563363 2563484-2563502	Chaignaud *et al.*, 2017
		TnA1r	ATCCGCATAGCTGACCTCAT	62.6			
	dcmB	DcmCf0	GCGACGCAATCTTACTTCGAT	61.8	150	2564955-2564974 264825-2564844	Chaignaud *et al.*, 2017
		TnB1f	AGGGCTTGCTCAATCAAAT	59.5			
	dcmC	TnCf	CGGCGTCTGGTCATTTGG	62.4	152	2565225-2565242 2565357-2565376	Chaignaud *et al.*, 2017
		DcmCr	ATATCACTGCCCAGGTGTCG	63.2			
	P_{R1} (*dcmR* promoter n°1)	TnR1f	AGGTGCACAAAGCCATTAAA	60.4	151	2561793-2561812 2561662-2561681	Chaignaud *et al.*, 2017
		TnR1r	GCCCACTCCATATCACCAAC	61.9			
	P_{R2} (*dcmR* promoter n°2)	TnR3f	TGGTCATCCGTTTCTTCCTC	61.5	148	2562260-2562279 2562390-2562407	Chaignaud *et al.*, 2017
		TnR3r	CAGCGTTTCGGAGTCACG	62.4			
	rrnA	Mex1369f	CGGTGAATACGTTCCCGG	61.4	138	5 copies (100% Id)	Chaignaud *et al.*, 2017
		Mex1492r	GGCTACCTTGTTACGACTT	58.6			
	METDI3043 (DCM-regulated)	3043f	TATCCTGGCAGAGAAAGGCG	63.5	224	2957473-2957492 2957677-2957696	This study
		3043r	CCTTCGTCGTCATGTGATGC	62.8			
For qPCR to assess plasmid stability							
	gmaS (chromosome reference)	Gmas-F	GCAGATCGGCTACCACTTCA	61.1	158	2225714-2225793 2225912-2225931	Gruffaz *et al.*, 2014 This study
		METDI2327R	TGTTGCCCGTATAGGTCACG	61.1			
	kanF (plasmid reference)	qPCR-KanF	TCCGACTCGTCCAACATCAA	60.6	146	2689-2708 [d] 2815-2834 [d]	This study
		qPCR-KanR	GGCCGTGTTGAACAAGTCTGG	60.3			
For plasmid-borne transcriptional fusion construction							
	P_A (*dcmA* promoter)	dcmA-SacI	ATATGAGCTCATGCTGGTTAATCGCCGA	58.7	447	2562623-2562640 2563032-2563049	This study
		dcmA-KpnI	ATATGGTACCGAATTGATGTGCTGAGCG	55.9			
	P_R	dcmR-SacI	ATATGAGCTCAACCAGCATATCGAAGGG	56.1	485	2562614-2562631	This study

promoter n°1)	3043-PI-SacI	ATAYGAGCTCCTCGTCCGACGCTTGCCAGC	66.5	1536	2957333-2957332	
For site-directed chromosomal mutant construction						
Mutant Δ*dcmA* (Upstream of *dcmA*)	DcmAf	ATCGtctagaCCGCAACGTCAAAAGTTTA	58.8	529	2562446-2562465	This study
	dcmAdel1	GCGGTctagTTAATCCTCCCTTACGTGTGG	56.4		2562937-2562956	
(Downstream of *dcmA*)	dcmAr	GACAtctagaGTGGCCTGCCTTACGATCTC	61.5	521	2564304-2564323	This study
	dcmAdel2	AGGATAACTAGACCGCGCTATCGCGCA	65.1		2563821-2563839	
Mutant STOP*dcmR* (Upstream of *dcmR*)	DcmRf2	Gacttctagactccggtgacatcagtcg	58.5	333	2562539-2562538	This study
	dcmRmut2	GCGCGTGATCTTTCTTTagGTCTtATCC	55.0		2561950-2561972	
(Downstream of *dcmR*)	dcmRr2	ttaatctagaATCGCGCAAATATGATCGTG	58.0	317	2561954-2561972	This study
	dcmRmut3	AAAGAAAAGATCACGCGCAGCGAAGCG	68.4		2562235-2562260	
Mutant Δ*dcmBC* (Upstream of *dcmB*)	dcmBf	ATCGtctagaAAGCCGTACAGTCCAAAATTCA	59.7	575	2563368-2563389	This study
	dcmBCdel1	ctggagacCGTCTAATTCGGGCTTCCGGATC	63.4		2563904-2563925	
(Downstream of IS*1354*)	dcmBCdel2	aatagacgGTCTCCAGTCCGACGCTTCA	62.8	510	2566983-2567002	This study
	dcmBCr	GACAtctagaCACCAAAAATTGGCGATAGCTC	59.8		2567453-2567474	
Mutant Δ*dcm* Upstream of *dcmB*)	dcmBf	ATCGtctagAAGCCGTACAGTCCAAAATTCA	61	575	2563368-2563389	This study
	dcmBdel1	CGACAtccaCGTCTAATTCGGGCTTCCGGATC	68		2563904-2563925	
(Downstream of *dcmB*)	dcmBdel2	AATAGACGtcgaTGTCCGAAATCCGAAAATAGAC	58	538	2564985-2565007	This study
	dcmBr	GACAtctagaTAGCCTCCCCAGTGAAAACCTAT	61		2565503-2565524	
For PCR validation of *dcmABC* operonic structure						
validation of *dcmABC* operonic structure	dcmBendR	cgatctaGATAGCCCGGTCTAACGACT	63.3	Fig. S1.	2563814-2563834	This study
	dcmBiF	gacttctagacGTGTCCGTCCGTTCTGTTC	59.5	Fig. S1.	2563926-2563944	This study
	dcmCendR	aaaatctAGACTGGATCCGACAAGCCCTA	64.0	Fig. S1.	2565598-2565619	This study
	dcmCr	atatcactgcccaggtgtcg	60.5	Fig. S1.	2565357-2565376	This study
	IS1354r	agatggccagcaatgaagcaag	62.1	Fig. S1.	2566771-2566792	This study

[a] In bold, degenerate nucleotides with B= C, G or T; D = ; A, G or T; S = G or C; R = A or G; y= C or T; W = A or T. Nucleotides in lowercase are not present in the genome of *M. extorquens* DM4. Underlined sequences correspond to restriction enzyme recognition sites.

[b] Tm was calculated with OligoAnalyzer (IDT; V3.1) at 0.2 µM final concentration of primers, 50 µ monovalent salt, 1.5 mM divalent salt for PCR; 2.5 mM for qPCR, and 1 mM dNTPs for PCR and 0.8 mM for qPCR.

[c] Nucleotide position in the *M. extorquens* DM4 genome (accession number FP103042.2).

[d] Nucleotide position in plasmid pLM-sYFP2 (Kaczmarczyk *et al.*, 2011).

Table S4 Peptides used for SRM

Targeted protein	Peptide				SRM data		Reference
	Sequence	Size (aa)	Position	Missed cleavage	Psm score	Correlation	
DcmA (33 kDa)	LLYHPASQPCR [A]	11	16-26	0	83.39	Yes	This study
	AQINQFLQWYAYTLR	15	105-119	0	80.82	No	This study
	TVSEWQYTNVGK	12	224-235	0	80.53	Yes	This study
	FDGAGNWFGR	10	90-99	0	58.76	Yes	This study
	VWQGFPK	7	202-208	0	43.66	Yes	This study
	FTAEQNK	7	144-150	0	40.65	No	This study
DcmB (40 kDa)	TWDSAGGLAQSNGR	14	270-283	0	93.73	Yes	This study
	DLASPPGVNDTLR	13	160-172	0	84.27	Yes	This study
	FALINVWR	8	197-204	0	65.6	No	This study
	SVESSLYR	8	16-23	0	62.17	No	This study
	IGENYFAK	8	240-247	0	56.72	Yes	This study
	DVVINDAR	8	49-56	0	54.99	Yes	This study
	NLDQSWR	7	59-65	0	47.38	Yes	This study
	DEVILIK	7	263-269	0	45.19	No	This study
DcmC (21 kDa)	VIQDVLTYPR	10	72-81	0	85.17	Yes	This study
	LEQEMNVR	8	169-176	0	56.02	No	This study
	GIGTTLR	7	91-97	0	39.94	Yes	This study
DcmR (30 kDa)	SEEELLDIK	9	12-20	0	72.23	No	This study
	LSVPLLR	7	115-121	0	46.67	Yes	This study
	QAASLLNVSEASLR	14	21-34	0	43.13	No	This study
	DGQLIFPALK	10	159-168	0	36.83	No	This study
	GILDALK	7	244-250	0	30.92	Yes	This study

LSFANFVVGR	10	165-174	0	69.73	No	This study
SNALAHAAAER	11	175-185	0	60.04	No	This study
SWIESHYIDR	10	69-78	0	46.48	Yes	This study
GSVDLPAAWTR	11	17-27	0	37.96	No	This study
SLPEIGR	7	465-461	0	NA	Yes	This study
GFGTTLGNSLR	11	38-48	0	86.3	No	This study
VATVVAEPLER	11	27-37	0	68.79	No	This study
EDVTDIVLNIK	11	80-90	0	62.08	No	This study
VDELELSVR	9	259-267	0	53.89	Yes	This study
GYVPAER	7	154-160	0	45.54	Yes	This study
NWQELIKPNK	10	6-15	0	37.13	Yes	This study

ethyl (C10) modification.

Bibliography

Aguado-Llera, D., Oyenarte, I., Martínez-Cruz, L.A., and Neira, J.L. (2010) The CBS domain protein MJ0729 of *Methanocaldococcus jannaschii* binds DNA. *FEBS Lett.* **584**: 4485–4489.

Anantharaman, V. and Aravind, L. (2005) MEDS and PocR are novel domains with a predicted role in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. *Bioinformatics* **21**: 2805–2811.

Anashkin, V.A., Baykov, A.A., and Lahti, R. (2017) Enzymes regulated via cystathionine β-synthase domains. *Biochem. Mosc.* **82**: 1079–1087.

Bibi-Triki, S., Husson, G., Maucourt, B., Vuilleumier, S., Carapito, C., and Bringel, F. (2018) N-terminome and proteogenomic analysis of the *Methylobacterium extorquens* DM4 reference strain for dichloromethane utilization. *J. Proteomics* **179**: 131–139.

Bringel, F., Postema, C.P., Mangenot, S., Bibi-Triki, S., Chaignaud, P., Haque, M.F.U., et al. (2017) Genome sequence of the dichloromethane-degrading bacterium *Hyphomicrobium* sp. strain GJ21. *Genome Announc.* **5**: e00622-17.

Brown, P.J.B., Kysela, D.T., Buechlein, A., Hemmerich, C., and Brun, Y.V. (2011) Genome sequences of eight morphologically diverse *Alphaproteobacteria*. *J. Bacteriol.* **193**: 4567–4568.

Chaignaud, P., Maucourt, B., Weiman, M., Alberti, A., Kolb, S., Cruveiller, S., et al. (2017) Genomic and transcriptomic analysis of growth-supporting dehalogenation of chlorinated methanes in *Methylobacterium*. *Front Microbiol* **8**:.

Chubiz, L.M., Purswani, J., Carroll, S.M., and Marx, C.J. (2013) A novel pair of inducible expression vectors for use in *Methylobacterium extorquens*. *BMC Res. Notes* **6**: 183.

Couñago, R.M., Chen, N.H., Chang, C.W., Djoko, K.Y., McEwan, A.G., and Kobe, B. (2016) Structural basis of thiol-based regulation of formaldehyde detoxification in *H. influenzae* by a MerR regulator with no sensor region. *Nucleic Acids Res.* **44**: 6981–6993.

Figurski, D.H. and Helinski, D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. *Proc. Natl. Acad. Sci. U. S. A.* **76**: 1648–1652.

Fulton, G.L., Nunn, D.N., and Lidstrom, M.E. (1984) Molecular cloning of a malyl coenzyme A lyase gene from *Pseudomonas* sp. strain AM1, a facultative methylotroph. *J. Bacteriol.* **160**: 718–723.

Gälli, R. and Leisinger, T. (1985) Specialized bacterial strains for the removal of dichloromethane from industrial waste. *Conserv. Recycl.* **8**: 91–100.

Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. U. S. A.* **87**: 4645–4649.

Green, P.N. and Ardley, J.K. (2018) Review of the genus *Methylobacterium* and closely related organisms: a proposal that some *Methylobacterium* species be reclassified into a new genus, *Methylorubrum* gen. nov. *Int. J. Syst. Evol. Microbiol.* **68**: 2727–2748.

Gruffaz, C., Muller, E.E.L., Louhichi-Jelail, Y., Nelli, Y.R., Guichard, G., and Bringel, F. (2014) Genes of the N-methylglutamate pathway are essential for growth of *Methylobacterium extorquens* DM4 with monomethylamine. *Appl Env. Microbiol* **80**: 3541–3550.

Henrysson, T. and Mattiasson, B. (1993) A microbial biosensor system for dihalomethanes. *Biodegradation* **4**: 101–105.

Hossaini, R., Chipperfield, M.P., Montzka, S.A., Leeson, A.A., Dhomse, S.S., and Pyle, J.A. (2017) The increasing threat to stratospheric ozone from dichloromethane. *Nat. Commun.* **8**: 15962.

Hovasse, A., Bruneel, O., Casiot, C., Desoeuvre, A., Farasin, J., Hery, M., et al. (2016) Spatio-temporal detection of the *Thiomonas* population and the *Thiomonas* arsenite oxidase involved in natural arsenite attenuation processes in the Carnoulès acid mine drainage. *Front. Cell Dev. Biol.* **4**:.

Kaczmarczyk, A., Campagne, S., Danza, F., Metzger, L.C., Vorholt, J.A., and Francez-Charlot, A. (2011) Role of *Sphingomonas* sp. strain Fr1 PhyR-NepR-σEcfG cascade in general stress response and identification of a negative regulator of PhyR. *J. Bacteriol. J. Bacteriol.* **193**, **193**: 6629, 6629–6638.

Kayser, M.F. and Vuilleumier, S. (2001) Dehalogenation of dichloromethane by dichloromethane dehalogenase/glutathione S-transferase leads to formation of DNA adducts. *J. Bacteriol.* **183**: 5209–5212.

King, N.P., Lee, T.M., Sawaya, M.R., Cascio, D., and Yeates, T.O. (2008) Structures and functional implications of an AMP-binding cystathionine beta-synthase domain protein from a hyperthermophilic archaeon. *J. Mol. Biol.* **380**: 181–192.

Kohler-Staub, D. and Leisinger, T. (1985) Dichloromethane dehalogenase of *Hyphomicrobium* sp. strain DM2. *J. Bacteriol.* **162**: 676–681.

Kozhevnikova, E.N., van der Knaap, J.A., Pindyurin, A.V., Ozgur, Z., van Ijcken, W.F.J., Moshkin, Y.M., and Verrijzer, C.P. (2012) Metabolic enzyme IMPDH is also a transcription factor regulated by cellular state. *Mol. Cell* **47**: 133–139.

La Roche, S.D. and Leisinger, T. (1991) Identification of *dcmR*, the regulatory gene governing expression of dichloromethane dehalogenase in *Methylobacterium* sp. strain DM4. *J. Bacteriol.* **173**: 6714–6721.

Lopes, N., Hawkins, S.A., Jegier, P., Menn, F.M., Sayler, G.S., and Ripp, S. (2012) Detection of dichloromethane with a bioluminescent (*lux*) bacterial bioreporter. *J. Ind. Microbiol. Biotechnol.* **39**: 45–53.

Marx, C.J. (2008) Development of a broad-host-range *sacB*-based vector for unmarked allelic exchange. *BMC Res. Notes* **1**: 1.

McLLean, J.E., Hamaguchi, N., Belenky, P., Mortimer, S.E., Stanton, M., and Hedstrom, L. (2004) Inosine 5′-monophosphate dehydrogenase binds nucleic acids *in vitro* and *in vivo*. *Biochem. J.* **379**: 243–251.

Mielecki, D. and Grzesiuk, E. (2014) Ada response – a strategy for repair of alkylated DNA in bacteria. *FEMS Microbiol. Lett.* **355**: 1–11.

Muller, E.E.L., Bringel, F., and Vuilleumier, S. (2011) Dichloromethane-degrading bacteria in the genomic age. *Res. Microbiol.* **162**: 869–876.

Muller, E.E.L., Hourcade, E., Louhichi-Jelail, Y., Hammann, P., Vuilleumier, S., and Bringel, F. (2011) Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4. *Environ. Microbiol.* **13**: 2518–2535.

R Core Team (2017) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: URL http://www. R-project.org/.

Ritz, C. and Spiess, A.-N. (2008) qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics* **24**: 1549–1551.

Roche, S.D.L. and Leisinger, T. (1990) Sequence analysis and expression of the bacterial dichloromethane dehalogenase structural gene, a member of the glutathione S-transferase supergene family. *J. Bacteriol.* **172**: 164–171.

Sambrook, J. (2001) Molecular cloning: A Laboratory manual, Third edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Schlosser, P.M., Bale, A.S., Gibbons, C.F., Wilkins, A., and Cooper, G.S. (2015) Human health effects of dichloromethane: Key findings and scientific issues. *Environ. Health Perspect.* **123**: 114–119.

Schmid-Appert, M. (1996) Untersuchungen zur regulation des dichlormethan-dehalogenase gens aus *Methylobacterium* sp. stamm DM4 und struktur des angrenzenden DNA-region, thèse de l'École polytechnique fédérale de Zurich.

Schmid-Appert, M., Zoller, K., Traber, H., Vuilleumier, S., and Leisinger, T. (1997) Association of newly discovered IS elements with the dichloromethane utilization genes of methylotrophic bacteria. *Microbiol. Read. Engl.* **143 (Pt 8)**: 2557–2567.

Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**: 1101–1108.

Scholtz, R., Wackett, L.P., Egli, C., Cook, A.M., and Leisinger, T. (1988) Dichloromethane dehalogenase with improved catalytic activity isolated from a fast-growing dichloromethane-utilizing bacterium. *J. Bacteriol.* **170**: 5698–5704.

Shestakova, M. and Sillanpää, M. (2013) Removal of dichloromethane from ground and wastewater: A review. *Chemosphere* **93**: 1258–1267.

Škulj, M., Okršlar, V., Jalen, Š., Jevševar, S., Slanc, P., Štrukelj, B., and Menart, V. (2008) Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. *Microb. Cell Factories* **7**: 6.

Stourman, N.V., Rose, J.H., Vuilleumier, S., and Armstrong, R.N. (2003) Catalytic mechanism of dichloromethane dehalogenase from *Methylophilus* sp. Strain DM11. *Biochemistry* **42**: 11048–11056.

Tsai, W.-T. (2017) Fate of chloromethanes in the atmospheric environment: Implications for human health, ozone formation and depletion, and global warming impacts. *Toxics* **5**:.

Vuilleumier, S., Chistoserdova, L., Lee, M.C., Bringel, F., Lajus, A., Zhou, Y., et al. (2009) *Methylobacterium* genome sequences: a reference blueprint to investigate microbial metabolism of C1 compounds from natural and industrial sources. *PLoS ONE* **4**: e5584.

Vuilleumier, S. and Leisinger, T. (1996) Protein engineering studies of dichloromethane dehalogenase/glutathione S-transferase from *Methylophilus* sp. strain DM11. Ser12 but not Tyr6 is required for enzyme activity. *Eur. J. Biochem.* **239**: 410–417.

4.3 Annexe du chapitre 4 : données non présentées dans l'article

Cette annexe regroupe un ensemble d'approches expérimentales et de résultats associés à l'étude de la régulation par DcmR qui ne sont pas présentés dans l'article. Ces résultats sont présentés en quatre axes thématiques :

i) Protéine DcmR portant une étiquette six-histidine : clonage, purification, cristallisation et complémentation de fonction d'un mutant *dcmR*

ii) Fixation de la protéine DcmR à l'ADN : technique de « retards sur gel » avec les fragments d'ADN portant les promoteurs des gènes *dcmR* et *dcmA*

iii) Détection de la production de la protéine DcmA par *western blot*

iv) Construction de plasmides bi-rapporteurs : transfert chez *M. extorquens* DM4 par conjugaison

4.3.1 Matériel et méthodes

4.3.1.A Production et purification de la protéine DcmR de *M. extorquens* DM4 avec une étiquette pour la purification

Pour produire la protéine DcmR avec une étiquette en N ou C terminal dans le contexte hétérologue d'*E. coli*, le gène *dcmR* a été cloné dans les plasmides pET16b::10His-*dcmR* (pME8308) et pET28b::*dcmR*-6His (pME8323) pour produire la protéine avec une étiquette en N ou C terminal chez *E. coli*, respectivement (Fig. 2). Pour l'expression hétérologue chez *E. coli* BL21 (DE3) portant les plasmides pME8308 ou pME8323, a été cultivée à 30 °C avec une agitation de 120 rpm dans un Erlenmeyer de 2,5 L avec 0,8 L de milieu LB supplémenté par un antibiotique (ampicilline 50 $\mu g.\mu L^{-1}$ ou kanamycine 50 $\mu g.\mu L^{-1}$) à partir d'un ensemencement avec une préculture de 5 mL d'une nuit à 30 °C du même milieu. Lorsque la densité optique à 600 nm atteint environ 0,5, l'expression de la protéine est induite pendant trois heures par ajout d'une solution d'IPTG à 1 M (concentration finale de 1 mM). Les cellules sont récoltées par une centrifugation de 10 minutes à 4 000 rpm et 4 °C (Sorvall® Super T21). Les culots cellulaires sont conservés à -25 °C.

Pour l'expression de la protéine DcmR dans le contexte homologue chez *M. extorquens* DM4, le plasmide d'expression pCM160 a été utilisé (Lidstrom & Marx, 2001). Le gène *dcmR* est cloné

pour être exprimé à partir du promoteur fort et constitutif du gène *mxaF* (Chaignaud *et al.*, 2017). Le plasmide pME8322 (pCM160::*dcmR*-6His) a ainsi été construit pour produire la protéine DcmR-6His (Fig. 1). Pour l'expression homologue chez la souche de *M. extorquens* DM4 STOP*dcmR* portant le plasmide pME8322, une culture dans un Erlenmeyer de 5 L est ensemencée à une DO_{600nm} de 0,02 dans 1,6 L de milieu M3 supplémenté avec une source de carbone (méthanol 80 mM) et par un antibiotique (kanamycine 50 $\mu g.\mu L^{-1}$). La culture est incubée à 30 °C sous agitation. Les cellules sont récoltées lorsqu'une DO_{600nm} de 0,5 (environ 24 h de culture) est obtenue. Les culots cellulaires sont conservés à -25 °C.

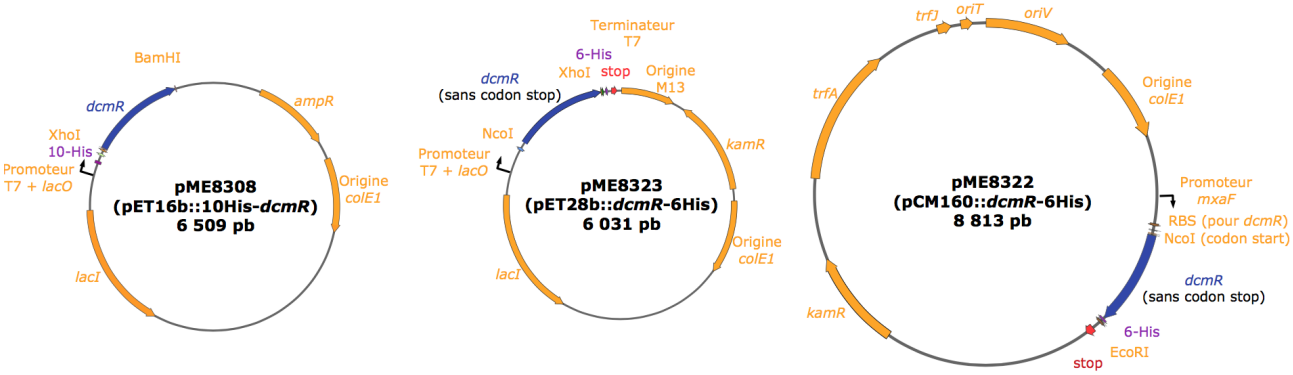


Figure 4.2: Cartes génétiques des plasmides construits en vue de la production de DcmR avec une étiquette histidine.
Le gène *dcmR* (en bleu) a été amplifié par PCR à partir du génome de *M. extorquens* DM4. Les plasmides pME8308 et pME8323 sont basés sur les plasmides pET16b (Merck™ Novagen) et pET28b (Merck™ Novagen), respectivement. La protéine LacI réprime l'expression du gène cloné en reconnaissant l'opérateur *lacO* présent au niveau du promoteur T7. Cette répression est levée en présence d'IPTG chez *E. coli*. Le plasmide pME8322 a été obtenu par le clonage du gène *dcmR* dans le site de clonage du plasmide pCM160, un plasmide qui peut se répliquer à la fois chez *E. coli* et *M. extorquens* DM4 et possède les gènes nécessaires à la conjugaison bactérienne (Fulton *et al.*, 1984). Le gène d'intérêt est exprimé de façon constitutive avec le promoteur *mxaF*. Les sites de restrictions utilisés pour les clonages sont indiqués sur les plasmides.

La protéine DcmR est purifiée à partir du culot cellulaire décongelé. Le culot est resuspendu avec 1 mL de lysozyme à 10 $mg.mL^{-1}$ dans du tampon (Tris 50 mM ; 300 NaCl ; pH 8,0), puis le volume est ajusté à 10 mL avec la même solution tampon. De l'antiprotéase (*Complete Protease inhibitor cocktail tablet* ; Roche) est ajoutée dans les 10 mL de suspension cellulaire selon les indications du fabricant. La suspension est incubée 30 minutes à 37 °C pour permettre la lyse cellulaire par le lysozyme. Une seconde étape de lyse cellulaire est réalisée par cinq cycles de sonication de

30 secondes (amplitude 30 % ; Vibra-cell 75041 Bioblock Scientific) en maintenant les tubes dans la glace. Un volume de 10 µL est récupéré comme témoin de lysat total. Le lysat est centrifugé à 4 °C pendant 30 minutes à 9 500 g puis filtré (0,22 µm) pour éliminer les cellules non lysées et les plus gros débits cellulaires. Une solution d'imidazole à 5 mM est ajoutée dans le surnageant pour limiter les fixations aspécifiques des protéines. Un volume de 800 µL de résine contenant du nickel (Protino® NiNTA agarose ; Macherey Nagel) est déposé dans une colonne (Pierce® *centrifuge columns*) puis équilibré par deux étapes de lavage avec 20 mL d'eau ultra-pure puis 20 mL de tampon (Tris 50 mM ; 300 NaCl ; pH 8,0). . Le lysat cellulaire filtré est versé trois fois de suite à travers la colonne pour optimiser la fixation de la protéine d'intérêt. La colonne est ensuite lavée avec 20 mL de solution d'imidazole croissante allant de 5 ; 20 ; 40 à 60 mM. L'élution est réalisée avec 5 mL de solution d'imidazole croissante allant de 100 ; 200 ; 300 à 400 mM. Au cours de la purification, la quantité de protéine éluée est estimée en déposant 15 µL de surnageant avec 5 µL de solution de Bradford. Les protéines sont conservées à 4 °C. La colonne est lavée avec de l'eau ultra-pure puis avec de l'éthanol à 20 % pour conserver la résine. La pureté des différentes fractions de protéines éluées est validée par une électrophorèse SDS-PAGE (tableau n°1).

Tableau n°1 : Préparation d'un gel SDS-PAGE

Solution	Séparation (15 %)	Concentration (4 %)
H_2O ultra-pure	2,268	3
Acrylamide 30 % (Rotiphorese® gel 30)	5	0,665
Tris HCl 1,5 M et pH 8,8	2,5	/
Tris HCl 0,5 M et pH 6,8	/	1,25
SDS 10 %	0,1	0,05
Ammonium persulfate 10 % (APS)	0,12	0,06
Tétraméthyléthylènediamine (TEMED)	0,012	0,006

Les protéines DcmR purifiées ont été quantifiées à l'aide du kit de dosage Qubit. Des tests ont été réalisés pour concentrer les protéines à l'aide de filtres (Amicon 10 K, Millipore ; *Protein concentrator*, Pierce) ou par dialyse (Slide A lyzer™ ThermoFischer). Aucune de ces méthodes n'a permis de concentrer les protéines purifiées.

Une étape de chromatographie d'exclusion stérique permet d'affiner la purification des protéines et d'estimer leur taille et leur état d'oligomérisation. La chromatographie d'exclusion stérique est basée sur l'utilisation d'une colonne ÄKTA de 25 mL (Superdex 200 *Increase* 10/300 GL ; Gelifesciences) montée sur un appareil BioLogic Duoflow *chromatography apparatus* (Bio-Rad) avec des adaptateurs ÄKTA *adapter to* Bio-Rad *material* (Bio-Rad 1/16" OD 1/4–28 *fitting*). L'ensemble des appareils est placé à 4 °C. La pression maximale de la colonne est fixée à 435 psi. La formation de bulles d'air dans la solution d'élution est évitée au maximum durant la chromatographie. Le liquide dans la colonne est en permanence en mouvement pour éviter des dépôts de sels. La colonne est tout d'abord lavée avec de l'eau ultra-pure filtrée (0,22 µm) représentant trois à cinq fois le volume de la colonne, à un débit de 0,25 $mL.min^{-1}$ pendant six à huit heures. La colonne est ensuite lavée avec du tampon d'élution (Tris HCl 50 mM ; NaCl 300 mM ; pH 7,0 ; filtré à 0,22 µm) pendant six à huit heures. Le tampon d'élution est injecté par la tête d'injection en amont de colonne. La solution (1 mL) de protéine est ensuite injectée de la même façon. L'élution est suivie par mesure de l'absorbance UV (280 nm). Les éluats sont récupérés dans des microtubes maintenus dans la glace. Après la chromatographie, la colonne est lavée avec de l'eau puis conservée dans de l'éthanol à 20 %. La colonne a été calibrée avec une solution de calibration contenant des protéines de taille connues incluant de la thyroglobuline 670 kDa, de la γ-globuline 158 kDa, de l'ovalbumine 44 kDa et de la myoglobine 17 kDa ainsi que de la vitamine B12 1,35 kDa qui n'est pas une protéine.

4.3.1.B Retard sur gel avec la protéine DcmR

Des expériences de retard sur gel ont été tentées pour mettre en évidence la fixation de DcmR sur les sites de fixation prédits (La Roche & Leisinger, 1991) dans la région intergénique entre *dcmR* et *dcmA*. Plusieurs fragments d'ADN amplifiés par PCR et purifiés (kit Nucleospin PCR *and gel clean-up*) ont été utilisés comme cible : la région en amont de *dcmA* (2562623-2563049 ; les chiffres indiquent la position sur la séquence du chromosome chez *M. extorquens* DM4), la région en amont de *dcmR* (2562631-2562165) et l'ensemble de la région intergénique entre *dcmR* et *dcmA*, (2562165-2563049). La région en amont de *rpoA* (position 287694-2876963) a été utilisée comme témoin négatif vu que ce gène n'est pas régulé en présence de DCM (Chaignaud *et al*., 2017). Les fragments d'ADN à 8 ng ont été mélangés à 500, 1 000 et 2 000 ng de protéines 10His-DcmR et DcmR-6His produites chez *E. coli*. Le milieu réactionnel (20 µL final) contient 4 µL de solution de

fixation 5 x (Tris HCl 10 mM, pH 7,5 ; 50 mM NaCl; 0,5 mM DTT ; 1 mM $MgCl_2$; glycérol 2,5 %). Cette solution est incubée 30 minutes à 30 °C avec ou sans DCM à 10 mM. Après incubation, un volume de 2 µL de solution de charge (Tris HCl 25 mM, pH 7,5 ; bleu de bromophénol 0,2 % ; glycérol 4 %) est ajouté. La solution est mélangée doucement (sans l'agitateur vortex) pour éviter de dissocier les complexes. Après une pré-électrophorèse de 40 minutes à 100 V sans échantillon, une électrophorèse est réalisée dans un gel d'acrylamide 6 % qui est mis à migrer à 100 V pendant 60 min. Après électrophorèse, le gel est mis dans une cuve contenant du bromure d'éthidium (BET) puis le gel est observé sous exposition UV.

4.3.1.C Cristallisation de la protéine DcmR

Un essai de cristallisation a été réalisé avec la protéine 10His-DcmR après purification par chromatographie d'exclusion stérique. Un robot Mosquito Crystal : TTPLabtech est utilisé pour distribuer la solution de protéine dans des kits de cristallisation (JBScreen JCSG++, CS-206L, Jena Bioscience ; Index HT, HR2134, Hampton research). Pour cela 0,1 µL de solution de protéine à 200 µL et 0,1 µL de solution de cristallisation. Les plaques sont mises à incubées à 30 °C sans agitation. Une loupe binoculaire avec un filtre polarisant (NIKON SMZ1500) est utilisée pour observer si des cristaux se forment.

4.3.1.D Étude simultanée des promoteurs de *dcmA* et *dcmR* avec des plasmides bi-rapporteurs

Dans le but de faciliter l'étude des promoteurs de *dcmR* et *dcmA*,et d'étudier l'impact du sur-enroulement de l'ADN lors de la transcription de ces gènes divergents, la construction d'un plasmide bi-rapporteur a été envisagée. Ce plasmide devait permettre de mesurer en même temps la fluorescence des promoteurs de *dcmR* et de *dcmA*. Le plasmide pPF1 (Masulis *et al.*, 2015) est un plasmide bi-rapporteur avec deux gènes rapporteurs en orientation inverse qui a servi de base pour la construction d'un plasmide pouvant se répliquer non seulement chez *E. coli* mais aussi chez *M. extorquens* DM4. Un clonage par amplification PCR et digestion enzymatique a été tenté mais l'amplification PCR de la région contenant les deux gènes rapporteurs a systématiquement conduit à des chimères, probablement dû aux structures répétées. Un autre modèle de plasmide a donc été envisagé et construit par la société e-Zyvec (France ; Lille) en utilisant une technologie brevetée basée sur la méthode de clonage Gibson (Gibson *et al.*, 2009).

Le plasmide bi-rapporteur pME8312 est composé du corps du plasmide pLM-sYFP2 qui contient l'origine de réplication, le gène de sélection (résistance à la kanamycine), les gènes de mobilisation et conjugaison ainsi que le gène rapporteur *yfp* (Kaczmarczyk *et al.*, 2011). Le gène rapporteur mCherry est ajouté en orientation divergente du gène *yfp*. Un site de clonage (n°1) est ajouté entre ces deux gènes rapporteurs (qui ne possèdent pas de promoteurs). Dans le site de clonage n°1, le site de fixation du ribosome (RBS) est GGAGG, choisi sur la base du motif GAGGAG présent chez 176 des 697 séquences 5'UTR avec un codon start expérimentalement validé (Bibi-Triki *et al*., 2018). En aval, trois codons stop dans les trois cadres de lecture ont été insérés. Le codon UAA a été utilisé bien qu'il soit le moins présent dans le génome de *M. extorquens* DM4 : UGA (3 910) > UAG (904) > UAA (640) (données fournies par la plateforme MaGe du Genoscope). Cette répartition est en accord avec la littérature où le codon stop UAA est peu retrouvé dans le génome de bactérie ayant un haut pourcentage en GC (Korkmaz *et al.*, 2014) comme c'est le cas de *M. extorquens* DM4 qui a un contenu en GC de 68 % (Vuilleumier *et al*., 2009). Toutefois, c'est le codon stop avec la meilleure efficacité car il est reconnu par les protéines *release factor* I et II et possède un faible taux de translecture (Wei *et al.*, 2016). Les séquences nucléiques sont légèrement différentes de part et d'autres des sites de restriction pour éviter les recombinaisons entre les deux parties du site de clonage.

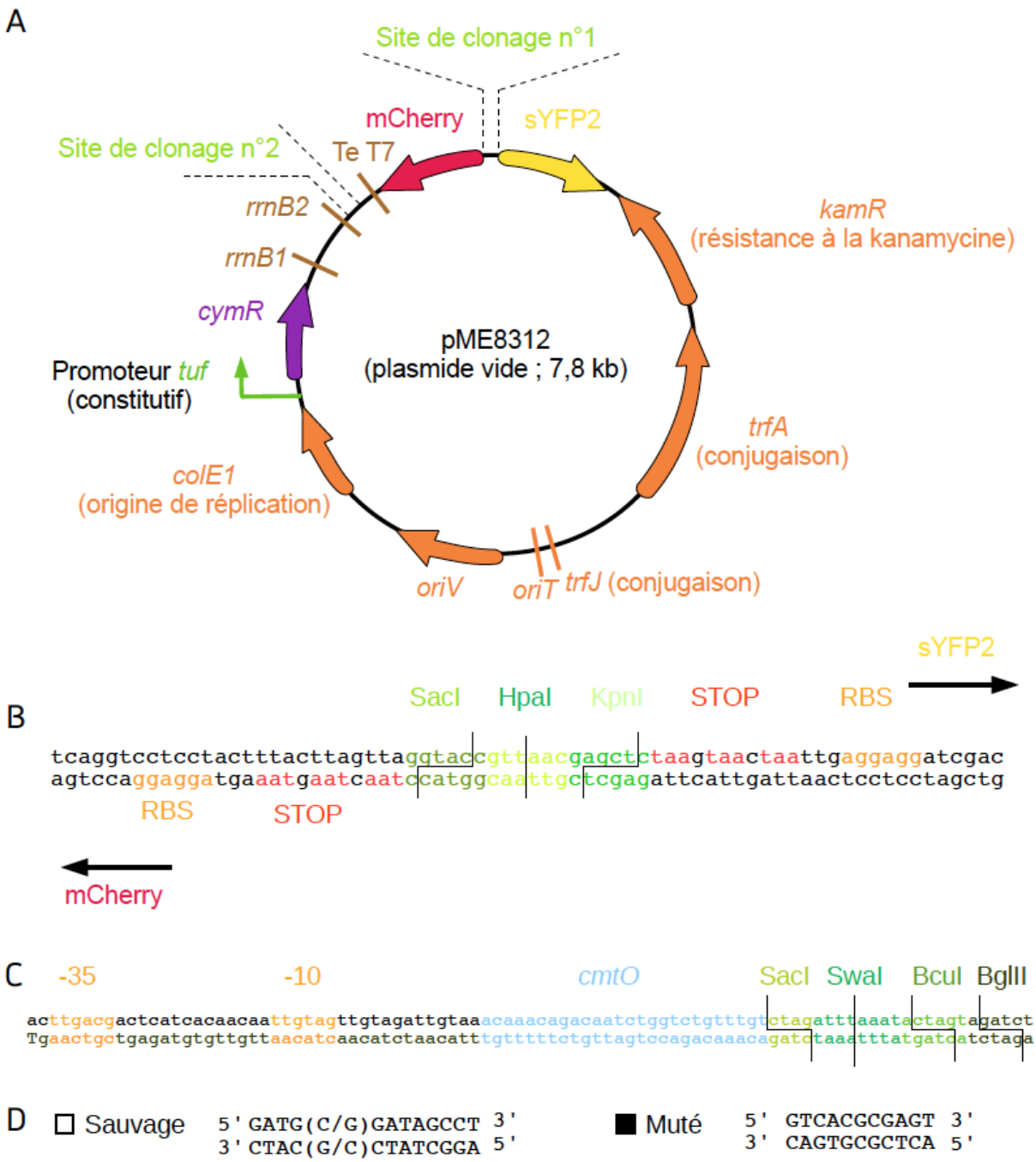


Figure 4.3: Carte génétique du plasmide pME8312.
A Représentation du plasmide avec ses deux gènes rapporteurs (mCherry et YFP), et les sites de clonages n°1 et 2. Les enzymes de restriction pouvant être utilisées pour ces deux sites de clonage sont indiquées ainsi que la position des sites de clivage. B Le site de clonage n° 1 est situé entre les deux gènes rapporteurs. Ce site est entouré par des codons STOP pour éviter la formation d'une protéine de fusion tandis que des séquences RBS permettent la fixation du ribosome. C Site de clonage n° 2 permettant l'expression d'un gène sous contrôle du facteur de transcription CymR (inductible au cumate ; (Chubiz *et al.*, 2013)). Ce site ne possède pas de séquence RBS, elle doit donc être apportée avec le gène d'intérêt. D Mutation de la séquence conservée entre les boites -35 et -10 des promoteurs de *dcmA* et *dcmR*.

Un second site de clonage a été prévu dans le plasmide pME8312 en vue de permettre l'expression d'un gène d'intérêt. Cette expression est régulée par le gène *cymR* qui code un facteur de transcription se fixant au niveau d'une séquence *cmtO* pour inhiber l'expression du promoteur associé (Chubiz *et al.*, 2013). Cette inhibition peut être levée par l'ajout de cumate. Ce système d'induction fonctionne chez *M. extorquens* AM1 (Chubiz *et al.*, 2013). Dans le plasmide pME8312, ce facteur de transcription est exprimé avec le promoteur du gène *tuf* (facteur d'élongation Tu) de *M. extorquens* DM4 (position 5371604 à 5371511 dans le génome de DM4). Ce promoteur a été choisi car il est exprimé de façon constitutive chez *M. extorquens* AM1 (Schada von Borzyskowski *et al.*, 2015) et chez *M. extorquens* DM4 (Chaignaud *et al*., 2017 ; Bibi-Triki *et al*., 2018). Un terminateur de la transcription (*rrnB* ; (Orosz *et al.*, 1991)) se trouve entre les gènes *cymR* et *dcmR* pour éviter que la transcription du premier entraine celle du second. En aval de ce site (et donc de *dcmR*) se trouve le terminateur de la transcription Te (*early terminator*) du bactériophage T7, indépendant du facteur rho (Briat & Chamberlin, 1984).

Tableau 2 : Liste des caractéristiques et des variants du plasmide pME8312 bi-rapporteurs

Nom	pME8312	pME8313	pME8314	pME8315	pME8316	pME8317	pME8318	pME8319	pME8320
Propriété des plasmides									
Taille (en kb)	7,8	8,4	8,5	9,2	8,4	8,4	8,4	8,4	9,2
Insert dans le site de clonage n°1	/	*dcmR* → *dcmA*	/	*dcmR* → *dcmA*	*dcmA* → *dcmR*	*dcmR* → *dcmA*	*dcmR* → *dcmA*	*dcmR* → *dcmA*	*dcmR* → *dcmA*
boite de fixation *dcm*	/	sauvage	/	sauvage	sauvage	P'_{R1}	P'_{A}	P'_{R1} et P'_{A}	P'_{R1} et P'_{A}
Insert dans le site de clonage n° 2	/	/	*dcmR*	*dcmR*	/	/	/	/	*dcmR*
Amorces pour vérifier les plasmides[a]									
MCherry1 / sYFP-r (taille en kb)	1,1	1,8	1,1	1,8	1,8	1,8	1,8	1,8	1,8
Mcherry2 / MF32 (taille en kb)	1,9	1,9	2,7	1,9	1,9	1,9	1,9	1,9	2,7
kanF1 / kanR2 (taille en kb)	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6

[a] Pour les PCR de vérification de la présence des différents plasmides, les couples d'amorce Mcherry/sYFPr pour le site de clonage n°1, Mcherry2/MF32 pour le site de clonage n°2 et couple d'amorce PCR kanF1/kanr2 pour la présence du plasmide, ont été utilisées.

Pour introduire ces plasmides chez *M. extorquens* DM4, une conjugaison tri-parentale a été tentée utilisant comme souche donneuse, le plasmide bi-rapporteur présent chez *E. coli* DH5α, une souche *Helper* de *E. coli* portant le plasmide pRK2013 (Figurski & Helinski, 1979) et *M. extorquens* DM4 sauvage comme souche receveuse d'après un protocole bien établi (Fulton *et al.*, 1984). Après conjugaison chez *M. extorquens* DM4 sauvage, la présence du plasmide est validée par plusieurs couples d'amorces (Tableau 2).

4.3.1.E Détection de la protéine DcmA par *western blot*

Pour déterminer la production de la protéine DcmA extraite de *M. extorquens* DM4 sauvage ou mutée notamment pour le gène *dcmR*, un *western blot* a été réalisé. Ces souches ont été cultivées en phase exponentielle de croissance (DO_{600nm} de 0,10 à 0,18) puis centrifugées 10 minutes à 11 600 g et 4 °C. Les culots cellulaires sont maintenus dans la glace et resuspendus dans un volume d'eau ultra-pure, à ajuster pour chaque culture de façon à normaliser la DO_{600nm} à une valeur de quatre pour avoir la même quantité de cellules dans tous les échantillons. Ces suspensions cellulaires sont à nouveau centrifugées puis les culots sont resuspendus dans 600 µL d'eau ultra-pure. Un gramme de bille de verre (0,1 mm, BioSpec) est ajouté à chaque échantillon. Les cellules sont lysées mécaniquement par trois cycles d'une minute d'agitation dans un agitateur mécanique (Retsch MM2). Après chaque cycle d'agitation, les tubes sont placés une minute dans la glace. Un volume de 150 µL de tampon de Laemmli (5 x) est ajouté à chaque tube. Le tampon de Laemmli (5 x) est composé de 2 mL de Tris-HCl à 1,5 M et pH 6,8, un gramme de SDS, 5 mg de bleu de bromophénol, 5 mL de glycérol, 3 mL d'eau ultra-pure et 50 µL de beta-mercaptoéthanol. Les tubes contenant les extraits protéiques sont ensuite portés à ébullition une minute dans un bain-marie, puis centrifugés 10 minutes à 6 600 g et 4 °C. Les protéines sont séparées dans un gel de polyacrylamide 12 % (Amersham ECL gel 12 % précoulé) en suivant les indications du fabricant. Un volume de 10 µL de marqueur de taille (PageRuler *prestained protein ladder* 10 *to* 180 kDa ; Life Technologies) est déposé sur le gel ainsi que 15 µL d'échantillon. Une migration d'une heure à 120 V est réalisée dans une cuve d'électrophorèse adaptée à ces gels précoulés (Amersham ECL *gel box*). Pour une détection directe des protéines par coloration au bleu de Coomassie, le gel est plongé pendant trois à quatre heures dans une solution de fixation (acide acétique 7 % et éthanol 40 %) avec une agitation lente, et ensuite laissé la nuit dans un bain de bleu de Coomassie (bleu brillant G

colloïdal 25 % et éthanol 75 %). Le gel est finalement décoloré pendant trois à quatre heures dans une solution de décoloration (acide acétique 10 % et éthanol 25 %).

Les protéines séparées selon leur taille par électrophorèse dans un gel de polyacrylamide peuvent être transférées sur une membrane en vue d'une révélation spécifique de protéines à l'aide d'un anticorps. Après migration, le gel est équilibré 15 minutes avec du tampon Tris-glycine (pH 8,5 ; Tris 25 mM et glycine 192 mM). Pendant ce temps, une membrane de polyfluorure de vinylidène (PVDF ; *Low fluorescence transfert membrane* 0,2 µm ; Pierce) est équilibrée une minute dans un bain de méthanol, lavée une minute dans un bain d'eau ultra-pure, puis équilibrée 15 minutes dans du tampon Tris-glycine. Le transfert des protéines sur la membrane se fait selon la méthode semi-sèche à l'aide d'un appareil Amersham TE 77 *semi-dry transfer unit* selon les recommandations du fabricant. Le transfert dure une heure lorsque l'on applique un courant de 0,8 $mA.cm^{-2}$. Après transfert, la membrane est plongée une heure avec une agitation lente dans une solution de blocage (SuperBlock (PBS) *blocking buffer* ; Pierce) afin de limiter la fixation aspécifique des protéines à la membrane. Après cinq lavages de cinq minutes dans du PBS avec du Tween 20 à 0,1 %, l'anticorps primaire dilué au 1 / 500 est ajouté dans un bain de tampon PBS-Tween 20 à 0,1 %, pendant une nuit à 4 °C avec une agitation lente. Un sérum de lapin anti-DcmA (obtenu par (Kohler-Staub *et al.*, 1986)) est utilisé. L'excès de ce sérum est éliminé par cinq lavages de deux à cinq minutes dans du PBS-Tween 20 à 0,1 %. Un anticorps secondaire de chèvre dirigé contre les IgG de lapin (*goat anti-rabbit* IgG (H+L) cross adsorbed secondary antibody ; Pierce) est ajouté à une dilution de 1 / 5 000 dans un bain de tampon PBS-Tween 20 à 0,1 %. La membrane est protégée de la lumière pour préserver la fluorescence du fluorochrome (DyLight® 488) couplé à l'anticorps secondaire lors d'une incubation d'une heure avec une agitation lente. La membrane est ensuite lavée selon la même méthode que pour l'anticorps primaire, mais à l'abri de la lumière. Pour la révélation, le fluorochrome est excité à 488 nm, la fluorescence émise est mesurée à 520 nm et numérisée avec une résolution de 100 µm par pixel à l'aide d'un Typhoon scanner Trio (Ge Healthcare). Le voltage du tube photo-multiplicateur est réglé à 400 V. Après révélation de la fluorescence, le gel d'électrophorèse et la membrane peuvent être colorés au bleu de Coomassie pour vérifier que le transfert s'est bien déroulé et que la quantité de protéine déposée était similaire dans les différents puits.

4.3.2 Résultats et discussions

4.3.2.A Résultats de la production de DcmR et des approches associées

La protéine DcmR a pu être produite avec une étiquette histidine en N ou C terminale. Cela a été confirmé notamment par gel SDS-PAGE après purification (Fig. 4). Suite à une recombinaison ayant eu lieu lors du clonage chez *E. coli*, un plasmide produisant juste le domaine MEDS de DcmR a pu être isolé (10His-DcmRΔ1-80). La protéine correspondante a été utilisée comme témoin négatif dans des expériences de retard sur gel.

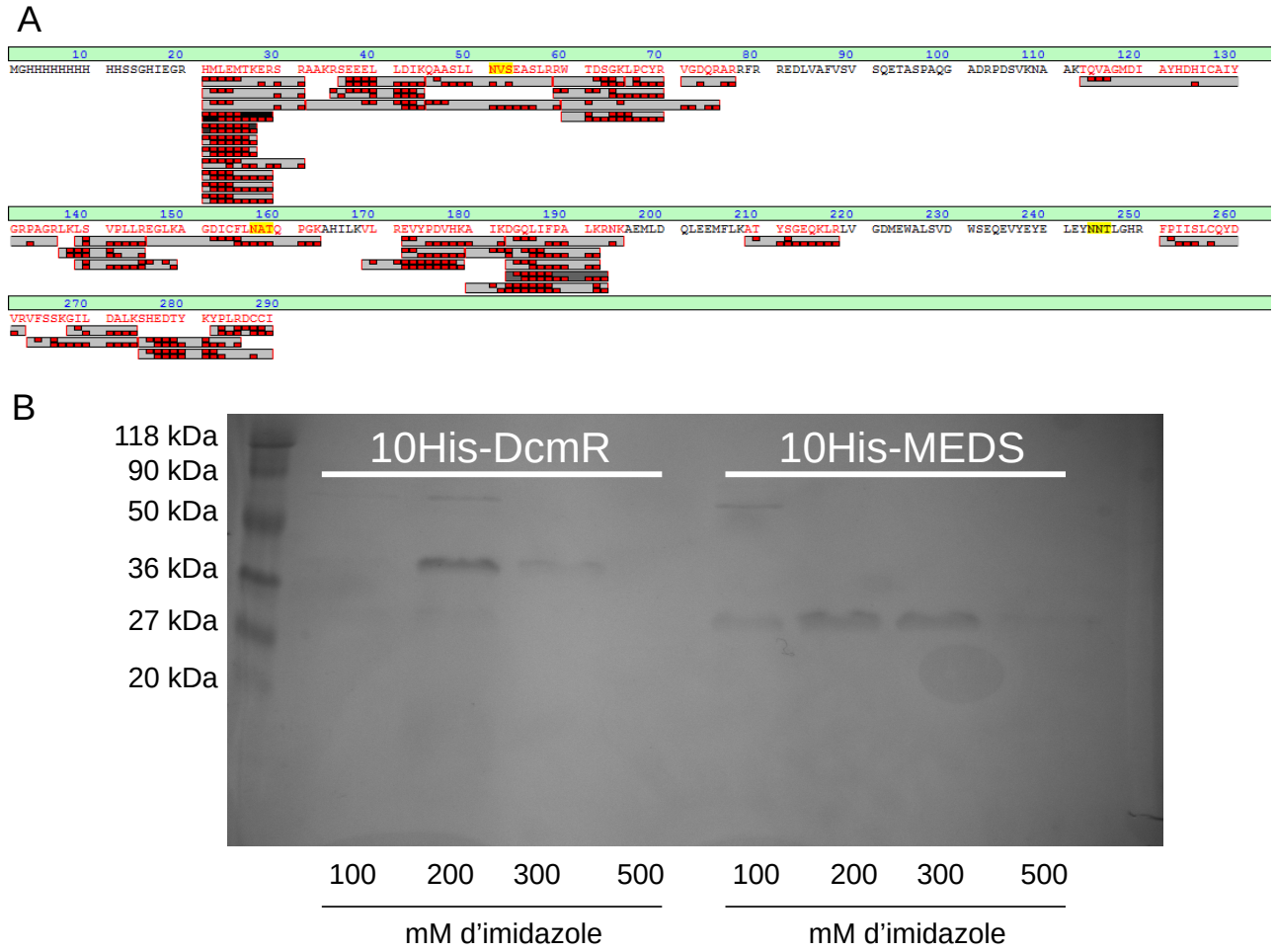


Figure 4.4: Analyse de la protéine 10His-DcmR produite chez *E. coli*.
A Confirmation par spectrométrie de masse que la protéine purifiée est bien DcmR. Les rectangles en dessous de la séquence protéique correspondent aux peptides qui ont été détectés. B Gel d'électrophorèse SDS-PAGE, confirmant la taille attendue de la protéine 10His-DcmR (32,96 kDa) et celle de la protéine tronquée 10His-MEDS (23,19 kDa). Pour chaque dépôt, la concentration d'imidazole qui a servi à l'élution est indiquée.

La protéine 10His-DcmR a été purifiée une seconde fois par chromatographie d'exclusion stérique, en condition non dénaturante. D'après l'estimation par un marqueur de taille, lors de la chromatographie d'exclusion stérique, la protéine 10His-DcmR aurait une taille inférieure à 1,35 kDa. Cela ne correspond pas à la taille attendue de DcmR qui est de 33 kDa. Une co-migration avec un marqueur interne comme l'acétone et le bleu dextran aurait permis de valider la taille estimée de la protéine DcmR produite. Pour vérifier la présence de la protéine DcmR, des éluats de cette chromatographie ont été déposés dans un gel SDS-PAGE (Fig. 4). Le pic observé par chromatographie d'exclusion stérique correspond bien à une protéine d'une taille de 33 kDa. Ces résultats suggèrent que la protéine 10His-DcmR purifiée est dénaturée (et non pas dégradée car elle possède la taille attendue). La présence d'une étiquette pourrait modifier la structure de DcmR, notamment en empêchant la protéine de prendre sa bonne conformation lors de sa synthèse. La dénaturation pourrait aussi résulter de la production en système hétérologue ou de l'utilisation d'un tampon non adapté lors de la purification. Plusieurs améliorations pourraient être apportées au protocole de purification telle qu'une étape de précipitation avec du sulfate d'ammonium pour concentrer les protéines vu que les colonnes commerciales et la dialyse n'ont pas fonctionné.

Plusieurs essais de retard sur gel ont été tentés dans différentes conditions avec la protéine 10His-DcmR et DcmR-6His, en absence de ligand ou en présence de DCM ou de méthanol. Aucun retard n'a été détecté pour le fragment d'ADN contenant la région intergénique *dcmR-dcmA* ou des versions tronquées de celle-ci. Cela est probablement lié à un ensemble d'éléments : la protéine DcmR avec une étiquette histidine n'est peut-être pas fonctionnelle, elle a pu être dénaturée lors de sa purification ; les conditions de fixation de DcmR sur l'ADN n'étaient pas optimales ; ou un défaut de sensibilité de visualisation des bandes, qui pourrait être améliorée par des techniques utilisant des sondes fluorescentes (Wang & Reed, 2012). Il serait également possible de répéter ces expériences avec une protéine DcmR extraite de lysats cellulaires d'*E. coli* portant le plasmide pME8315 puisqu'une répression du promoteur de *dcmA* est observée pour cette souche produisant DcmR sans étiquette.

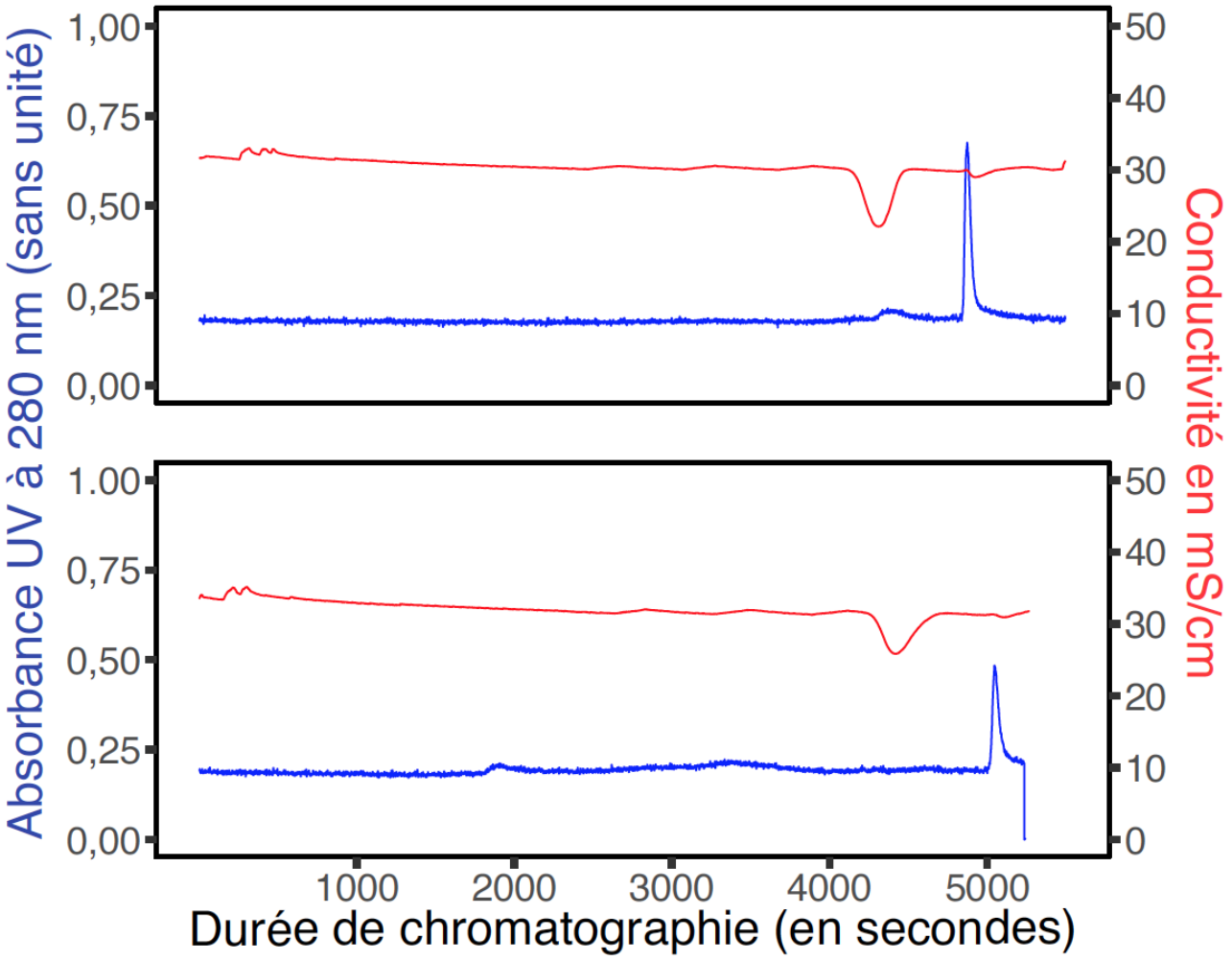


Figure 4.5: Profil de chromatographie d'exclusion stérique de la protéine 10His-DcmR.
La quantité de protéine est indiquée par l'absorbance à 280 nm. La conductivité est corrélée avec la concentration en sel dans le tampon. Les profils du haut et du bas correspondent respectivement à 370 et 570 µg.mL^{-1} de protéine injectée.

Pour vérifier si la protéine 10His-DcmR et DcmR-6His sont actives chez *M. extorquens* DM4, une complémentation de fonction a été réalisée chez le mutant STOP*dcmR* de *M. extorquens* DM4 à l'aide du plasmide pCM160 portant *dcmR* sous contrôle du promoteur *mxaF* (Fig. 6). Les résultats de croissance de la souche portant pCM160 sans insert sont très variables. Cette variation pourrait être due à un problème de stabilité du plasmide, malgré la présence de kanamycine comme marqueur de sélection. Cela pourrait être confirmé par qPCR. Néanmoins l'expression plasmidique de DcmR-6His ou 10His-DcmR ne restaure pas la croissance du mutant *dcmR* à un niveau similaire à celui de la souche sauvage (9,6 ± 0,6 h, avec 15 mM de DCM). Un témoin positif de complémentation pourrait être réalisé avec le gène *dcmR* natif sans étiquette exprimé à partir du plasmide pCM160.

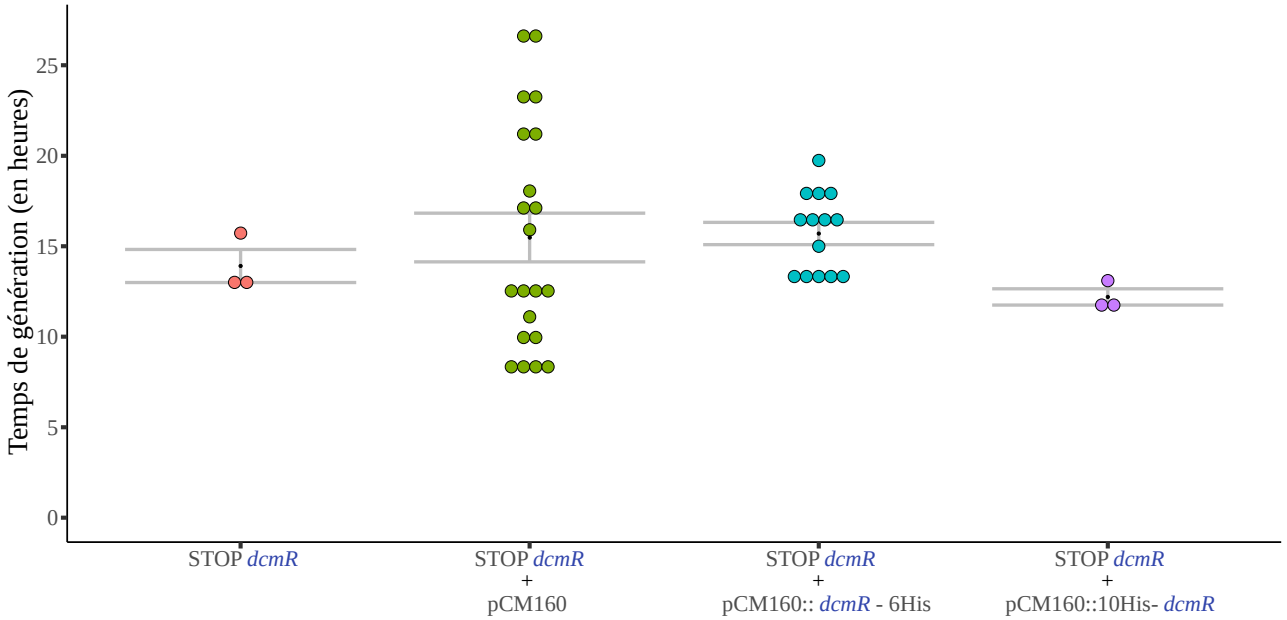


Figure 4.6: Complémentation de fonction avec 10His-DcmR ou DcmR-6His chez le mutant STOP*dcmR*.
Les cultures sont réalisées en présence de DCM 15 mM chez le mutant STOP*dcmR*. Le mutant STOP*dcmR* sans plasmide est cultivé en absence de kanamycine.

L'ensemble des résultats de chromatographie d'exclusion stérique, de retard sur gel et de complémentation de fonction tend à indiquer que les protéines DcmR avec une étiquette histidine ne sont pas fonctionnelles. Cela pourrait venir d'un encombrement stérique à cause de l'étiquette ou de mauvaises conditions de production.

4.3.2.B Conjugaison des plasmides bi-rapporteur chez *M. extorquens* DM4

Les plasmides bi-rapporteurs ont été utilisés avec succès chez *E. coli* (voir figure 3 de l'article) et leur utilisation chez *M. extorquens* DM4 était prévue. Or, il n'a pas été possible de les introduire par conjugaison chez la souche WT ou des mutants comme la souche Δ*dcm* (pour éviter les recombinaisons éventuelles entre les copies chromosomiques et plasmidiques des régions clonées). Après conjugaison, la présence du gène *kamR* porté par ces plasmides est systématiquement retrouvée dans les cellules transconjugantes. Par contre, il n'a pas été possible de détecter les deux sites de clonage malgré une optimisation du protocole de PCR d'amplification des régions clonées.

Pour confirmer ces résultats la fluorescence émise par les gènes rapporteurs YFP et mCherry a été mesurée lors de la croissance cellulaire par fusion transcriptionnelle. Cette fluorescence n'est pas détectée dans la majorité des souches malgré la présence du gène *kamR* de résistance à la kanamycine. On peut supposer qu'il y a un problème avec le plasmide, tel qu'une recombinaison avec l'ADN chromosomique, ou un problème de construction des parties permettant la réplication et la stabilité des plasmides chez *M. extorquens* DM4. Cette expérience pourrait être validée en testant directement la transformation du plasmide chez des cellules électro-compétentes de *M. extorquens* DM4.

4.3.2.C Détection de DcmA par *western blot*

La protéine DcmA a pu être détectée par *western blot* en utilisant un sérum dirigé contre cette protéine (Kohler-Staub *et al.*, 1986). En condition DCM, pour la souche sauvage et les mutants, il n'y a pas de variation significative d'abondance de DcmA (Fig. 7). Par contre, en condition méthanol et succinate, la protéine DcmA n'est pas détectée chez la souche sauvage. Elle est détectée en faible quantité chez les mutants STOP*dcmR*ΔP_{R1}, STOP*dcmR* et STOP*dcmR*Δ*dcmBC*. L'absence de la protéine DcmR peut donc être corrélée à une faible dérégulation de l'inhibition de la production de DcmA en condition de croissance sur méthanol ou succinate. Ces résultats sont moins précis que ceux obtenus en SRM. Par exemple, le *western blot* ne permet pas de détecter la protéine DcmA en présence de méthanol ou succinate chez la souche sauvage, contrairement à l'approche SRM. Néanmoins, il s'agit de réplicats biologiques indépendants qui confirment la même tendance.

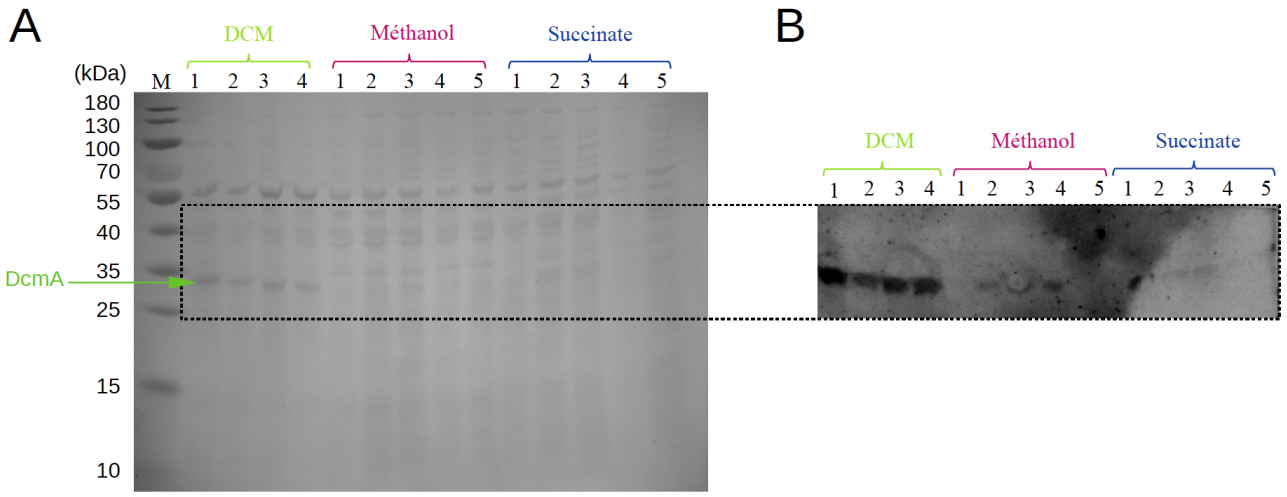


Figure 4.7: Détection de la protéine DcmA dans différents contextes génétiques.
Des extraits protéiques ont été réalisés à partir de la souche sauvage et des mutants cultivés en présence de DCM (10 mM), de méthanol (10 mM) ou de succinate (2,5 mM). Les pistes 1, 2, 3, 4 et 5 correspondent respectivement à la souche DM4 sauvage et aux mutants STOP*dcmR*ΔP$_{R1}$, STOP*dcmR*, STOP*dcmR*Δ*dcmBC* et DM4Δ*dcm*. La lettre M correspond au marqueur de taille de protéines de référence. La souche Δ*dcm* ne se développant pas en présence de DCM, il n'y a pas de colonne 5 dans cette condition. A) Protéines d'extraits cellulaires séparées par électrophorèse dans un gel de polyacrylamide 12 % en condition dénaturante, et visualisées par coloration au bleu de Coomassie. La protéine DcmA (indiquée par une flèche) d'une taille théorique de 33 kDa, migre à 35 kDa. B) *Western blot* avec un sérum anti-DcmA (KOHLER-STAUB *et al.*, 1986). Seule la portion de membrane d'hybridation correspondant à la zone de migration entre 25 à 40 kDa d'un gel d'électrophorèse ayant migré en parallèle avec les mêmes échantillons est montrée.

4.4 Perspectives de l'article *dcmR*

Les perspectives ouvertes par les nouveaux résultats présentés dans l'article et l'annexe sont abordés dans le chapitre 6.

5. Détection du gène *dcmR in situ* et diversité

Chapitre 5 : Détection du gène *dcmR in situ* et diversité

5.1 Introduction

L'étude de la diversité génétique des gènes *dcm* a été initiée et s'est poursuivie chez des souches DCM-dégradantes isolées à partir d'environnements contaminés. Ainsi, depuis la découverte de la présence de la séquence d'insertion IS*1354* flanquant une région conservée de 4,2 kb comprenant les gènes *dcm* chez plusieurs souches bactériennes dégradant le DCM (Schmid-Appert *et al.*, 1997), le séquençage du génome de trois souches dégradant le DCM a confirmé la grande conservation de la région des quatre gènes, y compris du gène *dcmR* (Vuilleumier *et al.*, 2009 ; Muller *et al.*, 2011 ; Bringel *et al.*, 2017). Dans les souches isolées au laboratoire, les gènes *dcmRABC* ont toujours été retrouvés regroupés dans le même ordre alors que les IS qui les entourent varient selon les souches (Schmid-Appert *et al.*, 1997). Chez les souches isolées dans les années 2000 et dont le génome n'a pas été séquencé, seul le gène *dcmA* a été recherché par PCR ou Southern *blot* (Nikolausz *et al.*, 2005 ; Wu *et al.*, 2007 ; Emanuelsson *et al.*, 2009). Chez ces bactéries, la présence d'autres gènes *dcm* et la nature des IS flanquantes n'a pas été déterminée. Quant aux cultures d'enrichissement et de l'ADN extrait d'échantillons de l'environnement, seule la présence du gène *dcmA* a été mise en évidence par PCR (Zuber *et al.*, 1997 ; Yu and Smith, 2000 ; Bailón *et al.*, 2009 ; Hermon *et al.*, 2018). Par contre, le gène *dcmR* n'avait à ce jour encore jamais été recherché à partir d'ADN métagénomique environnemental.

De surcroît, la diversité du contexte génétique de *dcmR* dans l'environnement n'a pas non plus été explorée alors même que de nombreuses questions restent ouvertes. Des variations de séquences sont-elles présentes chez les souches portant l'îlot *dcm* et dégradant le DCM ? Le gène *dcmR* peut-il être trouvé dans d'autres contextes génétiques ? Est-il associé systématiquement à des gènes du métabolisme méthylotrophe ou méthanotrophe comme le laisse supposer la présence de son domaine MEDS ? Pour répondre à ces questions, un séquençage d'amplicon PCR ciblant l'îlot *dcm* a été réalisé pour évaluer le degré de conservation de sa séquence et plus particulièrement celle du gène *dcmR* chez des souches issues de l'environnement. Pour cela, le génome de *Hyphomicrobium* sp. GJ21 a été séquencé (Bringel *et al.*, 2017). La publication décrivant le séquençage de ce génome et son annotation se trouve en annexe. Les souches *Hyphomicrobium* sp. T28.1 et T28.3, dégradant

le DCM et pour lesquelles le gène *dcmA* a été détecté (Hermon *et al.*, 2018), ont aussi été utilisées dans cette analyse. D'autre part, cette étude de la diversité du gène *dcmR* a été complétée par la recherche et l'analyse bio-informatique de séquences disponibles dans les banques de données telles que GenBank ou Uniprot. Enfin, la présence systématique des gènes *dcmR* et *dcmA* dans des environnements pollués a été testée. Pour cela, les gènes *dcmR* et *dcmA* ont été quantifiés conjointement à partir de deux échantillons d'environnements industriels pollués par le DCM à savoir un sol et une eau d'aquifère (sol saturé en eau). Rechercher et quantifier la présence de *dcmR* et de *dcmA* permettrait d'évaluer le ratio entre *dcmA* et *dcmR in situ*. La connaissance de ce ratio fournira des indices sur la co-occurrence dans l'environnement des gènes *dcmR* et *dcmA,* jusqu'ici toujours retrouvés côté à côte dans les souches DCM-dégradantes qui ont été caractérisées.

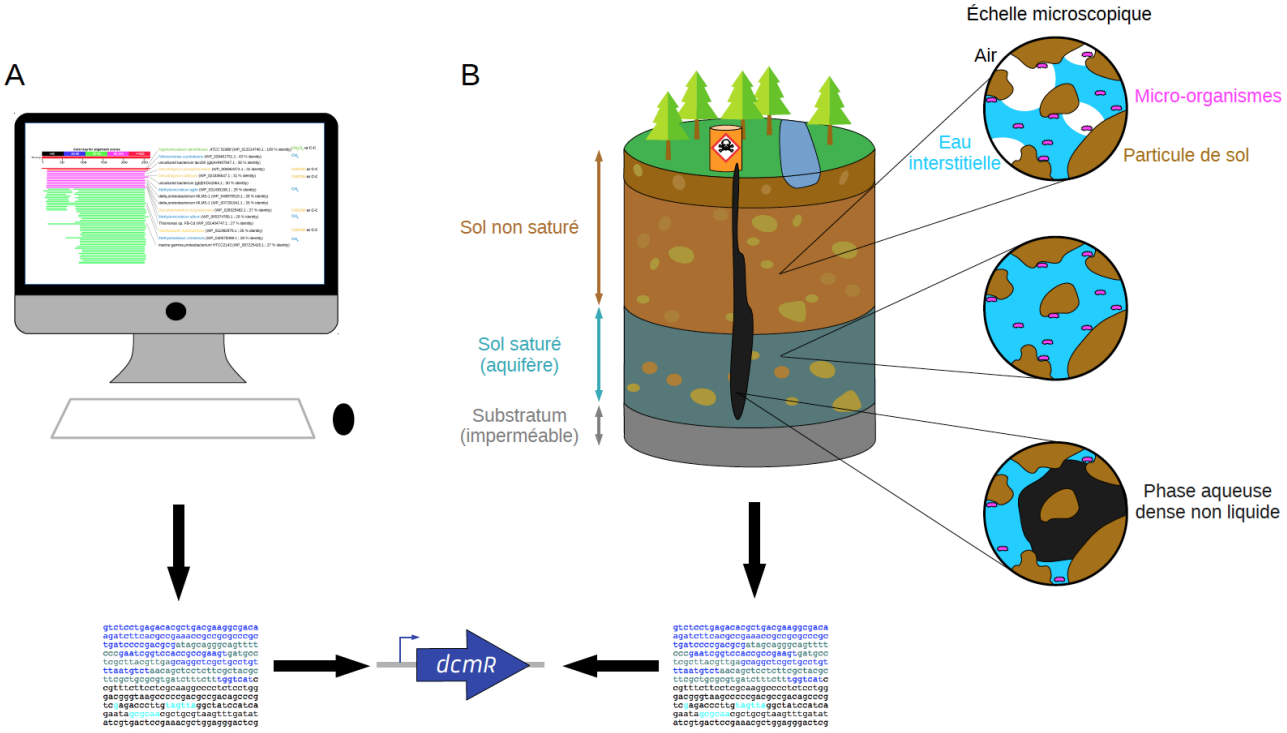


Figure 5.1 :Recherche de *dcmR* par bio-informatique et dans l'environnement.
A Des gènes et protéines similaires à *dcmR* sont recherchés dans les bases de données. B Le sol est un milieu complexe composé de particules minérales, d'air et d'eau ainsi que de micro- et macro-organismes. Les interstices entre les particules minérales composant le sol peuvent être remplies d'eau (zone saturée) ou alors d'eau et d'air (zone non saturée). Dans les deux cas, des micro-organismes peuvent se développer à l'intérieur des pores et se fixer sur les particules minérales (Beyer *et al.*, 2014). Le DCM peut être trouvé sous forme de gaz, soluble dans l'eau ou insoluble dans une phase aqueuse dense non liquide (DNAPL en anglais) qui va sédimenter au fond de la nappe phréatique (Seyedabbasi *et al.*, 2012). La couche de roche mère appelée substratum est imperméable.

5.2 Matériels et méthodes

5.2.1 Comparaison des séquences protéiques

La recherche de séquence homologues à DcmR a été faite sur la base de similarités en utilisant le logiciel BLASTp (Altschul *et al.*, 1990) dans plusieurs banques de données (GenBank ; Protein Data Bank ; SwissProt ; Protein research foundation). Seules les protéines possédant un domaine HTH et MEDS ont été retenues. Le degré de similarité des protéines DcmR-*like* par rapport à celles déjà connues, est visualisé par la construction d'arbres phylogénétiques. On peut également parler de phénogramme pour ces graphiques car ils sont construits uniquement sur la comparaison de la distance entre des séquences et non pas sur des théories évolutives comme un arbre phylogénétique (Sneath, 1989). Le logiciel MUSCLE (Edgar, 2004) a été utilisé pour réaliser des alignements multiples qui ont ensuite été affinés à l'aide du logiciel Gblock (Castresana, 2000). PhyML a été utilisé pour construire l'arbre par maximum de vraisemblance (Guindon *et al.*, 2010).

5.2.2 Description de l'origine des échantillons

5.2.2.A Échantillons de l'aquifère du site industriel de Themeroil et extraction d'ADN

L'eau d'une aquifère multi-contaminée par des polluants chlorés a été prélevée à partir de piézomètres sur le site industriel de Themeroil (France). Il s'agit d'une ancienne usine, spécialisée dans le traitement de solvants chlorés, ayant pollué l'environnement par un mauvais stockage des déchets. Le site est contaminé par des éthènes et des alcanes chlorés dont le DCM. Plusieurs piézomètres ont été installés pour prélever des échantillons sur le site et ses alentours (Fig. 2). La dégradation biologique du DCM a été mise en évidence *in situ* par des approches de chimie-isotopique (Hermon *et al.*, 2018). Deux nouvelles souches dégradant le DCM et possédant le gène *dcmA* ont été isolées à partir de ce site : *Hyphomicrobium* sp. T28.1 et T28.3.

A B

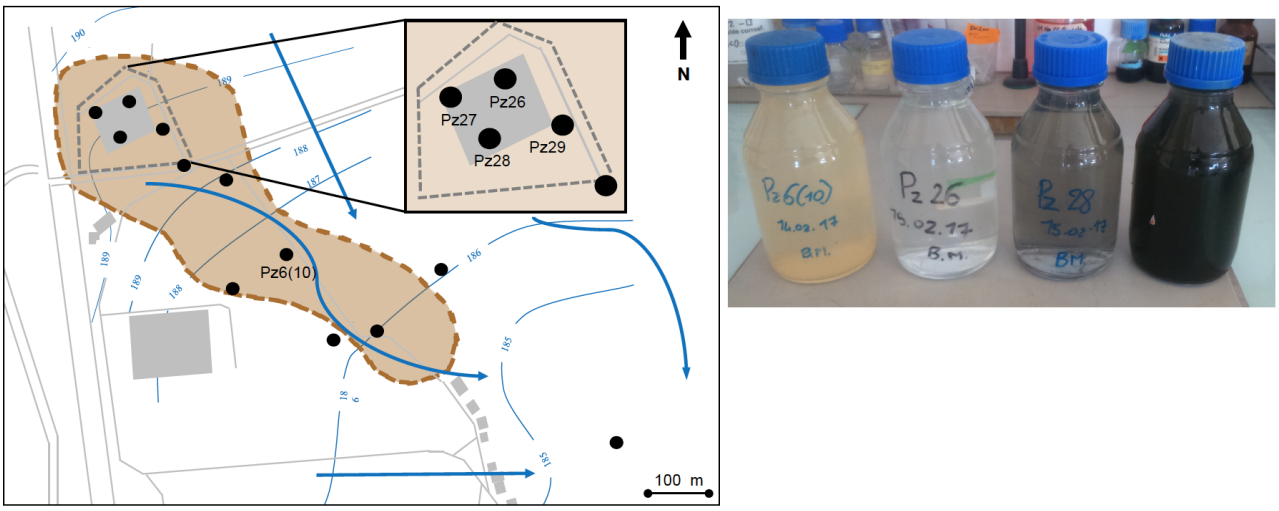


Figure 5.2: Description du site de Themeroil pollué par des solvants chlorés dont le dichlorométhane (Figure adaptée de Louis Hermon 2017 avec sa permission).
A Carte du site de Themeroil (en gris) avec le panache de pollution (en marron) et l'emplacement des différents piézomètres (Pz) construits pour le prélèvement d'échantillons d'aquifère. Le niveau piézométrique est indiqué en bleu sur la figure (en mètres). Les flèches bleues indiquent le sens d'écoulement de la nappe phréatique. Les souches *Hyphomicrobium* sp. T28.1 et T28.3 dégradant le DCM et possédant *dcmA* ont été isolées du piézomètre Pz28. B Photographie des échantillons d'eaux de la nappe phréatique échantillonnée de différents piézomètres utilisés pour la recherche du gène *dcmR*.

L'eau de l'aquifère de Themeroil a été prélevée le 14 février 2017, et a été stockée à 4 °C. L'extraction de l'ADN a été réalisée à partir d'un volume de 250 mL filtré le 16 février 2017, de façon stérile, sur des membranes de nylon GNWP de 0,2 µm (Millipore). L'échantillon Pz27 a nécessité un passage au préalable au travers d'une feuille de Wattman pour enlever les particules en suspension qui saturaient le filtre. Les micro-organismes récoltés sur les membranes de nylon ont ensuite été conservés à -25 °C. Un témoin d'extraction a été réalisé en filtrant une suspension bactérienne de *M. extorquens* DM4 (250 mL) à une densité de 10^7 cellules par mL. L'ADN a été extrait directement à partir des membranes de filtration avec le kit PowerWater (MOBio) selon les indications du fabricant.

5.2.2.B Echantillons prélevés de microcosmes exposés ou non au DCM

Des microcosmes ont été réalisés dans le cadre du projet TOP-DCM, financé par l'INSU-CNRS. Le projet a pour objectif l'étude de l'adaptation du compartiment microbien d'un microcosme de sol surfacique multi-contaminé exposé à des polluants industriels. Dans ce projet, le sol contaminé par

le lindane provient d'un ancien site de production industriel disponible suite à un accord de confidentialité entre les gérants du site de la friche industrielle et le LHyGeS (laboratoire d'hydrologie et de géochimie de Strasbourg, Dr. Imfeld). Ce sol a été exposé au DCM qui a servi de solvant pour le lindane (gamma-hexachlorocyclohexène) dans sa formulation commerciale. L'analyse physico-chimique de ce sol a montré une contamination multiple, à des concentrations supérieures aux normes, de plusieurs polluants chlorés (isomères de l'hexacyclohexène, 3,4,5,6-tétrachlorocyclohexène et 1,3,4,5,6-pentachlorocyclohexène) ou non (naphtalène et anthracène). À partir de ce sol multi-contaminé, trois microcosmes ont été réalisés et étudiés en parallèle selon un dispositif détaillé dans la Figure 3. Le dispositif expérimental incluait une approche de bio-augmentation par ensemencement d'une souche bactérienne *Hyphomicrobium* sp. GJ21. Cette souche est connue pour sa capacité à se développer dans des microcosmes ou des biofiltres (Ottengraf *et al.*, 1986 ; Hermon *et al.*, 2018), en aérobie ou en condition anaérobie en utilisant l'oxygène ou le nitrate comme accepteur d'électrons. Des détails supplémentaires sur la souche GJ21 sont fournis dans la publication du séquençage de son génome (annexe chapitre 5, (Bringel *et al.*, 2017)). La bio-augmentation avait pour but de servir de témoin positif d'une part, et d'avoir un comparatif de dégradation pour les souches endogènes potentiellement DCM-dégradantes dans le microcosme non bio-augmenté et exposé au DCM d'autre part.

Du milieu M3 (Muller *et al.*, 2011) dilué à 1 % avec ou sans DCM est apporté via une pompe péristaltique à la base de la colonne. L'apport du liquide par le bas de la colonne permet d'obtenir deux zones, l'une saturée en eau interstitielle et une autre non-saturée plus haut dans la colonne. Plusieurs valves permettent de récupérer le liquide interstitiel (l pour *low* et m pour *middle*) ou la phase gazeuse (h pour *high* et t pour *top*) à différents niveaux de la colonne (Fig. 3). La dégradation du DCM a été suivie par chromatographie en phase gazeuse (Varian CP-3800). Des dénombrements bactériens ont été réalisés après croissance à 30 °C sur différents milieux de culture gélosé : Nutrient agar ; M3 + DCM 8 mM ± acétate 5 mM en présence d'un indicateur de pH (Bleu de bromothymol à 12,8 mM). Les microcosmes exposés au DCM sont les microcosmes D et G. La souche déchlorante *Hyphomicrobium* sp. GJ21 a été ensemencée dans les microcosmes C et G à raison de 80 mL de suspension à 5.105 CFU par mL. Du sol a été prélevé au lancement de la culture (-31 jours) puis après 61 jours lorsque les colonnes ont été démantelées. Ces expériences de chromatographies en phase gazeuse et de cultures ont été réalisées par une étudiante, Farah Hajjar, en 2015. L'extraction d'ADN a été réalisée à partir d'échantillons congelés à -25 °C. Une masse de

0,25 g de sol a été prélevée pour une extraction avec le kit DNeasy PowerSoil (Quiagen). Contrairement aux échantillons de sol, il n'a pas été possible d'extraire suffisamment d'ADN à partir des prélèvements du liquide interstitiel des microcosmes (filtres réalisés à partir de 50 mL) en utilisant le kit PowerWater (MOBio).

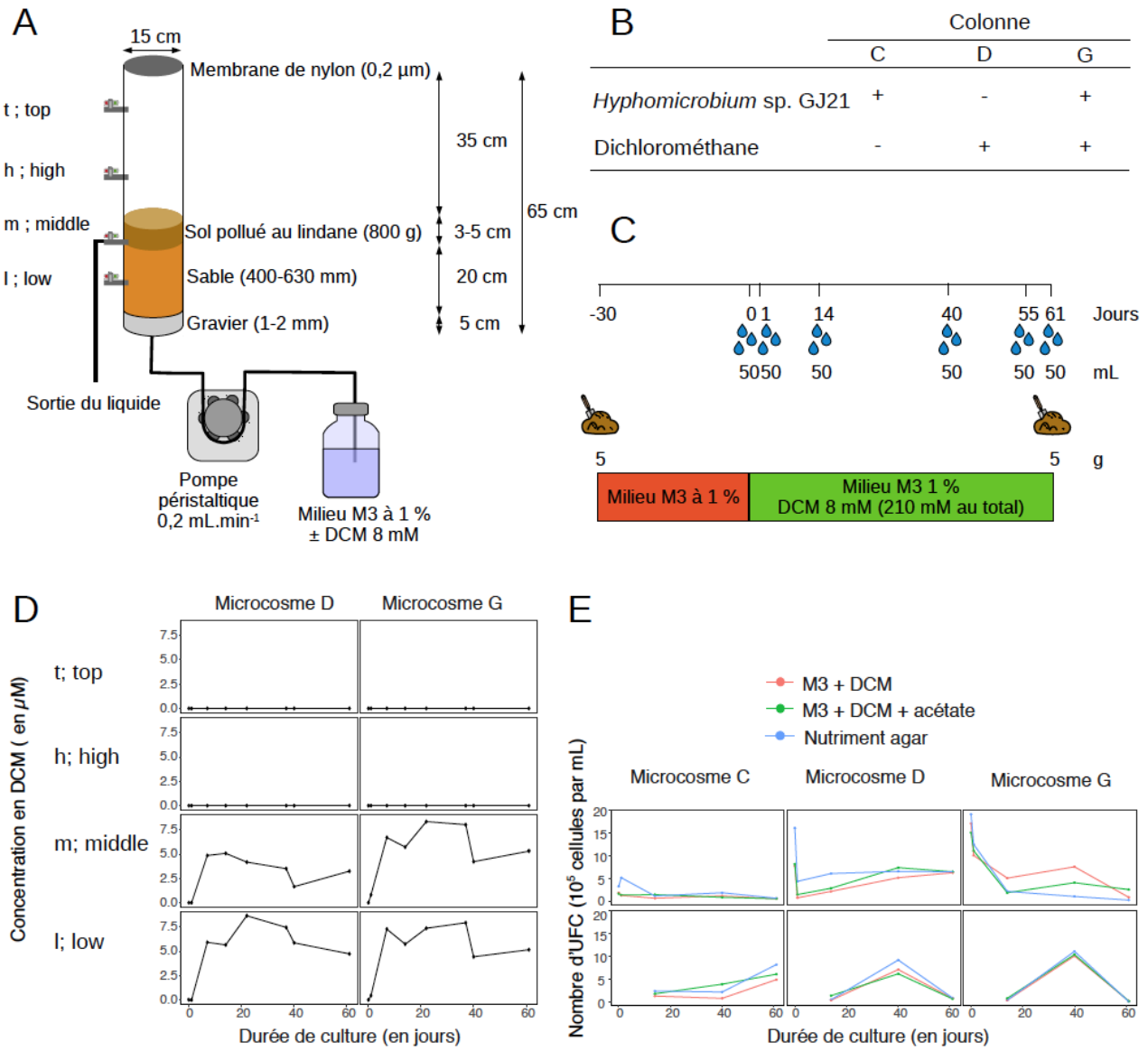


Figure 5.3: Description des microcosmes du projet TOP-DCM.
A Un sol pollué au lindane et au dichlorométhane est placé dans des colonnes en verre sur un lit de sable et de gravier stérilisés préalablement par autoclavage. Du milieu nutritif auquel du DCM a été ajouté ou non alimente chaque colonne. B Plusieurs conditions de culture ont été réalisées : avec ou sans ajout de DCM ou de la souche souche DCM-dégradante *Hyphomicrobium* sp. GJ21. C Chronologie de l'expérience. D Quantification du DCM par chromatographie en phase gazeuse à partir d'échantillons sous forme gazeuse ou liquide. E Dénombrement bactérien, sur différents milieux de culture, à partir du liquide interstitielle prélevé.

5.2.3 Quantification des gènes *dcmA* et *dcmR* par qPCR dans l'environnement

Les gènes *dcmA*, *dcmR* et celui de l'ARNr 16S ont été quantifiés par PCR quantitative (qPCR) par référence à une droite d'étalonnage (méthode absolue). L'amplification des gènes a été suivie en temps réel par la fixation de SYBR® Green, un agent intercalant dont la fluorescence change selon s'il est fixé ou non à l'ADN. Les résultats sont normalisés par rapport à un second fluorochrome, le 6-carboxy-X-rhodamine (ROX), qui ne se lie pas à l'ADN et permet ainsi de comparer la fluorescence entre les puits (témoin du volume réactionnel). Le suivi en temps réel permet de déterminer le cycle seuil (Ct) par la méthode d'ajustement de courbes sigmoïdes (Spiess *et al.*, 2008) avec la librairie R qpcR (Ritz & Spiess, 2008). Des inhibiteurs tels que les acides humiques peuvent être extraits en même temps que l'ADN à partir d'échantillons environnementaux (Schrader *et al.*, 2012). Leur présence a été évaluée par deux approches : i) Les échantillons sont dilués en cascades dans de l'eau ultra-pure à plusieurs concentrations pour voir s'il y a une diminution du nombre de Ct à fortes concentrations ; ii) Un ADN témoin de concentration connue est ajouté à certains échantillons.

Les réactions qPCR se font directement dans des plaques 96 puits translucides pour qPCR (Agilent) avec un volume réactionnel de 15 µL comprenant 3,75 µL d'ADN matrice, 3,0 µL d'eau ultra-pure, 0,225 µL de chaque amorce (à 20 µM ; Eurofins Genomics ; Tab 2), 0,3 µL de l'albumine de sérum bovin à 50 $mg.L^{-1}$ (Ambion) ainsi que 10 µL de Brilliant III SYBR® Green *low* ROX qPCR *master mix* (Agilent). Cette dernière solution contient les dNTP, le tampon et l'ADN polymérase ainsi que du SYBR® Green pour détecter l'ADN et du ROX. Pour limiter l'évaporation des mélanges réactionnels et la contamination entre les puits, la plaque 96 puits est scellée avec un film plastique transparent (Microseal® 'B' *adhesive seals* ; Bio-Rad) puis centrifugée pendant une minute à 130 g avant l'étape d'amplification PCR. Le programme de qPCR est similaire à celui décrit précédemment (Chaignaud *et al.*, 2017). Brièvement, le programme du thermocycleur (Stratagene Mx3005P Agilent Technologies) commence par trois minutes à 95 °C, se poursuit par quarante cycles alternant 20 secondes à 95 °C pour la dénaturation et 20 secondes à 60 °C pour l'hybridation et l'élongation. Après l'amplification, une courbe de dissociation est réalisée avec une minute à 95 °C puis 30 secondes à 60 °C et 30 secondes à 95 °C. Cette dernière étape permet de vérifier la spécificité des produits PCR en mesurant la dissociation de l'ADN. La courbe de dissociation

obtenue permet d'observer s'il y a un ou plusieurs produits de PCR ainsi que la présence de dimères d'amorces. Un témoin négatif sans ADN est également réalisé. Les produits de qPCR sont déposés dans un gel d'agarose 1 % puis séparés par électrophorèse pour valider leur pureté et leur taille.

Le plasmide pME8309 a été construit par Ludovic Besaury par clonage dans un plasmide pGEM®-T Easy Vector (Promega) d'un fragment du gène codant l'ARNr16S d'une souche de la phyllosphere. Cette séquence a été amplifiée avec les amorces 799f/1193 et possède 98 % d'identité avec *Chryseobacterium* sp. Y85J12 accession n° KT201627.

Le ratio entre le nombre de copies de *dcmA* et *dcmR* est adapté d'une méthode décrite (Škulj *et al.*, 2008). Pour un même échantillon d'ADN, le nombre de Ct est calculé à partir de droite d'étalonnage pour le gène *dcmA* et *dcmR* et un ratio normalisé est calculé en tenant compte du facteur d'efficacité spécifique des réactions PCR de chaque gène.

$$Ratio\,normalisé_{\,dcmA/dcmR} \frac{Efficacité\ pour\,dcmA \times Ct_{\,dcmA}}{Efficacité\ pour\,dcmR \times Ct_{\,dcmR}}$$

Tableau 1 : Liste des amorces et plasmides utilisées pour la qPCR

Gène	Amorce	Séquence 5'-> 3'	Tm (en °C) [a]	Taille des produits (bp)	Référence des amorces	Plasmide pour l'étalonnage	Référence du plasmide
dcmR (domaine MEDS)	tnR1r	gcccactccatatcaccaac	60,4	151	(Chaignaud *et al*., 2017)	pME8308	Non publié
	tnR1f	acgtgcacaaagccattaaa	61,9				
dcmA	tnA1f	ttcggctgtttgatctatgg	59,5	159	(Chaignaud *et al*., 2017)	pME1944	(Vuilleumier *et al*., 2001)
	tnA1r	atccgcatagctgacctcat	62,6				
ADNr 16S (région V5-V7)	799f	aac**m**ggattagataccc**k**g[b]	58,1	394	(Bodenhausen *et al.*, 2013)	pME8309	Non publié
	1193r	acgtcatccccaccttcc	63				

[a] La température de fusion des amorces a été calculée avec le logiciel OligoAnalyzer (version 3.1 ; IDT®) en utilisant des concentrations de 0,2 µM d'amorces, 50 µM de sel monovalent, 2,5 mM de sel divalent et 0,8 mM de dNTPs.

[b] Les lettres en gras correspondent à une ambiguïté de nucléotide selon la nomenclature IUPAC. La lettre K correspond à G ou T tandis que M correspond à A ou C.

5.3 Résultats et discussions

Dans un premier temps, le gène *dcmR* est recherché dans les différents génomes disponibles dans les banques de données d'acides nucléiques. Cette recherche a permis de déterminer si *dcmR* est toujours trouvé dans le contexte génétique de l'îlot *dcm*. Étudier la diversité de séquence de *dcmR* et de la protéine qu'il code permettra également de déterminer des régions conservées ou non et d'orienter de futures études.

5.3.1 Séquence conservée du gène *dcmR* au sein du transposon *dcm*

L'étude du *locus dcmR* chez les nouvelles souches dégradant le dichlorométhane confirme la très grande conservation des gènes *dcm*. En effet, aucune différence de séquence pour le gène *dcmR* et la région intergénique entre les gènes *dcmR*-*dcmA* n'a été observée pour les trois souches d'*Hyphomicrobium* sp. analysées (Fig. 5). En plus des deux souches d'*Hyphomicrobium* sp. T28.1 et T28.3 du site de Themeroil, le séquençage du génome de *Hyphomicrobium* sp. GJ21 a fait l'objet d'une publication dans le journal *Genome Announcement* (Cf. annexe ; (Bringel *et al.*, 2017)). Cette souche possède un îlot *dcm* très conservé par rapport à celui de référence de *M. extorquens* DM4. Sur les 4,2 kb de l'îlot, on dénombre 11 variations non synonymes dans le gène *dcmA* et 21 mutations dont une délétion de 6 nucléotides pour le gène *dcmC* qui ne modifie pas le cadre de lecture (Fig. 5). Les données de séquences obtenues chez les trois souches *d'Hyphomicrobium* ont été comparées à celles disponibles pour les autres souches DCM-dégradantes isolées et possédant l'îlot *dcm* (Tab. 2 ; Fig. 5 et 6).

Tableau 2 : Souches isolées dont l'îlot *dcm* a été séquencé

Souche	Origine géographique	Site d'isolement	Numéro d'accession des séquences	Référence
Methylorubrum extorquens DM4	ND [a]	Sol pollué d'une usine chimique	NC_012988.1	(Gälli & Leisinger, 1985; Vuilleumier *et al.*, 2009)
Gottschalkia methylica DM15	ND	Sol de jardin pollué au DCM	FJ235914.1	(Firsova *et al.*, 2010)
Ancylobacter dichloromethanicus DM16	Russie ; Volgograd	Sol pollué d'une usine chimique	EU596479.1	(Firsova *et al.*, 2009, 2010)
Methylorubrum extorquens DM17	ND	Sol pollué d'une usine chimique	ND	(Firsova *et al.*, 2010)
Hyphomicrobium denitrificans ATCC51888	ND	ND	NC_014313	(Brown *et al.*, 2011)
Hyphomicrobium sp. GJ21	ND	Biofiltre	CDHO00000000	(Ottengraf *et al.*, 1986 ; Bringel *et al.*, 2017)
Hyphomicrobium sp. T28.1	France ; Varennes-le-Grand	Usine Themeroil ; aquifère polluée	ND	(Hermon *et al.*, 2018)
Hyphomicrobium sp. T28.3	France ; Varennes-le-Grand	Usine Themeroil ; aquifère polluée	ND	(Hermon *et al.*, 2018)

[a] Non disponible

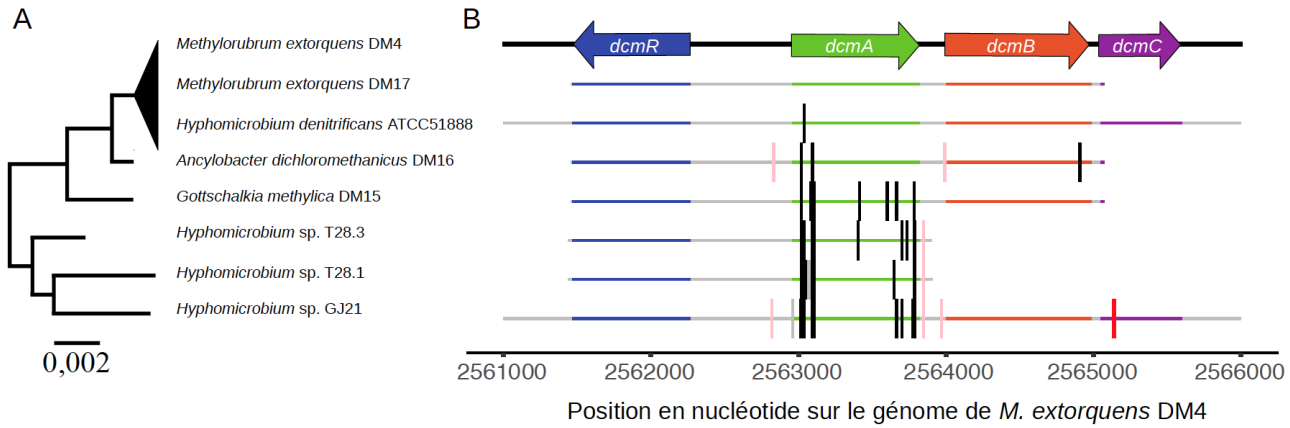


Figure 5.4: Conservation de la séquence de l'îlot *dcm* chez différentes souches bactériennes dégradant le DCM.
A Phénogramme construit à partir de la séquence nucléique de l'îlot *dcm*. B Positionnement des mutations par rapport à la référence *Methylorubrum extorquens* DM4 qui est considérée comme référence. Les lignes horizontales représentent la partie séquencée. Les traits verticaux gris, noirs, rouges et roses correspondent respectivement aux mutations synonymes, non-synonymes, aux délétions et aux variations dans les régions non codantes, variations synonymes, non-synonymes, aux délétions et aux mutations dans les régions non-codantes.

L'étude de l'îlot *dcm* confirme qu'il est très conservé en termes d'organisation et de séquences (Fig. 4). Le gène *dcmR* est le seul à ne pas présenter de variation parmi les gènes *dcm* de séquences connues. Deux mutations sont aussi trouvées dans la région intergénique *dcmR-dcmA* mais ne se situent pas dans les séquences putatives de régulation (La Roche & Leisinger, 1991). Les variations détectées se trouvent principalement dans le gène *dcmA*. Une hypothèse serait que ces mutations permettraient de modifier l'activité enzymatique de DcmA. Néanmoins, des expériences de biochimie sur plusieurs variants de DcmA ont montré que ces mutations ont peu d'effets en terme d'activité enzymatique ou de propriétés physico-chimiques (Vuilleumier *et al.*, 2001). La souche *Methylophilus leisingerii* DM11 fait cependant exception car elle présente 332 variations dans son gène *dcmA* comparé à la séquence de *M. extorquens* DM4 (Bader & Leisinger, 1994). La protéine DcmA codée par ce gène présente 56 % d'identité avec l'enzyme de DM4 et possède des propriétés enzymatiques différentes. Il a donc été proposé de classer les enzymes DcmA en deux groupes : « A » qui comprend *M. extorquens* DM4 et la majorité des enzymes DcmA ; et « B » qui ne comprend à ce jour que la séquence DcmA de *Methylophilus leisingerii* DM11. Chez DM11, le promoteur de *dcmA* présente également des mutations par rapport à *M. extorquens* DM4. La séquence conservée chez *M. extorquens* DM4 n'est pas retrouvée entre les boîtes -10 et -35. Les gènes *dcmR*, *dcmB* et *dcmC* n'ont pas encore été déterminés pour la souche DM11. Le séquençage

du génome de la souche DM11 permettrait de savoir si les mutations se limitent à *dcmA* ou sont aussi présentes dans les autres gènes.

Si l'on considère l'ensemble des souches possédant *dcmA*, connues à ce jour, une particularité est à noter pour le gène *dcmA* de *Bacillus circulans* WZ-12 (Wu *et al.*, 2007, 2009). Son gène *dcmA* possède 209 variations par rapport au gène *dcmA* chez *M. extorquens* DM4 qui n'apportent cependant aucune modification de la séquence protéique (variations synonymes). Aucune variation non synonyme n'est observée. De plus et alors que les protéines DcmA chez *B. circulans* WZ-12 et *M. extorquens* DM4 sont prédites comme identiques sur la base de leur séquence, les protéines purifiées ont été décrites comme possédant des propriétés enzymatiques et une taille différente (20 kDa chez WZ-12 au lieu de 33 kDa chez DM4). Une régulation post-traductionnelle telle qu'un clivage par une protéase pourrait expliquer ce résultat. Ces données restent pour le moins mystérieuses et nécessiteraient une étude approfondie pour voir comment classer ce gène parmi les autres variants *dcmA*. Les auteurs de cet article n'ont jamais répondu à une demande répétée d'envoi de la souche.

5.3.2 Diversité du contexte génétique des gènes des protéines homologues à DcmR

Le gène *dcmR* a été recherché par le logiciel BLASTn à partir de la séquence nucléotidique du gène de *M. extorquens* DM4. Outre les protéines DcmR de séquence identique et associées au transposon *dcm* chez différentes souches dégradant le DCM, 10 protéines avec une architecture similaire à celle de DcmR, à savoir la présence des domaines HTH et MEDS, ont été détectées. Le pourcentage d'identité protéique entre chacune de ces protéines s'échelonne de 36 à 82 % mais ne dépasse pas 42 % avec DcmR de *M. extorquens* DM4 (Figure 5B).

L'alignement de séquences permet d'observer deux zones conservées qui correspondent respectivement aux domaines HTH et MEDS de DcmR (Fig. 6). Entre les deux se trouve une région très variable. Le domaine MEDS présentes de nombreuses régions conservées comparées en plus de celles observées précédemment (Anantharaman & Aravind, 2005). Certains acides aminés qui étaient prédits comme conservés ne le sont pas dans cette étude. Cela peut signifier qu'il y aurait plusieurs groupes de domaines MEDS avec potentiellement des fonctions différentes.

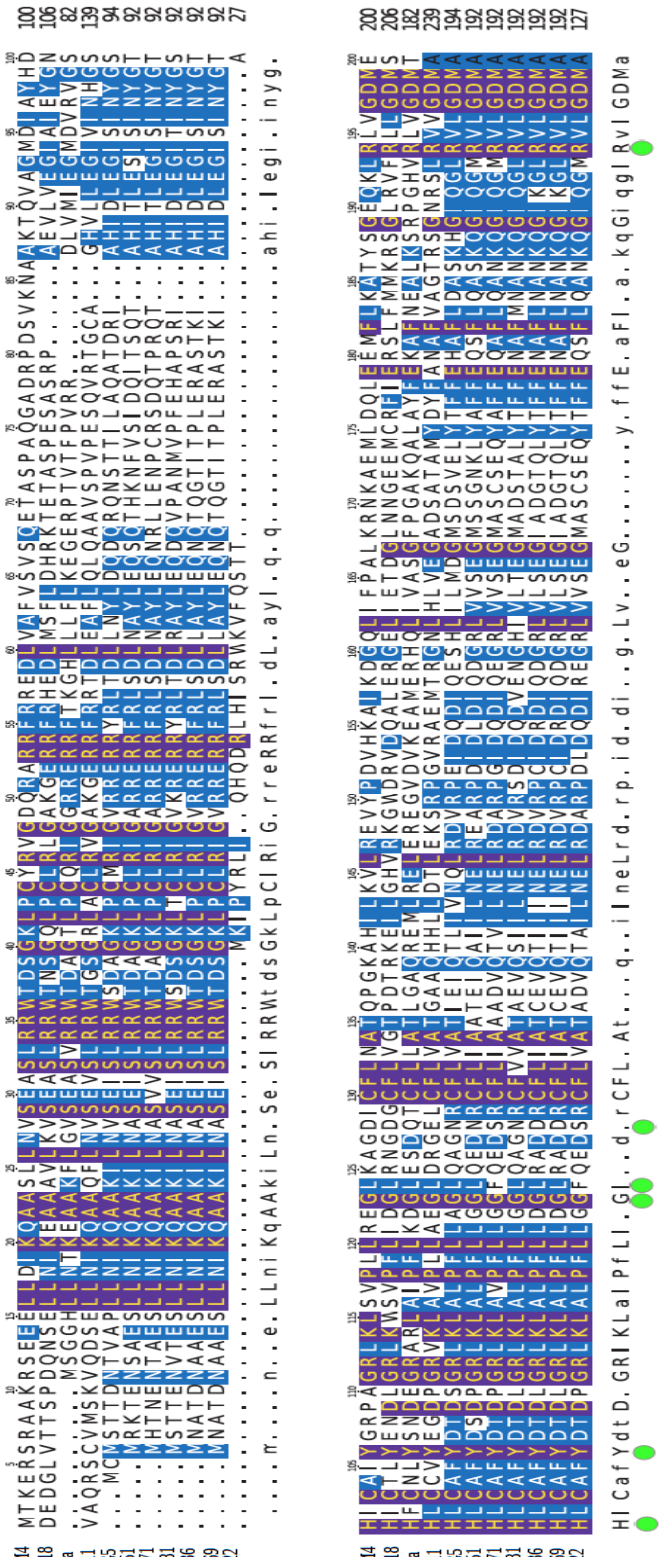

Mextorquens extorquens DM4
Rhodospirillaceae HHTR118
Candidatus Methylomirabilis limnetica
Azoarcus sp. DN11
Nitrosomonas sp. APG5
Nitrosomonas sp. NM51
Nitrosomonas marina NM71
Nitrosomonas cryotolerans ATCC49181
Nitrosomonas aestuarii NM36
Nitrosomonas aestuarii NM69
Nitrosomonas marina NM22

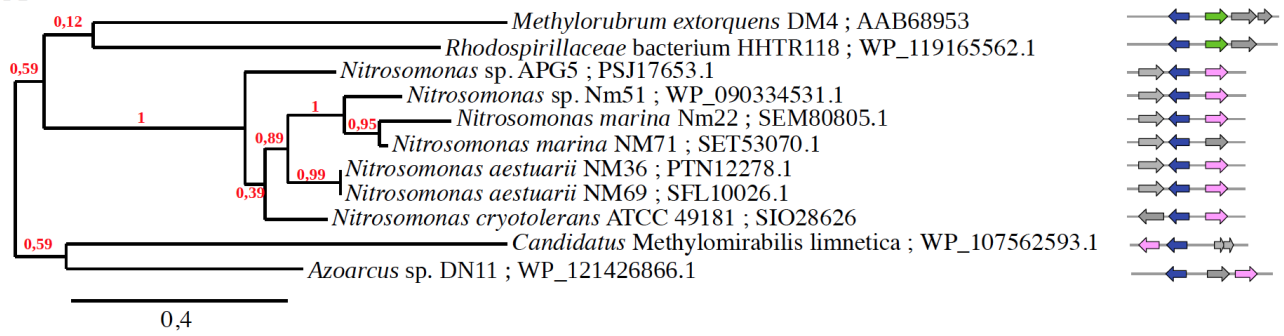
Figure ci-contre

Figure 5.5: Alignement des séquences protéiques de *DcmR-like*.
Les points verts correspondent aux acides aminés qui avaient été prédits comme conservés dans le domaine MEDS (Anantharaman & Aravind, 2005).

En outre, plusieurs génomes portent un gène *dcmR-like* localisé à proximité d'un gène potentiellement impliqué dans la synthèse de la biotine, à l'exception de *Nitrosomonas marina* NM71 et *Rhodospirillaceae bacterium* HHTR118 (Fig. 6A). Ce gène coderait une enzyme avec un domaine pimeloyl-ACP methyl ester carboxyl esterase (EC 3.1.1.85) qui produit du méthanol au cours de sa réaction. Cela suggère un lien potentiel avec les protéines similaires à DcmR et le domaine MEDS qui est supposé reconnaître des petits composés hydrocarbonés (Anantharaman & Aravind, 2005). Curieusement, ce gène est trouvé uniquement dans les génomes de souches possédant un gène *dcmR-like*. En plus de ce gène potentiel de synthèse de biotine, *dcmR* est rencontré dans des contextes génétiques variés associés à des gènes du système toxine/antitoxine *hciBA*, une pyruvate kinase ou une endonucléase (Fig. 6A).

Le contexte génétique du gène *dcmR-like* chez *Rhodospirillaceae bacterium* HHTR118 est le plus proche de celui rencontré dans l'îlot *dcm* (Fig. 6A). En effet, le gène *dcmR-like* se trouve aussi en orientation divergente avec un gène possédant une forte identité de séquence au niveau protéique avec DcmA de *M. extorquens* DM4 (Fig. 7). Il s'agirait ici d'un premier exemple d'une souche dans laquelle les gènes *dcmR* et *dcmA* sont détectés en absence de *dcmB* et *dcmC*. De plus et alors que ces deux gènes possèdent une très faible conservation au niveau nucléotidique avec *M. extorquens* DM4, une forte conservation au niveau protéique est détectée : 40 % pour DcmR et 60 % pour DcmA. Une comparaison des séquences protéiques de DcmA de *M. extorquens* DM4 et *Rhodospirillaceae bacterium* HHTR118 a été réalisée (Fig. 8). Pour cela d'autres protéines avec des séquences proches ont été incluses à l'analyse. La protéine DcmA prédite chez la souche *Rhodospirillaceae bacterium* HHTR118 est assez différente de celle de *M. extorquens* DM4 du groupe A, mais aussi de celle de *Methylophilus leisingerii* DM11 du groupe B (Fig. 8). Cette enzyme pourrait ainsi définir un nouveau troisième groupe de protéines DcmA, ici nommé « C ». Cela rend d'autant plus intéressante l'étude de cette souche afin de savoir si elle est bien capable de dégrader le DCM ou d'autres méthanes halogénés.

A



B

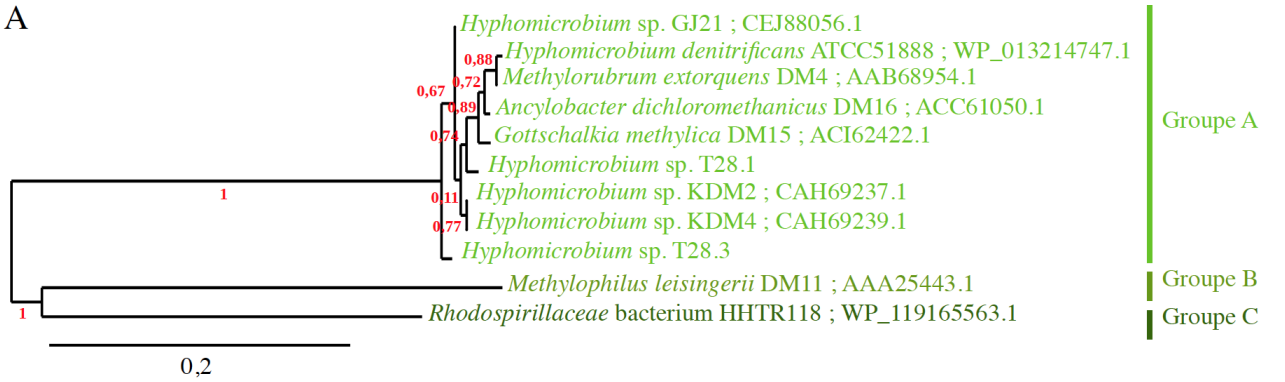
	1	2	3	4	5	6	7	8	9	10	11
1: *Candidatus* Methylomirabilis limnetica	100,00										
2: *Methylorubrum extorquens* DM4	36,18	100,00									
3: *Rhodospirillaceae* bacterium HHTR118	40,65	40,54	100,00								
4: *Azoarcus* sp. DN11	42,28	41,22	45,19	100,00							
5: *Nitrosomonas* sp. Nm51	41,46	43,53	41,57	46,69	100,00						
6: *Nitrosomonas marina* Nm22	35,83	37,50	35,79	41,67	72,02	100,00					
7: *Nitrosomonas marina* NM71	40,65	43,14	41,96	47,86	81,01	82,90	100,00				
8: *Nitrosomonas* sp. APG5	42,28	42,35	42,80	46,72	68,60	64,25	67,83	100,00			
9: *Nitrosomonas cryotolerans* ATCC 49181	40,65	41,57	43,14	49,81	70,54	65,28	70,54	73,26	100,00		
10: *Nitrosomonas aestuarii* NM36	42,28	44,71	42,75	48,25	74,81	67,36	75,19	72,87	77,91	100,00	
11: *Nitrosomonas aestuarii* NM69	42,28	44,71	42,75	48,25	74,81	67,36	75,19	72,87	77,91	100,00	100,00

Figure 5.6: Diversité de séquences et de contexte génétique de protéines homologues au régulateur DcmR

A Phénogramme basé sur la séquence protéique de DcmR. Les valeurs de *bootstrap* sont indiquées en rouge. Le contexte génétique du gène *dcmR* (en bleu) est indiquée à droite du phénogramme. Un gène de synthèse de la biotine (rose) coderait une enzyme pimeloyl-ACP methyl ester carboxyl esterase tandis que les gènes en gris ont des fonctions diverses. B Pourcentage d'identité de séquence entre les protéines DcmR lors d'un alignement multiple.

En aval de du gène *dcmA* chez le génome *Rhodospirillaceae bacterium* HHTR118 (Fig. 7A) se trouve un gène annoté comme codant une enzyme impliquée dans l'oxydation du formaldéhyde en formiate : S-(hydroxymethyl)glutathion déshydrogénase (Zhang *et al.*, 2015). Cette enzyme pourrait métaboliser le formaldéhyde produit par DcmA (Fig. 7B). Une expression simultanée de ces deux enzymes semble donc logique. Un troisième gène est prédit comme codant une formyl-CoA transférase. Il y a potentiellement un lien entre ce gène et la troisième enzyme de cette voie qui est la S-formylglutathionhydrolase. L'annotation de ce gène ne permet pas d'avoir plus d'informations.

A



B

	1	2	3	4	5	6	7	8	9	10	11
1: *Gottschalkia methylica* DM15	100,00										
2: *Ancylobacter dichloromethanicus* DM16	98,26	100,00									
3: *Methylorubrum extorquens* DM4	97,92	98,96	100,00								
4: *Hyphomicrobium denitrificans* ATCC51888	97,57	98,61	99,65	100,00							
5: *Hyphomicrobium* sp. T28.3	97,22	96,88	96,88	96,88	100,00						
6: *Hyphomicrobium* sp. GJ21	97,22	97,57	96,88	96,88	98,26	100,00					
7: *Hyphomicrobium* sp. KDM4	97,22	97,57	96,88	96,88	97,57	99,31	100,00				
8: *Hyphomicrobium* sp. KDM2	97,22	97,57	96,88	96,88	97,57	99,31	100,00	100,00			
9: *Hyphomicrobium* sp. T28.1	97,57	97,22	97,22	97,22	97,92	98,26	98,26	98,26	100,00		
10: *Rhodospirillaceae bacterium* HHTR118	60,74	60,37	60,37	60,00	61,11	61,11	60,37	60,37	60,74	100,00	
11: *Methylophilus leisingerii* DM11	56,18	55,81	55,81	55,81	55,43	56,18	56,18	56,18	56,55	58,80	100,00

Figure 5.7: Diversité de la protéine DcmA.
A Arbre phylogénétique de la protéine DcmA. Deux groupes « A » et « B » ont pu être définis sur la base ds propriétés enzymatiques de DcmA et la comparaison de séquences. La souche *Rhodospirillaceae bacterium* HHTR118 pourrait correspondre à un troisième groupe « C ». Les valeurs de *bootstrap* sont indiquées en rouge. B Pourcentage d'identité de séquence entre les protéines dans un alignement multiple.

Le génome de la souche bactérienne *Rhodospirillaceae bacterium* HHTR118 à été rendu public en septembre 2018 (Institut de bioénergie et bioprocédés de Qing Dao, Chine) à partir d'un isolat d'une algue verte (laitue de mer ; *Ulva*). Sachant que des macro-algues peuvent produire des méthanes chlorés dont le DCM et des dérivés avec du brome ou de l'iode (Laturnus *et al.*, 2004, 2010 ; Bravo-Linares *et al.*, 2010), on peut émettre l'hypothèse que ce cette bactérie se nourrit ou se protège des émissions des algues. À ce jour, la capacité de dégradation du DCM (ou d'autres méthanes chlorés) de cette souche n'a pas été déterminée.

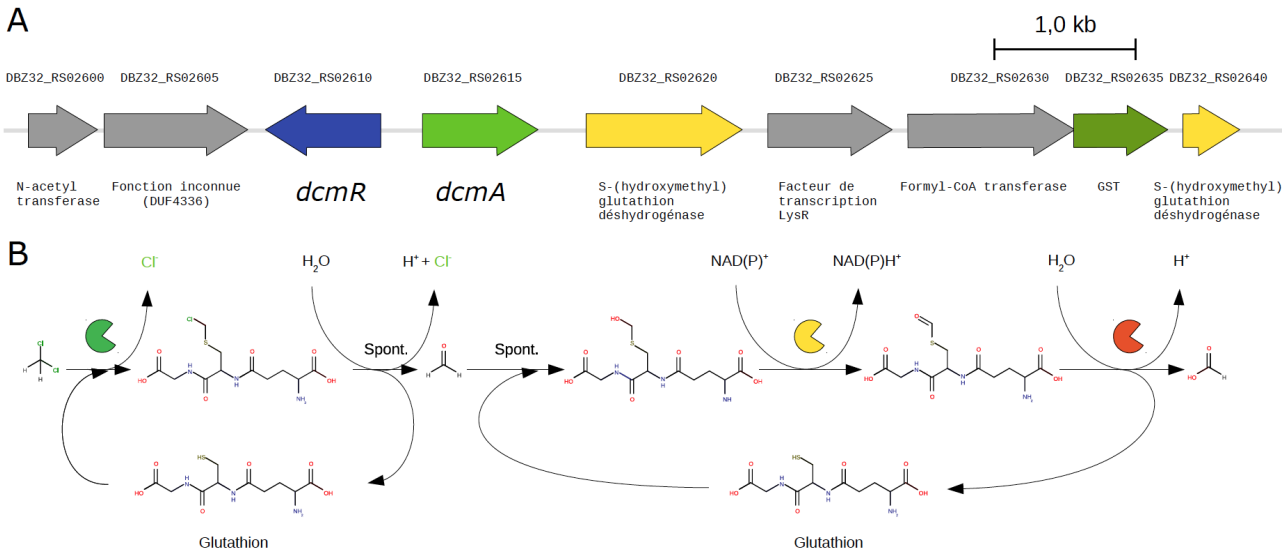


Figure 5.8: Environnement génétique des gènes *dcmA* et *dcmR-like* et voie métabolique potentielle chez *Rhodospirillaceae bacterium* HHTR118.
A Locus du génome de *Rhodospirillaceae bacterium* HHTR118 (position 226 158 à 234 930 ; numéro d'accession NZ_QXNY01000002.1). B Le formaldéhyde produit par DcmA pourrait réagir spontanément avec du glutathion et être métabolisé par une enzyme S-(hydroxymethyl)glutathion déshydrogénase encodée par un gène (en jaune) en aval de *dcmA*. Un gène avec une fonction prédite similaire est présent dans ce même groupe de gènes à côté d'une GST potentielle (en vert). La dernière enzyme, de cette voie métabolique, produisant du formiate, la S-formylglutathion hydrolase n'est pas détectée dans ce groupe de gènes.

En comparant les régions intergéniques et les promoteurs de *dcmR* et *dcmA* dans différents contextes, on note que le motif conservé de 12 pb entre les boites -35 et -10 n'est détecté chez certaines souches ayant l'îlot *dcm* avec une enzyme DcmA de groupe « A ». Chez *M. leisingerii* DM11, où seul *dcmA* a été séquencé, le promoteur ne possède pas ce motif. Cela peut laisser supposer un mécanisme de régulation différent. Il est également possible que chez *M. leisingerii* DM11, la protéine DcmR soit adaptée pour reconnaître une autre séquence génétique. Dans le cas de *Rhodospirillaceae bacterium* HHTR118, et en utilisant le logiciel Promoter Hunter (Klucar *et al.*, 2010), un chevauchement des promoteurs des gènes *dcmR* et *dcmA* est prédite, une situation fréquente qui favorise la synchronisation de la régulation des deux gènes du fait de l'encombrement stérique de l'ARN polymérase lorsqu'elle se fixe à l'un des promoteurs (Beck & Warren, 1988).

L'ensemble des analyses bioinformatiques indique que le gène *dcmR* peut être retrouvé dans plusieurs contextes génétiques, associé à des métabolismes autres que la dégradation du DCM, et

faisant souvent intervenir une association avec un gène potentiel de synthèse de la biotine, dont un lien fonctionnel avec le gène *dcmR* n'a pas encore été établi. Cependant, un lien possible entre des fonctions associés à la biotine (Lin & Cronan, 2012) et le métabolisme méthylotrophe est le rôle de la biotine comme cofacteur pour des réactions de carboxylation et de décarboxylation (Chapman-Smith & Cronan, 1999), comme précédemment suggéré lorsque des gènes *bio* on été associés à l'îlot génomique portant le transposon *dcm* (Muller *et al.*, 2011).

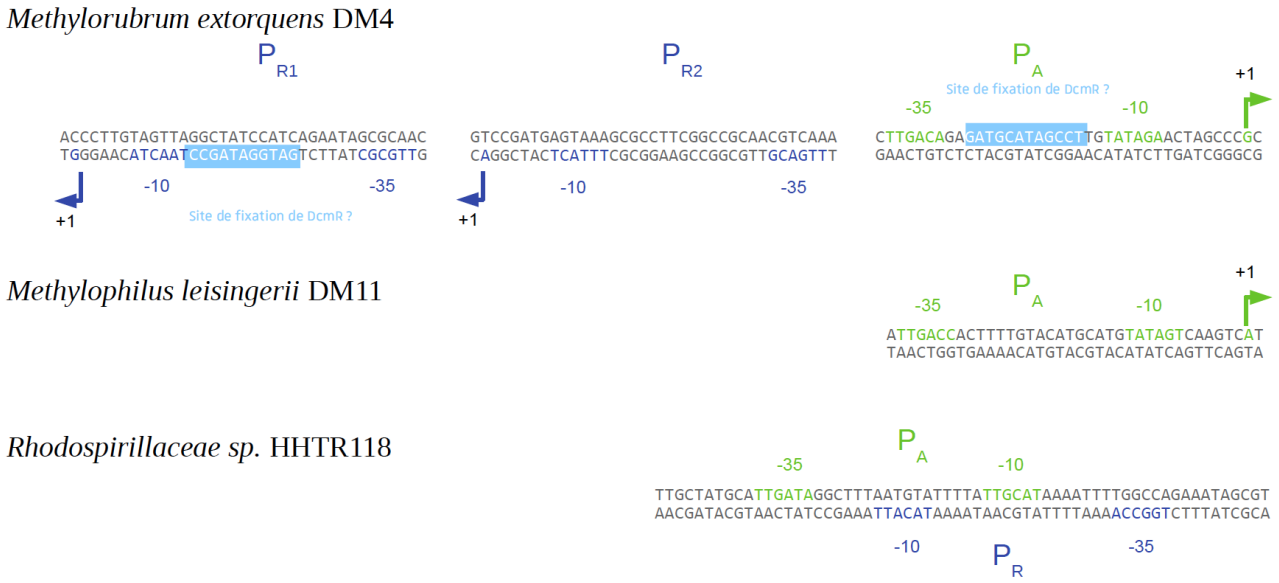


Figure 5.9: Diversité des promoteurs des gènes *dcmR* et *dcmA*.
Les promoteurs de *M. extorquens* DM4 et *M. leisingerii* DM11 ont été déterminés expérimentalement par la détection du site d'initiation de la transcription (La Roche & Leisinger, 1991 ; Bader & Leisinger, 1994). Les promoteurs de *Rhodospirillaceae bacterium* HHTR118 ont été prédits avec le logiciel Promoteur Hunter (Klucar *et al.*, 2010), mais n'ont pas encore été validés expérimentalement.

5.3.3 Quantification du gène *dcmR* dans l'environnement

Le gène *dcmR* a été recherché et quantifié à partir de l'ADN métagénomique extrait d'échantillons de deux sites industriels en France, géographiquement distincts et indépendants. Les couples d'amorces TnA1r/f et TnR1r/f ont été utilisées (Chaignaud *et al.*, 2017). Il s'agit d'amorces spécifiques aux gènes *dcmA* et *dcmR* qui ont été dessinés pour s'hybrider avec les différentes séquences connues de ces deux gènes. Les gènes *dcmR* et *dcmA* ont pu être quantifiés à partir des deux environnements pollués (Fig. 10).

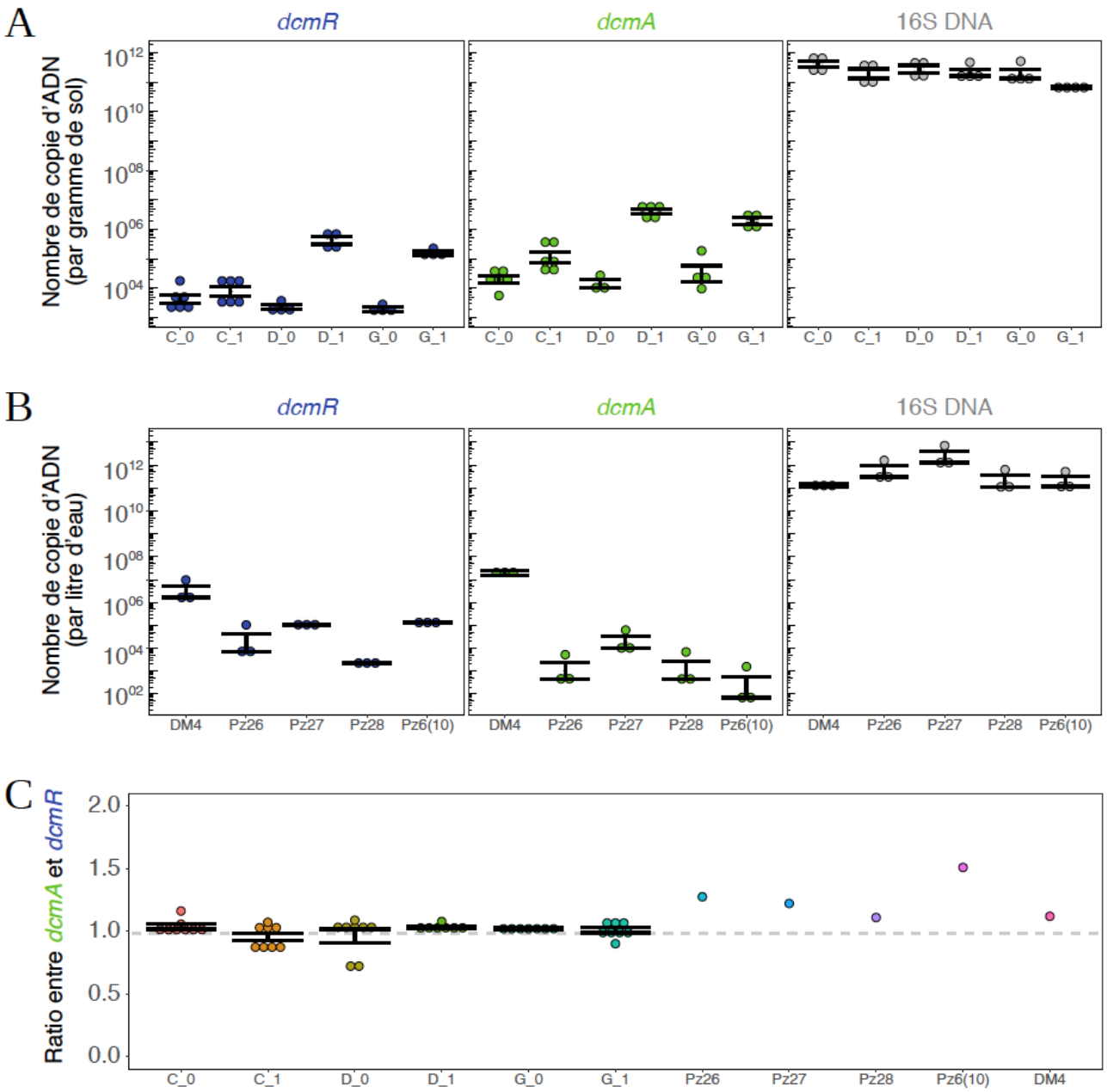


Figure 5.10: Quantification des gènes *dcmR* et *dcmA* à partir d'échantillons environnementaux.
A Quantification des gènes *dcmR* et *dcmA* dans le sol des microcosmes TOP-DCM. Deux réplicats techniques ont été réalisés à partir de six extractions d'ADN indépendantes. Les suffixes « _0 » et « _1 » sont ajoutés au nom des microcosmes pour indiquer si l'échantillon a été prélevé au début ou à la fin de l'expérience respectivement. B Quantification des gènes *dcmR* et *dcmA* dans l'eau de l'aquifère contaminée du site de Themeroil. Trois réplicats techniques ont été réalisés à partir d'une seule extraction d'ADN. Le témoin DM4 correspond à la suspension de *M. extorquens* DM4 filtrée. C Ratio des gènes *dcmA* / *dcmR* dans différents échantillons environnementaux.

Le gène *dcmR* a été détecté dans tous les échantillons testés. Pour les échantillons de l'eau de l'aquifère de Themeroil, les résultats pour le gène *dcmA* confirment ceux obtenus lors des campagnes précédentes (Hermon *et al.*, 2018). Dans les deux cas, le puits Pz27, dans la zone source, est celui qui présente la plus grande quantité de *dcmA* tandis que Pz6(10), dans le panache de pollution, présente la quantité la plus faible. Les résultats pour le gène de l'ARNr 16S sont plus élevés (10^{11} copies par L) que ceux obtenus lors des campagnes précédentes (moyenne de 10^{9} copies par L lors des campagnes en septembre 2015 et 2016 ; février 2016 et 2017). Cela pourrait être dû à la différence des amorces qui ont été utilisées : 799f / 1193r contrairement à BACT1369f / PROK1492r (Hermon *et al.*, 2018). Elles ciblent chacune des régions variables de l'ADN différentes: V2 à V4 pour BACT1369f / PROK1492r et V5 à V7 pour 7999f / 1193r. Une correction pourrait être réalisée en tenant compte de l'efficacité d'amplification de chaque amorce. Le couple d'amorces 799f / 1193r a été développé pour limiter les contaminations par les séquences d'ARN ribosomiques des chloroplastes (Bodenhausen *et al.*, 2013). Elles ont été choisies dans le but d'éviter des contaminations dans le cas d'échantillons venant de sol en contact avec des matières végétales. La détection des gènes *dcmA* et *dcmR* dans tous les microcosmes du projet TOP-DCM, à savoir avant et après la culture, indique qu'ils étaient présents dans la communauté bactérienne naturelle du sol. Dans les microcosmes D et G où du DCM a été ajouté, le nombre de copie de *dcmA* et *dcmR* augmente de façon significative entre le début et la fin de l'expérience. Cela signifie qu'en présence de DCM, les souches portant ces gènes ont pu se développer.

Comme les gènes *dcmA* et *dcmR* ont pu être détectés dans tous les échantillons testés, le ratio entre *dcmR* et *dcmA* a été calculé en tenant compte de la quantité d'ADN détectée et du facteur d'efficacité de la qPCR (Škulj *et al.*, 2008). Ce ratio varie de 0,72 à 1,5 et il est très proche de 1 pour les échantillons des microcosmes TOP-DCM (Fig. 10C). Les résultats sont plus variables pour les échantillons de Themeroil et pourraient être mieux validés par d'autres extractions d'ADN suivies d'expériences d'amplification complémentaires. Pour le témoin avec la suspension de *M. extorquens* DM4 possède un ratio proche de 1. Il semble donc probable que, dans ces environnements comme dans les souches caractérisées précédemment, les gènes *dcmR* et *dcmA* soient associés.

5.4 Conclusions et perspectives

Dans le cadre de cette étude, le gène *dcmR* a été mis en évidence pour la première fois dans des échantillons d'ADN issus de l'environnement. Le gène *dcmR* pourrait être envisagé comme biomarqueur pour l'étude de la dégradation du DCM en complément de *dcmA*. En effet, ce gène n'a été trouvé que dans le génome de souches dégradant le DCM même si une protéine a été prédite dans d'autres contextes, il n'y a pas de conservation au niveau nucléotidique. Par ailleurs, la présence de régions de la protéine DcmR très conservées dans le domaine MEDS laisse supposer qu'elles remplissent un rôle dans la fonction de ce domaine. Ces données seront utiles pour une analyse biochimique de ce domaine. En revanche, le gène *dcmR* ne semble donc pas être associé systématiquement au gène *dcmA*, et pourrait être trouvé dans d'autres contextes génétiques. En effet, il s'avère que ces deux gènes sont peu conservés au niveau nucléotidique. Ainsi, il faudrait sans doute dessiner de nouvelles amorces pour les détecter par PCR et les étudier. D'autres sites pollués pourraient être étudiés pour confirmer la présence simultanée des gènes *dcmR* et *dcmA* et de l'intérêt de ces gènes comme biomarqueurs de la dégradation du DCM.

Au-delà de la recherche des gènes et bactéries connues pour leur rôle dans la dégradation du DCM, il serait également possible, par métagénomique, de déterminer quelles sont les bactéries présentes dans les microcosmes du projet TOP-DCM, voire d'identifier celles qui sont impliquées dans la dégradation du DCM. De fait, une simple approche de métagénomique ciblant le gène codant l'ARNr 16S permettrait de déterminer si la composition microbienne des microcosmes change lors d'une exposition au DCM. En revanche, ces informations ne permettraient pas de déterminer quelles sont les espèces bactériennes qui dégradent le DCM même si des changements de fréquences sont détectés. Une approche de *stable isotope probing* (SIP) (Radajewski *et al.*, 2000) en utilisant du DCM marqué avec un isotope stable (^{13}C) permettrait de répondre à cette question. Une telle approche a déjà été réalisée pour un autre méthane chloré : le chlorométhane dans un sol forestier (Chaignaud *et al.*, 2018). Dans ce contexte, l'utilisation de nouvelles approches permettant d'extraire simultanément de l'ADN, des ARN et des protéines à partir d'un seul échantillon (Roume *et al.*, 2015) semblerait la plus prometteuse. En effet, une telle approche combinant la détection de ces trois types de molécules permettrait d'observer non seulement le potentiel variant des gènes *dcmR* et *dcmA* et de leurs produits protéiques, mais aussi de détecter une éventuelle régulation de

l'expression génétique *in situ*, non seulement des gènes associés à l'utilisation du DCM connus comme *dcmA* et *dcmR*, mais de l'ensemble des gènes de la communauté microbienne des environnements étudiés.

5.5 Annexe : article de Bringel *et al*. paru en 2017

L'article de Bringel *et al*., 2017 publié dans le journal *Genome Announcement* est présent ci-dessous. Dans le cadre de cette thèse, le transposon catabolique *dcm* chez *Hyphomicrobium* sp. GJ21 a été étudié et annoté.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Genome Sequence of the Dichloromethane-Degrading Bacterium *Hyphomicrobium* sp. Strain GJ21

Françoise Bringel,[a] Christiaan P. Postema,[b] Sophie Mangenot,[c] Sabrina Bibi-Triki,[a] Pauline Chaignaud,[a] Muhammad Farhan Ul Haque,[a] Christelle Gruffaz,[a] Louis Hermon,[a] Yousra Louhichi,[a] Bruno Maucourt,[a] Emilie E. L. Muller,[a] Thierry Nadalig,[a] Aurélie Lajus,[d,e,f] Zoé Rouy,[d,e,f] Claudine Médigue,[d,e,f] Valérie Barbe,[c] Dick B. Janssen,[b] Stéphane Vuilleumier[a]

Université de Strasbourg, CNRS, GMGM UMR 7156, Department of Microbiology, Genomics and the Environment, Strasbourg, France[a]; Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands[b]; Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Institut de Biologie François-Jacob, Laboratoire de Biologie Moléculaire pour l'Etude des Génomes, Évry, France[c]; Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Institut de Biologie François-Jacob, Laboratoire d'Analyse Bioinformatique en Génomique et Métabolisme, Évry, France[d]; Centre National de la Recherche Scientifique, UMR 8030, Évry, France[e]; UEVE, Université d'Évry, Évry, France[f]

ABSTRACT The genome sequence of *Hyphomicrobium* sp. strain GJ21, isolated in the Netherlands from samples of environments contaminated with halogenated pollutants and capable of using dichloromethane as its sole carbon and energy source, was determined.

Received 15 May 2017 **Accepted** 30 May 2017 **Published** 27 July 2017

Citation Bringel F, Postema CP, Mangenot S, Bibi-Triki S, Chaignaud P, Farhan Ul Haque M, Gruffaz C, Hermon L, Louhichi Y, Maucourt B, Muller EEL, Nadalig T, Lajus A, Rouy Z, Médigue C, Barbe V, Janssen DB, Vuilleumier S. 2017. Genome sequence of the dichloromethane-degrading bacterium *Hyphomicrobium* sp. strain GJ21. Genome Announc 5:e00622-17. https://doi.org/10.1128/genomeA.00622-17.

Copyright © 2017 Bringel et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Stéphane Vuilleumier, vuilleumier@unistra.fr.

Dichloromethane (DCM) is one of the most used industrial halogenated solvents worldwide, and its toxicity and substantial release into the environment are of serious concern (1, 2). Various methylotrophic bacterial strains capable of growing with DCM as the sole source of carbon and energy have been isolated over the years (3). All such strains for which the dehalogenase has been characterized, including strain GJ21 (4), contain the *dcmA* gene for DCM dehalogenase, an enzyme of the glutathione *S*-transferase family (5, 6). Other DCM-degrading systems, particularly from anaerobic environments, are currently being investigated in detail (7–10), including at the genome level. Whereas natural sources of DCM exist, such as production by microalgae (11), most DCM in the environment is of industrial origin, raising questions about the origin, evolution, and distribution of DCM dehalogenase genes in microorganisms.

Hyphomicrobium sp. strain GJ21, isolated after prolonged adaptation from a mixture of activated sludge from a wastewater treatment plant and soil samples from polluted sites (12), has been extensively used as a model for bioremediation of DCM in bioreactor studies (e.g., see reference 13 and references therein). The genome of *Hyphomicrobium* sp. strain GJ21 was sequenced using Illumina technology. A mate-paired library (7-kb insert size) and a paired-end library (340-bp insert size) were produced and sequenced using HiSeq2000 (2 × 100 nucleotides), yielding ~4 Gb and ~1.85 Gb of useful reads after selection and trimming, respectively. Sequence data were assembled with the Velvet assembler (https://www.ebi.ac.uk/~zerbino/velvet). Gap filling was performed using GapCloser (http://soap.genomics.org.cn/soapdenovo.html) on scaffolds larger than 2 kb. The final assembly resulted in 1 circular scaffold comprising 4 contigs with a cumulative size of 3.84 Mb and a G+C content of 60.8%. Automatic and manual gene annotations were performed using the MicroScope platform (14). The genome contains a single rRNA operon and encodes 47 tRNAs for all

Downloaded from http://genomea.asm.org/ on June 21, 2018 by SCD UNIV LOUIS PASTEUR

amino acids. It features genes for enzymes participating in methanol, methylamine, dimethylamine, and trimethylamine oxidation and associated accessory genes; for enzymes involved in formaldehyde oxidation by the tetrahydromethanopterin pathway, enzymes of the tetrahydrofolate-dependent C1 pathway, and corresponding cofactor biosynthesis genes; and for enzymes involved in formate oxidation, as well as genes of the serine and ethylmalonyl-CoA cycles for carbon assimilation (15). The genome also possesses the strongly conserved *dcm* islet of *dcmRABC* genes found in all bacteria with *dcmA*-encoded DCM dehalogenase (3), including *Hyphomicrobium denitrificans* ATCC 51888 of known genome sequence (16). Genes associated with dissimilatory nitrate reduction and complete denitrification to N_2 were also identified, as expected from the known ability of *Hyphomicrobium* strains to grow with DCM with nitrate as the terminal electron acceptor in the absence of oxygen (17).

Several other *Hyphomicrobium* genome sequences of isolated strains (16, 18–21) (http://www.genoscope.cns.fr/agc/microscope) or reconstructed from metagenomes (22) are now publicly available. Comparative genomics of the *Hyphomicrobium* genus will support ongoing experimental studies on bacterial adaptation to growth with halogenated methanes for bioremediation applications.

Accession number(s). The *Hyphomicrobium* sp. strain GJ21 genome sequence was deposited in GenBank under the accession number CDHO00000000. The version described here is the first version, CDHO01000000.

ACKNOWLEDGMENTS

This work was supported by a GIS IBiSA grant-in-kind to S.V. The LABGeM (CEA/IG/Genoscope and CNRS UMR 8030) and the France Génomique national infrastructure (funded as part of the Investissements d'avenir program managed by the Agence Nationale pour la Recherche, contract ANR-10-INBS-09) are acknowledged for support within the MicroScope annotation platform.

REFERENCES

1. Shestakova M, Sillanpää M. 2013. Removal of dichloromethane from ground and wastewater: a review. Chemosphere 93:1258–1267. https://doi.org/10.1016/j.chemosphere.2013.07.022.
2. Schlosser PM, Bale AS, Gibbons CF, Wilkins A, Cooper GS. 2015. Human health effects of dichloromethane: key findings and scientific issues. Environ Health Perspect 123:114–119. https://doi.org/10.1289/ehp.1308030.
3. Muller EEL, Bringel F, Vuilleumier S. 2011. Dichloromethane-degrading bacteria in the genomic age. Res Microbiol 162:869–876. https://doi.org/10.1016/j.resmic.2011.01.008.
4. Vuilleumier S, Ivoš N, Dean M, Leisinger T. 2001. Sequence variation in dichloromethane dehalogenases/glutathione S-transferases. Microbiology 147:611–619. https://doi.org/10.1099/00221287-147-3-611.
5. Vuilleumier S, Pagni M. 2002. The elusive roles of bacterial glutathione S-transferases: new lessons from genomes. Appl Microbiol Biotechnol 58:138–146. https://doi.org/10.1007/s00253-001-0836-0.
6. Allocati N, Federici L, Masulli M, Di Ilio C. 2009. Glutathione transferases in bacteria. FEBS J 276:58–75. https://doi.org/10.1111/j.1742-4658.2008.06743.x.
7. Lee M, Wells E, Wong YK, Koenig J, Adrian L, Richnow HH, Manefield M. 2015. Relative contributions of *Dehalobacter* and zerovalent iron in the degradation of chlorinated methanes. Environ Sci Technol 49:4481–4489. https://doi.org/10.1021/es5052364.
8. Kleindienst S, Higgins SA, Tsementzi D, Konstantinidis KT, Mack EE, Löffler FE. 2016. Draft genome sequence of a strictly anaerobic dichloromethane-degrading bacterium. Genome Announc 4(2):e00037-16. https://doi.org/10.1128/genomeA.00037-16.
9. Kleindienst S, Higgins SA, Tsementzi D, Chen G, Konstantinidis KT, Mack EE, Löffler FE. 2017. "*Candidatus* Dichloromethanomonas elyunquensis" gen. nov., sp. nov., a dichloromethane-degrading anaerobe of the *Peptococcaceae* family. Syst Appl Microbiol 40:150–159. https://doi.org/10.1016/j.syapm.2016.12.001.
10. Trueba-Santiso A, Parladé E, Rosell M, Lliros M, Mortan SH, Martínez-Alonso M, Gaju N, Martín-González L, Vicent T, Marco-Urrea E. 2017. Molecular and carbon isotopic characterization of an anaerobic stable enrichment culture containing *Dehalobacterium* sp. during dichloromethane fermentation. Sci Total Environ 581:640–648. https://doi.org/10.1016/j.scitotenv.2016.12.174.
11. Paul C, Pohnert G. 2011. Production and role of volatile halogenated compounds from marine algae. Nat Prod Rep 28:186–195. https://doi.org/10.1039/c0np00043d.
12. Ottengraf SPP, Meesters JJP, Vandenoever AHC, Rozema HR. 1986. Biological elimination of volatile xenobiotic compounds in biofilters. Bioproc Eng 1:61–69. https://doi.org/10.1007/bf00387497.
13. Diks RMM, Ottengraf SPP, Vrijlnad S. 1994. The existence of a biological equilibrium in a trickling filter for waste-gas purification. Biotechnol Bioeng 44:1279–1287. https://doi.org/10.1002/bit.260441103.
14. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, Mercier J, Renaux A, Rollin J, Rouy Z, Roche D, Scarpelli C, Médigue C. 2017. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res 45:D517–D528. https://doi.org/10.1093/nar/gkw1101.
15. Ochsner AM, Sonntag F, Buchhaupt M, Schrader J, Vorholt JA. 2015. *Methylobacterium extorquens*: methylotrophy and biotechnological applications. Appl Microbiol Biotechnol 99:517–534. https://doi.org/10.1007/s00253-014-6240-3.
16. Brown PJB, Kysela DT, Buechlein A, Hemmerich C, Brun YV. 2011. Genome sequences of eight morphologically diverse alphaproteobacteria. J Bacteriol 193:4567–4568. https://doi.org/10.1128/JB.05453-11.
17. Kohler-Staub D, Frank S, Leisinger T. 1995. Dichloromethane as the sole carbon source for *Hyphomicrobium* sp. strain DM2 under denitrification conditions. Biodegradation 6:229–235. https://doi.org/10.1007/BF00700462.
18. Vuilleumier S, Nadalig T, Ul Haque MF, Magdelenat G, Lajus A, Roselli S, Muller EEL, Gruffaz C, Barbe V, Médigue C, Bringel F. 2011. Complete genome sequence of the chloromethane-degrading *Hyphomicrobium* sp. strain MC1. J Bacteriol 193:5035–5036. https://doi.org/10.1128/JB.05627-11.
19. Venkatramanan R, Prakash O, Woyke T, Chain P, Goodwin LA, Watson D, Brooks S, Kostka JE, Green SJ. 2013. Genome sequences for three denitri-

Downloaded from http://genomea.asm.org/ on June 21, 2018 by SCD UNIV LOUIS PASTEUR

fying bacterial strains isolated from a uranium- and nitrate-contaminated subsurface environment. Genome Announc 1(4):e00449-13. https://doi.org/10.1128/genomeA.00449-13.
20. Martineau C, Villeneuve C, Mauffrey F, Villemur R. 2014. Complete genome sequence of *Hyphomicrobium nitrativorans* strain NL23, a denitrifying bacterium isolated from biofilm of a methanol-fed denitrification system treating seawater at the Montreal Biodome. Genome Announc 2(1):e01165-13. https://doi.org/10.1128/genomeA.01165-13.
21. Martineau C, Mauffrey F, Villemur R. 2015. Comparative analysis of denitrifying activities of *Hyphomicrobium nitrativorans*, *Hyphomicrobium denitrificans*, and *Hyphomicrobium zavarzinii*. Appl Environ Microbiol 81:5003–5014. https://doi.org/10.1128/AEM.00848-15.
22. Fujinawa K, Asai Y, Miyahara M, Kouzuma A, Abe T, Watanabe K. 2016. Genomic features of uncultured methylotrophs in activated-sludge microbiomes grown under different enrichment procedures. Sci Rep 6:26650. https://doi.org/10.1038/srep26650.

Downloaded from http://genomea.asm.org/ on June 21, 2018 by SCD UNIV LOUIS PASTEUR

6. Conclusions et perspectives

Chapitre 6 : Conclusions et perspectives

6.1 Conclusions

L'ensemble des résultats obtenus au cours de cette thèse indique que la régulation du métabolisme du DCM chez *M. extorquens* DM4 est plus complexe que précédemment pensé. Il a également été démontré que la réponse adaptative à la croissance sur le DCM fait intervenir des mécanismes de régulation bien au-delà des quatre gènes portés par le transposon catabolique *dcm* en impliquant une centaine de gènes de fonctions diverses (Chaignaud *et al.*, 2017 ; Bibi-Triki *et al.*, 2018) qui confirment un premier modèle d'adaptation à la croissance sur le DCM proposé précédemment (Muller *et al*., 2011). D'autre part, ce travail a également démontré une diversité de séquences et de contextes génétiques bien au-delà de celle connue au départ de cette thèse, avec une conservation de séquence nucléotidique du gène *dcmR* au sein du transposon *dcm* catabolique chez diverses bactéries. Alors que le gène *dcmR* a pu être détecté dans des échantillons environnementaux pollués au DCM dans un rapport un à un avec le gène *dcmA*, l'interrogation des banques de données a permis de détecter des gènes homologues à *dcmR* associés ou non au gène *dcmA*. Ces résultats permettent d'ouvrir de nouvelles pistes de recherches qui peuvent s'articuler en trois volets selon des postulats de travail qui seront discutés dans la suite de ce chapitre.

1 proposition d'un nouveau modèle d'étude de la régulation du transposon catabolique *dcm* chez *M. extorquens* DM4 intégrant l'ensemble des données acquises au cours de cette thèse

2 proposition de nouvelles pistes de recherches de fonctions adaptatives à la croissance avec du DCM basées sur l'identification de gènes ou de protéines lors des approches « -omiques »

3 proposition d'utilisation de DcmR voire d'autres fonctions de régulation pour le suivi et l'identification des verrous limitant la décontamination de sites pollués au DCM

6.2 Nouveau modèle d'étude de la régulation du transposon catabolique *dcm* chez *M. extorquens* DM4

Ce modèle considère deux situations opposées : la faible expression des gènes *dcm* en absence de DCM (répression) et la forte expression des gènes *dcm* (activation).

En absence de DCM, le rôle de *dcmR* dans la répression de sa propre expression et celle de *dcmA* a été confirmé (Fig. 1). Il a été démontré que pour le gène *dcmR*, cette répression fait intervenir P_{R1}, un seul des deux promoteurs de *dcmR*., le promoteur P_{R2} étant exprimé de façon constitutive en présence ou absence de DcmR. Chez *E. coli*, la répression des gènes *dcmR* et *dcmA* requiert la présence de DcmR. Mais elle est indépendante de la présence des séquences conservées *dcm* dans les promoteurs de *dcmR* et *dcmA, s*uggérant qu'elles n'interviennent pas dans la répression. Cette répression pourrait faire intervenir *dcmABC* puisqu'un mutant Δ*dcmA* n'est plus régulé, avec un phénotype similaire à celui du mutant *dcmR*. Il pourrait s'agir de DcmA qui aurait une activité de régulation en plus de son activité enzymatique ou alors d'une déstabilisation de l'ARNm *dcmABC* qui impacterait en plus de *dcmA*, l'expression des gènes en aval de *dcmA*, à savoir les gènes *dcmBC*. Le rôle du domaine MEDS dans l'activité de répression de DcmR et la nature de son ligand n'ont pas pu être élucidés au cours de cette thèse. L'ensemble des données tend à proposer un nouveau modèle de la régulation dépendante de *dcmR* en condition de répression des gènes *dcm*.

Lors de croissance en condition DCM, le gène *dcmR* n'est pas essentiel mais son absence ralentit la vitesse de croissance dans cette condition de culture. Ce ralentissement n'est pas dû à une différence de production des protéines DcmA, DcmB ou DcmC. Il est peut-être dû à un effet sur des gènes en dehors de l'îlot *dcm*, tel que MEDTI3043, dont l'expression est régulée par le DCM (Fig. 1). L'activation de l'expression des gènes *dcmABC* nécessite la présence de toute la région intergénique entre *dcmR* et *dcmA* pour avoir lieu. Cela pourrait être dû au besoin d'une présence simultanée de sites de régulation au niveau des promoteurs des gènes *dcmR* et *dcmA* pour la fixation d'un facteur de transcription. En effet, deux sites distants peuvent intervenir dans la régulation de la transcription en formant une boucle au niveau de l'ADN (Hensel *et al.*, 2013). Chez les bactéries, ce mécanisme a été décrit pour plusieurs facteurs de transcription : AraC,CI, LacI et NtrC (Cournac & Plumbridge, 2013). Les deux séquences de fixation peuvent être séparées de 58 pb à plus de 5 kb (Hensel *et al.*, 2013). La présence des séquences conservées de 12bp dans la région promotrice des promoteurs P_A

et P_{R1} pourrait être impliquée dans l'activation de l'expression des gènes *dcm*. Pour tester cette hypothèse et identifier les séquences de régulation impliquées dans l'activation des gènes *dcm*, l'expression de fusion transcriptionnelle à partir de plasmides dérivés de pME8312 et portant des séquences de 12 bp mutées ou non, pourront être testées chez *M. extorquens* DM4 sauvage et les mutants de *dcmR* et d'autres gènes *dcm*. Pour valider ce modèle de régulation, plusieurs pistes de recherche sont proposées en perspectives.

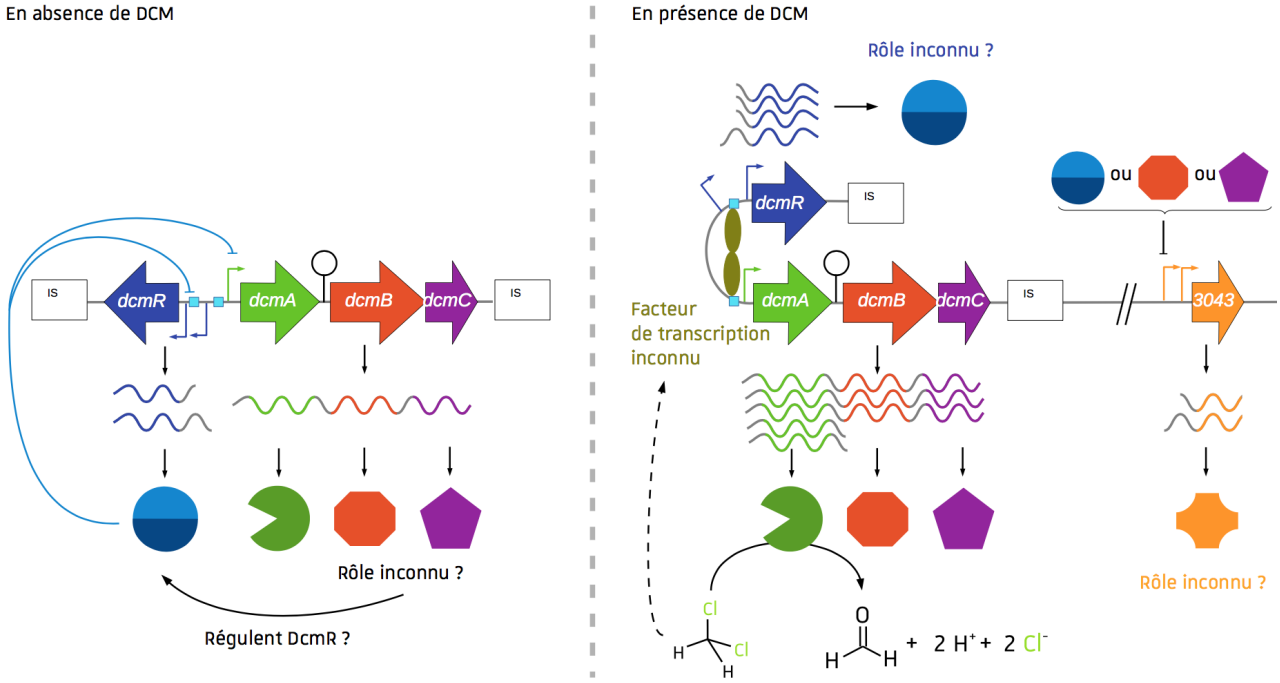


Figure 6.1: Mécanismes potentiels de régulation du transposon catabolique *dcm*.
DcmR agit comme répresseur de son propre promoteur et celui de *dcmA* en absence de DCM. Par contre en présence de DCM, la répression est abolie par un ligand encore inconnu tandis qu'une activation de l'expression de *dcmA* indépendante de DcmR a lieu par un facteur inconnu.

6.2.1 De nouvelles stratégies pour purifier la protéine DcmR ?

Produire la protéine DcmR permettra de mieux étudier son fonctionnement et notamment déterminer quel est le ligand impliqué dans cette régulation. La protéine DcmR n'a pas pu être purifiée chez *E. coli* ou *M. extorquens* DM4, de façon fonctionnelle par l'utilisation d'étiquette histidine en N ou C terminal. Une purification avec une étiquette *Mannose Binding Protein* (MalE) en N-ter s'était également soldée par un échec (Schmid-Appert, 1996). Pour purifier cette protéine, d'autres étiquettes telles que SUMO ou calmoduline (Kimple *et al.*, 2013) pourraient être testées

ainsi que la production chez d'autres souches (*E. coli* ou *M. extorquens* spp.). Une approche alternative serait de purifier la protéine DcmR de façon native et sans étiquette, en utilisant par exemple une chromatographie d'affinité avec des fragments d'ADN fixés à une colonne (Tacheny *et al.*, 2013). Pour DcmR, la séquence entière ou partielle de la région intergénique entre *dcmR* et *dcmA* pourrait être fixée sur une colonne. L'ajout d'ADN compétiteurs permettrait d'éviter les fixations aspécifiques (ex : poly (dI-dC)). La réussite de cette approche dépend de la capacité à reconstituer *in vitro* les bonnes conditions de fixation à l'ADN de DcmR.

6.2.2 Impact des métabolites du DCM dans la régulation transcriptionnelle ?

L'activation de l'expression de *dcmA* indépendante de *dcmR* a lieu seulement en présence de DCM comme seule source de carbone et n'a pas lieu en présence de méthanol, ni lorsque le DCM est présent avec un autre substrat comme le méthanol (chapitre 4). Il est donc possible que le signal reconnu dans ce cas ne soit pas directement le DCM mais l'un de ses intermédiaires métaboliques, dont certains sont très réactifs. Il est peu probable que le S-chlorométhylglutathion (Fig. 2) soit directement reconnu comme ligand car c'est un composé réactif avec une durée de demie-vie qui serait inférieure à cinq millisecondes (Kayser & Vuilleumier, 2001). Lors de l'étude de la toxicité du DCM, plusieurs substituts avec des durées de vie plus longues et un effet toxique similaire ont été synthétisés (Blocki *et al.*, 1994 ; Marsch *et al.*, 2001). Une hypothèse pourrait être que ce n'est pas le métabolite intermédiaire en lui-même qui est reconnu comme ligand mais l'alkylation qu'il provoque qui est importante dans cette régulation. Un tel mécanisme est connu pour Ada, un facteur de transcription qui reconnaît l'alkylation de l'ADN et répare l'ADN en éliminant ces adduits (Mielecki & Grzesiuk, 2014). Une fois le groupement méthyl détecté, l'enzyme Ada le transfère sur sa propre surface ce qui modifie sa conformation et lui permet d'acquérir sa fonction de régulateur en se fixant au niveau de son propre promoteur et celui d'*alkAB* et *aidB* (Mielecki & Grzesiuk, 2014).

Un tel rôle dans la réparation de l'ADN pourrait aussi s'appliquer à DcmR et expliquer pourquoi les mutants du gène *dcmR* ont une croissance ralentie en présence de DCM alors qu'il y a autant de protéines DcmA, DcmB et DcmC. Dans ce cas, ce ne serait pas la présence de DCM qui lèverait la répression de DcmR mais la détection d'adduits à l'ADN liés au métabolisme de ce polluant. Pour vérifier cette hypothèse, il serait possible de comparer les dommages à l'ADN chez la souche

sauvage et chez un mutant du gène *dcmR*. S'il est impliqué dans la réparation en plus de la régulation, on s'attendrait à voir plus de dommages chez le mutant ne produisant plus cette protéine. Pour quantifier ces dommages, il est par exemple possible d'utiliser une protéine rapportrice détectant les cassures doubles brins (Shee *et al.*, 2013).

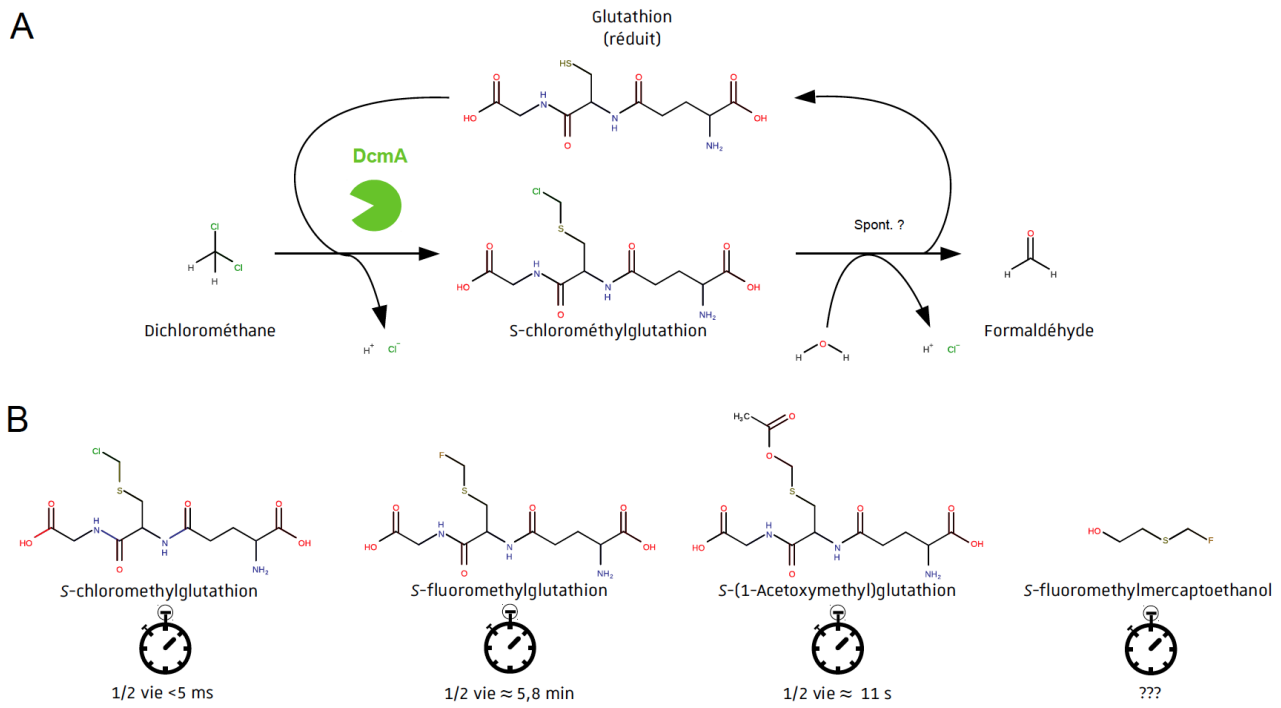


Figure 6.2: Ligands possibles impliqués dans la régulation de l'îlot *dcm*.
A Réaction de dégradation du DCM catalysée par DcmA B Durée de vie d'agents alkylants mimant les effets du S-chloromethylglutathion : S-fluorométhylglutathion (Kayser & Vuilleumier, 2001), S-(1-acétméthyl)glutathion (Marsch *et al.*, 2001), S-fluorométhylmercaptoéthanol (Blocki *et al.*, 1994).

Par ailleurs, les alkylations provoquées par la dégradation du DCM par les GST de la famille de DcmA et *theta* (Marsch *et al.*, 2001 ; Watanabe *et al.*, 2007) pourraient également avoir un impact sur la régulation de l'îlot. Ces adduits à l'ADN sont l'un des facteurs principaux de la toxicité du DCM observés *in vitro* et *in vivo* (Evans *et al.*, 2000 ; Kayser *et al.*, 2000 ; Kayser & Vuilleumier, 2001). Ces alkylations pourraient aussi modifier l'expression de certains gènes par des mécanismes d'épigénétique en changeant l'appariement entre deux nucléotides et ainsi l'ouverture de l'ADN lors de l'initiation de la transcription (Adhikari & Curtis, 2016). Cet effet pourrait être testé. Il serait par exemple possible d'utiliser un composé, tel que le méthanesulfonate de méthyle, réalisant des alkylations similaires à celles du S-chlorométhyglutathion (Beranek, 1990) puis de comparer par RT-qPCR ou fusion transcriptionnelle l'expression de certains gènes en présence de ces autres

agents alkylants. Les nouvelles techniques de séquençage sur une molécule unique et native d'ADN (telle que le MinION de Nanopore Oxford) permettent de mettre en évidence des modifications épigénétiques (Rand *et al.*, 2017). Cette technique permettrait d'identifier si les sites d'alkylation à l'ADN associés au métabolisme du DCM sont répartis de façon homogène sur l'ensemble du génome ou concentrés sur une zone particulière, comme par exemple du tranposon catabolique *dcm*.

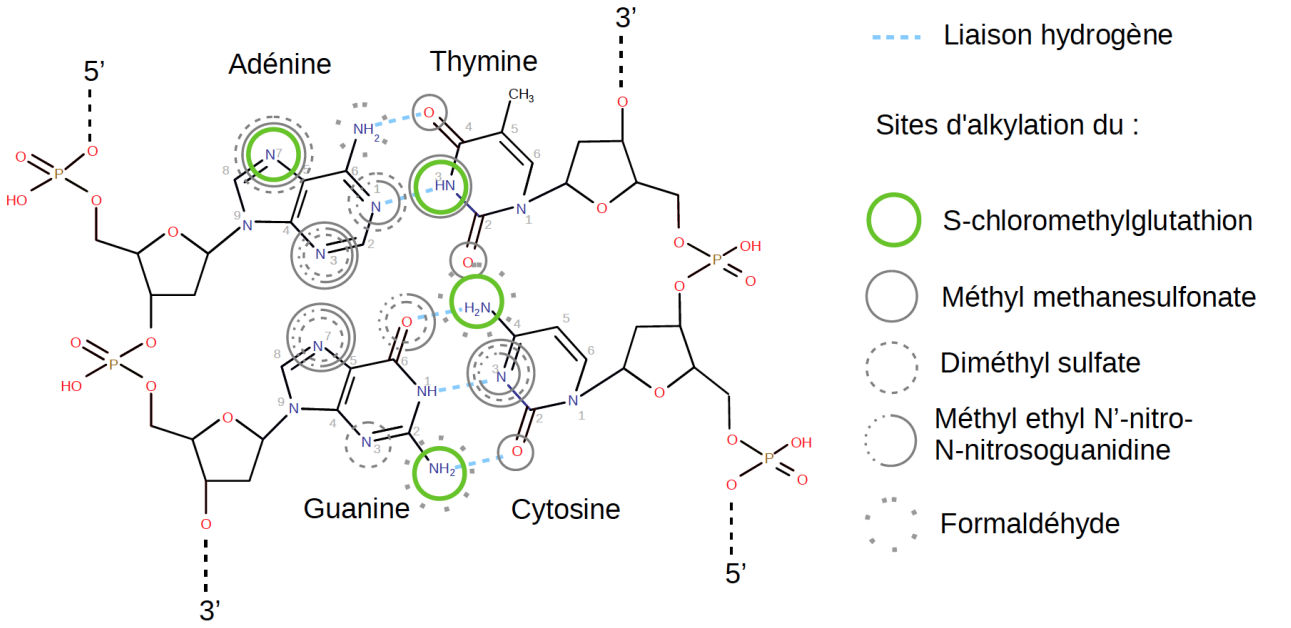


Figure 6.3: Alkylation de l'ADN par des agents chimiques.
Représentation d'une molécule d'ADN avec les sites d'alkylation connus du S-cholorométhyglutathion (Marsch *et al.*, 2001) ainsi que d'autres agents alkylants (Beranek, 1990). L'alkylation peut conduire à des mutations mais aussi bloquer le passage de certaines protéines telles que l'ADN polymérase ou l'ARN polymérase.

L'alkylation du S-chlorométhylglutathion pourrait également avoir lieu au niveau de DcmA ou d'autres protéines cellulaires. L'alkylation des protéines peut conduire à une inactivation de l'activité enzymatique et empêcher des modifications post-traductionnelles (Liebler, 2008). De tels dommages ont déjà été observés suite au métabolisme de substrats variés par différents types d'enzymes (Asimov & Dawson, 1950 ; Liebler, 2008 ; Codreanu *et al.*, 2014). Des méthodes de spectrométrie de masse permettraient de détecter de tels adduits sur des protéines. Néanmoins, des méthodes de purification sont souvent nécessaires au préalable car les alkylations n'ont lieu que sur une fraction des protéines et qu'il est difficile de différencier celles alkylées ou non (Liebler, 2008 ; Codreanu *et al.*, 2014). Ces analyses permettront de savoir si la grande quantité de protéine DcmA produite sert à compenser une perte de fonction d'une partie des protéines par l'alkylation.

L'alkylation de facteurs de transcription pourrait peut-être expliquer en partie les régulations au niveau transcriptionnel observées en présence de DCM.

6.3 Nouvelles pistes de recherches de fonctions adaptatives à la croissance sur le DCM

Les approches de « -omiques » ont permis de montrer que l'adaptation au DCM faisait intervenir des gènes et protéines en dehors du transposon catabolique *dcm*. Cela a pu être confirmé pour le gène METDI3043. En plus de permettre une meilleure croissance en présence de DCM, ils pourraient aussi avoir un lien, direct ou indirect, avec les gènes du transposon *dcm* comme METDI3043.

6.3.1 METDI3043 un nouveau gène impliqué dans l'adaptation en présence de DCM ?

Le gène METDI3043 a été mis en évidence comme régulé en présence de DCM : avec un $\log_2$ fc de 4,4 en transcriptomique tandis qu'en protéomique il n'a été détecté qu'en présence de DCM (Chaignaud *et al*., 2017; Bibi-Triki *et al*., 2018). Des résultats de RT-qPCR chez les mutants des gènes *dcm* ont confirmé ces résultats et indiqué que *dcmR* ou *dcmBC* pourraient être impliqués dans sa répression. La protéine codée par ce gène METDI3043 n'a pas de fonction connue à ce jour mais possède deux domaines CBS (cystathione β-synthase) qui sont connus pour intervenir dans la régulation de certaines enzymes en fixant de ligands dérivés de l'adénine tels que l'AMP ou ATP (Ereño-Orbea *et al.*, 2013 ; Anashkin *et al.*, 2017). Cela est confirmé par la présence de ces deux ligands au niveau des protéines cristallisées possédant ce domaine (King *et al.*, 2008). Il est supposé que ce domaine permette de détecter le niveau d'énergie cellulaire par la fixation d'AMP ou d'ATP. Lorsque le ligand est fixé, la protéine pourrait transmettre un signal à une autre protéine par le biais d'interactions protéines - protéines. Le domaine CBS permettrait également la fixation à de l'ADN et l'ARN simple brin pour l'enzyme IMPDH (inosine 5'-monophosphate déshydrogénase) chez *E. coli* (McLLean *et al.*, 2004) et chez la drosophile chez qui elle rentre dans le noyau cellulaire (Kozhevnikova *et al.*, 2012). Cette fixation est due uniquement aux domaines CBS et non pas aux domaines catalytiques de l'enzyme. Il existe certaines protéines, comme METDI3043, qui possèdent uniquement ce domaine (dupliqué) sans autre domaine (King *et al.*, 2008 ; Anashkin *et al.*, 2017). Une telle protéine chez *Methanocaldococcus jannaschii* (MJ0729) est capable de se fixer

à l'ADN double brin (Aguado-Llera *et al.*, 2010). Ainsi, le domaine CBS dupliqué permet aussi la fixation à l'ADN. L'ensemble de ces informations permet d'émettre l'hypothèse que la protéine METDI3043 serait impliquée dans la détection de l'état énergétique de la cellule par fixation d'AMP/ATP et la transmission de ce signal par une interaction avec une autre protéine ou l'ADN

La première étape pour étudier ce gène chez *M. extorquens* DM4 serait de le muter par délétion ou par remplacement allélique du gène sauvage par un gène muté (Fig. 4). Cela permettra d'observer si l'absence de ce gène confère un phénotype lors de la croissance avec du DCM. Plusieurs protéines avec des domaines CBS ont été détectées comme fixant l'ADN. Cela pourrait être confirmé chez *M. extorquens* DM4 avec la protéine METDI3043 en la produisant et purifiant puis en faisant des retards sur gel avec de l'ADN simple ou double brin (par exemple avec ses promoteurs). Le ligand de METDI3043 pourrait également être testé par des approches de protéomique basées sur la différence de stabilité d'une protéine en présence / absence de ligands (Chang *et al.*, 2012). Dans ce cas, il serait possible d'étudier la protéine en présence de plusieurs molécules dérivées de l'adénine ou d'extraits cellulaires de *M. extorquens* DM4 fractionnés par chromatographie.

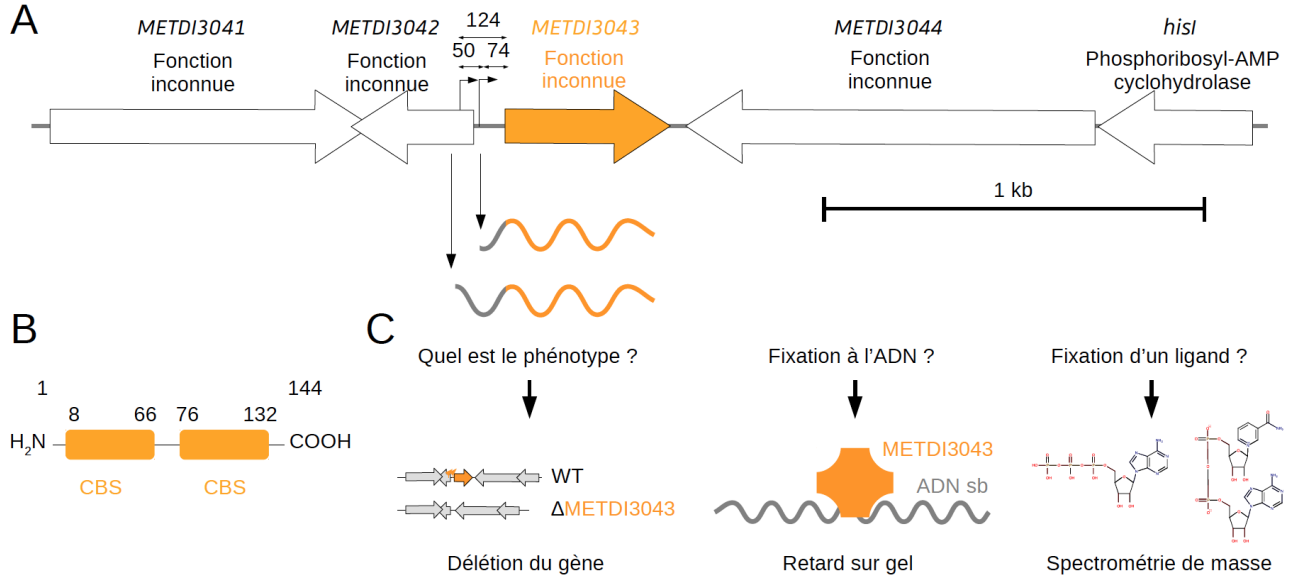


Figure 6.4: Perspectives pour l'étude de METDI3043.
A. Contexte génétique du gène METDI3043 dans le génome de *M. extorquens* DM4. La présence de deux promoteurs permettrait de produire deux ARNm pour ce gène. B La protéine METDI3043 est composée de deux domaines CBS. C Questions scientifiques et expériences proposées pour y répondre.

6.3.2 Les métabolites cellulaires ont-ils un rôle dans l'adaptation à la croissance avec le DCM ?

Les approches de « -omiques » réalisées chez *M. extorquens* DM4 ont permis de mettre en évidence des différences en terme d'expression d'ARNm et de production de protéines selon les conditions de culture (Chaignaud *et al.*, 2017 ; Bibi-Triki *et al.*, 2018). Il reste encore une inconnue, les métabolites. En effet, la production différentielle de certaines enzymes peut moduler les flux métaboliques ce qui peut modifier à son tour la croissance bactérienne ou la transcription de certains gènes. Par exemple le ribulose-1,5-bisphosphate, un intermédiaire du cycle de Calvin (fixation du CO_2), a été récemment détecté comme régulant le métabolisme du méthanol (méthylotrophie) chez *M. extorquens* PA1 (Ochsner *et al.*, 2017). Une approche de métabolomique permettrait de doser ces différents composés et éventuellement de détecter d'autres métabolites différentiellement produits (Fig. 5). La construction de mutants des gènes du métabolisme de ces métabolites permettrait de voir s'ils ont un effet lors de la croissance en présence de DCM.

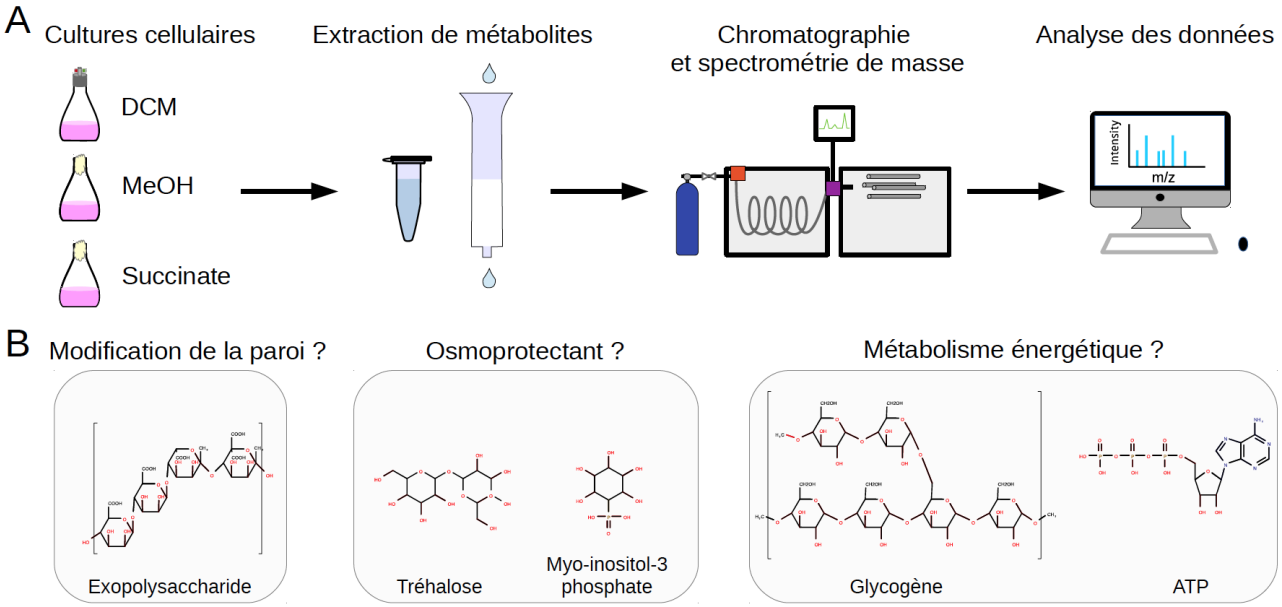


Figure 6.5: Approche de métabolomique pour mieux comprendre l'adaptation au DCM.
A Principe simplifié de la métabolomique. Les métabolites sont identifiés par spectrométrie de masse après différentes méthodes d'extraction et de séparation. B Molécules pouvant être recherchées lors de l'approche de métabolomique par rapport aux enzymes détectées comme plus abondantes en condition de croissance sur le DCM (Bibi-Triki *et al.*, 2018).

Cette approche de métabolomique pourrait aussi servir à quantifier l'AMP et l'ATP dans plusieurs conditions de culture pour voir si l'état énergétique de la cellule change (Fig. 5). Il est supposé que METDI3043 reconnaisse ces molécules comme ligands donc savoir s'il y a un changement du ratio

ATP/AMP en présence de DCM serait un indice sur la fonction de ce gène. Il serait possible de quantifier ces molécules à partir d'extraits cellulaires par chromatographie liquide (Jenkins *et al.*, 2013) ou par métabolomique (Weckmann *et al.*, 2018).

6.4 Intérêts de l'utilisation de DcmR dans le suivi de décontamination de sites pollués

Des facteurs de l'environnement tels que la bio-disponibilité du DCM, la présence d'autres substrats ou de contaminants, l'interaction avec d'autres microorganismes pourraient modifier la régulation des gènes *dcm* dans l'environnement par rapport aux facteurs mis en évidence *in labo*. Comprendre comment est régulé le génome de souches utilisées pour la dépollution est une étape qui permettra de mieux optimiser les stratégies de bioréhabilitation.

L'étude de *M. extorquens* DM4 *in labo* a permis d'établir un modèle de régulation du transposon catabolique *dcm*. Néanmoins, les conditions de cultures utilisées étaient simplifiées par rapport à celles trouvées dans l'environnement (Cf. chapitre 2). Dans un environnement pollué, de nombreux signaux et stress vont être détectés par les bactéries et par conséquent peuvent influencer la transcription des gènes *dcm* et d'autres gènes du génome. Cela pourrait se traduire par une mauvaise efficacité de la dégradation du DCM *in situ*, qui est non prévisible par des études faites seulement *in labo*.

La dépollution du DCM peut se faire *in situ* au niveau de sols ou d'aquifères contaminés mais aussi directement au niveau des effluents industriels en sortie d'usine. Les effluents de DCM étant principalement gazeux, les biofiltres sont l'une des approches couramment utilisées pour limiter la contamination de l'environnement (Shestakova & Sillanpää, 2013). Les effluents gazeux pollués passent à travers des biofiltres constitués d'une matrice minérale sur laquelle des micro-organismes se développent en dégradant les polluants (Yang *et al.*, 2018). La capacité de souches bactériennes à dégrader le DCM dans ces biofiltres a été démontrée (Ottengraf *et al.*, 1986 ; Hartmans and Tramper, 1991 ; Abtahi *et al.*, 2014 ; Ravi *et al.*, 2015 ; Naddafi *et al.*, 2016 ; Han *et al.*, 2018) mais aucune étude n'a étudié la transcription des gènes de dégradation. Dans un biofiltre, les conditions sont plus complexes qu'en condition d'étude *in labo* notamment car la composition de l'effluent (polluants et concentrations) peut varier au cours du temps et que les paramètres d'exploitation tels

que l'humidité, la recirculation de l'air ou du milieu nutritif peuvent impacter la physiologie bactérienne et donc l'expression des gènes de biodégradation. Au fur et mesure de l'utilisation du biofiltre, une communauté bactérienne complexe peut s'établir à l'intérieur (et même conduire à la perte de la souche d'intérêt).

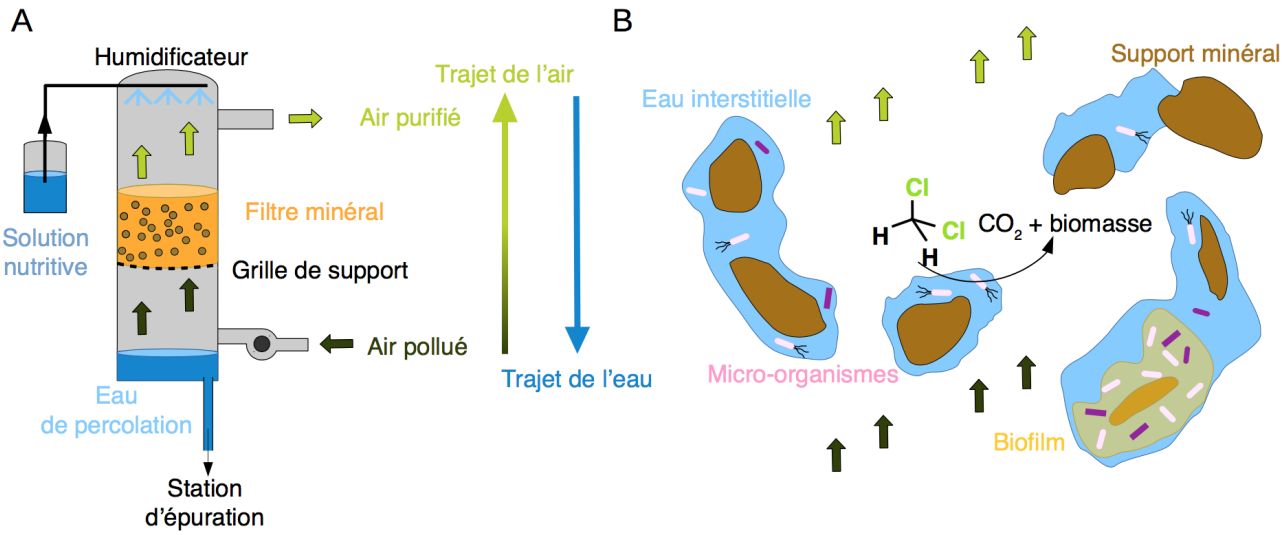


Figure 6.6: Élimination d'un polluant gazeux avec un biofiltre.
A Schéma d'un biofiltre proprement dit (sans recirculation de l'eau de percolation). L'air pollué passe par un filtre minéral dans lequel se développent des bactéries dégradant le polluant. Une solution nutritive est apportée par un humidificateur pour aider au développement bactérien. B Schéma de la croissance bactérienne sur le support minéral.

L'étude de l'expression du génome de *M. extorquens* DM4, ou d'une autre souche dégradant le DCM, pourrait se faire dans un tel biofiltre au cours de la bioréhabilitation d'effluents pollués au DCM. Une approche de méta-transcriptomique différentielle (Klingenberg & Meinicke, 2017) pourrait être mise en place pour comparer deux biofiltres avec des conditions différentes, et voir si un lien peut être établi entre la biodégradation et la régulation de la transcription et étudier l'apport du gène *dcmR* dans la régulation de ce métabolisme.

6.5 Conclusions à plus long terme

Ces dernières années, les efforts pour comprendre le métabolisme du DCM et les mécanismes moléculaires d'adaptation ont été centrés principalement sur la souche de référence *M. extorquens* DM4. Néanmoins, certains indices suggèrent que l'adaptation à l'utilisation DCM semble spécifique à chaque souche. Par exemple des structures membranaires composées de chlorure ont

été observée chez les souches *A. helveticus* DM6 et *A. methylovorans* DM10 mais pas chez *M. extorquens* DM4 (Torgonskaya *et al.*, 2011). L'apport du gène *dcmA*, par un plasmide, chez plusieurs bactéries du genre *Methylorubrum* ne permet pas forcément la croissance avec ce polluant et repose donc sur d'autres mécanismes spécifiques à chaque souche (Michener *et al.*, 2014). Découvrir de nouveaux gènes d'adaptation permettrait de faciliter l'ingénierie de nouvelles souches dégradantes plus efficaces (Dvořák *et al.*, 2017). Par exemple, une forte expression du transporteur de chlorure *clcA* favorise l'utilisation du DCM chez des souches du genre *Methylorubrum* (Michener *et al.*, 2014). Ainsi, l'étape de régulation peut être limitante et certains mécanismes d'adaptation impliquer une régulation au niveau de la transcription.

Les conditions physico-chimiques très différentes au laboratoire par rapport à celles observées dans l'environnement sont souvent invoquées pour expliquer la faible efficacité des projets de bioréhabilitation, mis en place à l'aide de souches développées au laboratoire(Govantes *et al.*, 2010 ; Vymazal, 2018). Une meilleure connaissance du métabolisme du DCM et de sa régulation dans la souche modèle *M. extorquens* DM4, mais aussi d'autres souches déchlorantes permettra d'avancer dans la compréhension des facteurs déterminant l'efficacité du métabolisme de dégradation du DCM voire d'autres polluants chlorés, et par là d'orienter à terme les stratégies de bioréhabilitation de sites pollués à mettre en place dans les années à venir.

7. Annexes

Bibliographie

Abhilash, P.C., Jamil, S., and Singh, N. (2009) Transgenic plants for enhanced biodegradation and phytoremediation of organic xenobiotics. *Biotechnol. Adv.* **27**: 474–488.

Adams, P.P., Flores Avile, C., Popitsch, N., Bilusic, I., Schroeder, R., Lybecker, M., and Jewett, M.W. (2017) *In vivo* expression technology and 5′ end mapping of the *Borrelia burgdorferi* transcriptome identify novel RNAs expressed during mammalian infection. *Nucleic Acids Res.* **45**: 775–792.

Adhikari, S. and Curtis, P.D. (2016) DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol. Rev.* **40**: 575–591.

Adrian, L., Rahnenfuhrer, J., Gobom, J., and Holscher, T. (2007) Identification of a chlorobenzene reductive dehalogenase in *Dehalococcoides* sp. strain CBDB1. *Appl. Environ. Microbiol.* **73**: 7717–7724.

Agarwal, V., Miles, Z.D., Winter, J.M., Eustaquio, A.S., El Gamal, A.A., and Moore, B.S. (2017) Enzymatic halogenation and dehalogenation reactions: Pervasive and mechanistically diverse. *Chem. Rev.* **117**: 5619–5674.

Aguado-Llera, D., Oyenarte, I., Martínez-Cruz, L.A., and Neira, J.L. (2010) The CBS domain protein MJ0729 of *Methanocaldococcus jannaschii* binds DNA. *FEBS Lett.* **584**: 4485–4489.

Albers, P., Weytjens, B., De Mot, R., Marchal, K., and Springael, D. (2018) Molecular processes underlying synergistic linuron mineralization in a triple-species bacterial consortium biofilm revealed by differential transcriptomics. *MicrobiologyOpen*.

Albersmeier, A., Pfeifer-Sancar, K., Rückert, C., and Kalinowski, J. (2017) Genome-wide determination of transcription start sites reveals new insights into promoter structures in the *Actinomycete Corynebacterium glutamicum*. *J. Biotechnol.* **257**: 99–109.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.

Alves, I.R., Lima-Noronha, M.A., Silva, L.G., Fernández-Silva, F.S., Freitas, A.L.D., Marques, M.V., and Galhardo, R.S. (2017) Effect of SOS-induced levels of *imuABC* on spontaneous and damage-induced mutagenesis in *Caulobacter crescentus*. *DNA Repair* **59**: 20–26.

Amman, F., Wolfinger, M.T., Lorenz, R., Hofacker, I.L., Stadler, P.F., and Findeiß, S. (2014) TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics* **15**: 89.

Amos, B.K., Ritalahti, K.M., Cruz-Garcia, C., Padilla-Crespo, E., and Loffler, F.E. (2008) Oxygen effect on *Dehalococcoides* viability and biomarker quantification. *Environ. Sci. Technol.* **42**: 5718–5726.

Anantharaman, V. and Aravind, L. (2005) MEDS and PocR are novel domains with a predicted role

in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. *Bioinformatics* **21**: 2805–2811.

Anashkin, V.A., Baykov, A.A., and Lahti, R. (2017) Enzymes regulated via cystathionine β-synthase domains. *Biochem. Mosc.* **82**: 1079–1087.

Andersen, M.E., Black, M.B., Campbell, J.L., Pendse, S.N., Clewell, H.J., Pottenger, L.H., *et al.* (2017) Combining transcriptomics and PBPK modeling indicates a primary role of hypoxia and altered circadian signaling in dichloromethane carcinogenicity in mouse lung and liver. *Toxicol. Appl. Pharmacol.* **332**: 149–158.

Ang, T.F., Maiangwa, J., Salleh, A., Normi, Y., and Leow, T. (2018) Dehalogenases: From improved performance to potential microbial dehalogenation applications. *Molecules* **23**: 1100.

Arnold, T.E., Yu, J., and Belasco, J.G. (1998) mRNA stabilization by the *ompA* 59 untranslated region: Two protective elements hinder distinct pathways for mRNA degradation. *RNA* **4**: 12.

Bader, R. and Leisinger, T. (1994) Isolation and characterization of the *Methylophilus* sp. strain DM11 gene encoding dichloromethane dehalogenase/glutathione S-transferase. **176**: 8.

Baharoglu, Z. and Mazel, D. (2014) SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol. Rev.* **38**: 1126–1145.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., *et al.* (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**: W202-208.

Bailón, L., Nikolausz, M., Kästner, M., Veiga, M.C., and Kennes, C. (2009) Removal of dichloromethane from waste gases in one- and two-liquid-phase stirred tank bioreactors and biotrickling filters. *Water Res.* **43**: 11–20.

Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B., and Breaker, R.R. (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**: 233–235.

Bartels, I. (1984) Suicide inactivation of catechol 2,3-dioxygenase from *Pseudomonas putida* mt-2 by 3-halocatechols. *Appl. Environ. Microbiol.* **47**: 500–505.

Beck, C.F. and Warren, R.A. (1988) Divergent promoters, a common form of gene organization. *Microbiol Rev* **52**: 318–326.

Beck, H.J. and Moll, I. (2018) Leaderless mRNAs in the spotlight: Ancient but not outdated! *Microbiol. Spectr.* **6**:.

Becker, J.G. (2006) A modeling study and implications of competition between *Dehalococcoides ethenogenes* and other tetrachloroethene- respiring bacteria. *Environ. Sci. Technol.* **40**: 4473–4480.

Beggah, S., Vogne, C., Zenaro, E., and van der Meer, J.R. (2007) Mutant HbpR transcription

activator isolation for 2-chlorobiphenyl via green fluorescent protein-based flow cytometry and cell sorting. *Microb. Biotechnol.* **0**: 071018044055002.

Belchik, S.M., Schaeffer, S.M., Hasenoehrl, S., and Xun, L. (2010) A β-barrel outer membrane protein facilitates cellular uptake of polychlorophenols in *Cupriavidus necator*. *Biodegradation* **21**: 431–439.

Benbrahim-Tallaa, L., Lauby-Secretan, B., Loomis, D., Guyton, K.Z., Grosse, Y., Ghissassi, F.E., *et al.* (2014) Carcinogenicity of perfluorooctanoic acid, tetrafluoroethylene, dichloromethane, 1,2-dichloropropane, and 1,3-propane sultone. *Lancet Oncol.* **15**: 924–925.

Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6**: 167.

Bernat, P., Nykiel-Szymanska, J., Stolarek, P., Slaba, M., Szewczyk, R., and Rozalska, S. (2018) 2,4-dichlorophenoxyacetic acid-induced oxidative stress: Metabolome and membrane modifications in *Umbelopsis isabellina*, a herbicide degrader. *PLoS ONE* **13**: e0199677.

Bers, K., Leroy, B., Breugelmans, P., Albers, P., Lavigne, R., Sorensen, S.R., *et al.* (2011) A novel hydrolase identified by genomic-proteomic analysis of phenylurea herbicide mineralization by *Variovorax* sp. strain SRS16. *Appl. Environ. Microbiol.* **77**: 8754–8764.

Bertaccini, D., Vaca, S., Carapito, C., Arsène-Ploetze, F., Van Dorsselaer, A., and Schaeffer-Reiss, C. (2013) An improved stable isotope N-terminal labeling approach with light/heavy TMPP to automate proteogenomics data validation: dN-TOP. *J. Proteome Res.* **12**: 3063–3070.

Beyer, A., Weist, A., Brangsch, H., Stoiber-Lipp, J., and Kothe, E. (2014) Aquifer microbiology at different geogenic settings for environmental biogeotechnology. In, *Microbial Biodegradation and Bioremediation*. Elsevier, pp. 555–575.

Bhandari, V. and Houry, W.A. (2015) Substrate interaction networks of the *Escherichia coli* chaperones: Trigger Factor, DnaK and GroEL. In, Krogan,N.J. and Babu,M. (eds), *Prokaryotic Systems Biology*. Springer International Publishing, Cham, pp. 271–294.

Bibi-Triki, S., Husson, G., Maucourt, B., Vuilleumier, S., Carapito, C., and Bringel, F. (2018) N-terminome and proteogenomic analysis of the *Methylobacterium extorquens* DM4 reference strain for dichloromethane utilization. *J. Proteomics* **179**: 131–139.

Bodenhausen, N., Horton, M.W., and Bergelson, J. (2013) Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS ONE* **8**: e56329.

Boopathy, R. (2000) Factors limiting bioremediation technologies. *Bioresour. Technol.* **74**: 63–67.

Bourne, W. and Stehle, R.L. (1923) Methylene chloride in anaesthesia. *Can. Med. Assoc. J.* **13**: 432–433.

Bradley, A.S., Swanson, P.K., Muller, E.E.L., Bringel, F., Caroll, S.M., Pearson, A., *et al.* (2017) Hopanoid-free *Methylobacterium extorquens* DM4 overproduces carotenoids and has widespread growth impairment. *PLoS ONE* **12**:.

Bravo-Linares, C.M., Mudge, S.M., and Loyola-Sepulveda, R.H. (2010) Production of volatile organic compounds (VOCS) by temperate macroalgae: The use of solid phase micro-extraction (SPME) coupled to GC-MS as method of analysis. *J. Chil. Chem. Soc.* **55**: 227–232.

Briat, J.F. and Chamberlin, M.J. (1984) Identification and characterization of a new transcriptional termination factor from *Escherichia coli*. *Proc. Natl. Acad. Sci.* **81**: 7373–7377.

Bringel, F., Postema, C.P., Mangenot, S., Bibi-Triki, S., Chaignaud, P., Haque, M.F.U., *et al.* (2017) Genome sequence of the dichloromethane-degrading bacterium *Hyphomicrobium* sp. strain GJ21. *Genome Announc.* **5**: e00622-17.

Brown, P.J.B., Kysela, D.T., Buechlein, A., Hemmerich, C., and Brun, Y.V. (2011) Genome sequences of eight morphologically diverse *Alphaproteobacteria*. *J. Bacteriol.* **193**: 4567–4568.

Brunner, W., Staub, D., and Leisinger, T. (1980) Bacterial degradation of dichloromethane. *Appl. Environ. Microbiol.* **40**: 950–958.

Bush, M. and Dixon, R. (2012) The role of bacterial enhancer binding proteins as specialized activators of sigma 54-dependent transcription. *Microbiol Mol Biol Rev* **76**: 497–529.

Byers, H.K. and Sly, L.I. (1993) Toxic effects of dichloromethane on the growth of methanotrophic bacteria. *FEMS Microbiol. Ecol.* 35–38.

Cai, M. and Xun, L. (2002) Organization and regulation of pentachlorophenol-degrading genes in *Sphingobium chlorophenolicum* ATCC 39723. *J Bacteriol* **184**: 4672–4680.

Cain, J.A., Solis, N., and Cordwell, S.J. (2014) Beyond gene expression: the impact of protein post-translational modifications in bacteria. *J. Proteomics* **97**: 265–286.

Cases, I. and de Lorenzo, V. (2005a) Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. *Int. Microbiol.* **8**: 213–222.

Cases, I. and de Lorenzo, V. (2005b) Promoters in the environment: transcriptional regulation in its natural context. *Nat. Rev. Microbiol.* **3**: 105–118.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**: 540–552.

Chae, J.C. and Zylstra, G.J. (2006) 4-Chlorobenzoate uptake in *Comamonas* sp. strain DJ-12 is mediated by a tripartite ATP-independent periplasmic transporter. *J. Bacteriol.* **188**: 8407–8412.

Chaignaud, P., Maucourt, B., Weiman, M., Alberti, A., Kolb, S., Cruveiller, S., et al. (2017) Genomic and transcriptomic analysis of growth-supporting dehalogenation of chlorinated methanes in *Methylobacterium*. *Front Microbiol* **8**:.

Chaignaud, P., Morawe, M., Besaury, L., Krober, E., Vuilleumier, S., Bringel, F., and Kolb, S. (2018) Methanol consumption drives the bacterial chloromethane sink in a forest soil. *ISME J.*

Chapman-Smith, A. and Cronan, J.E. (1999) The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem. Sci.* **24**: 359–363.

Charif, D., Thioulouse, J., Lobry, J.R., and Perrière, G. (2005) Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics* **21**: 545–547.

Chen, G., Kleindienst, S., Griffiths, D.R., Mack, E.E., Seger, E.S., and Loffler, F.E. (2017) Mutualistic interaction between dichloromethane- and chloromethane-degrading bacteria in an anaerobic mixed culture: Mutualistic degradation of chlorinated methanes. *Environ. Microbiol.*

Chen, G., Murdoch, R.W., Mack, E.E., Seger, E.S., and Löffler, F.E. (2017) Complete genome sequence of *Dehalobacterium formicoaceticum* strain DMC, a strictly anaerobic dichloromethane-degrading bacterium. *Genome Announc.* **5**:.

Chen, G., Shouakar-Stash, O., Phillips, E., Justicia-Leon, S.D., Gilevska, T., Sherwood Lollar, B., *et al.* (2018) Dual carbon–chlorine isotope analysis indicates distinct anaerobic dichloromethane degradation pathways in two members of *Peptococcaceae*. *Environ. Sci. Technol.*

Chen, K., Mu, Y., Jian, S., Zang, X., Chen, Q., Jia, W., *et al.* (2018) Comparative transcriptome analysis reveals the mechanism underlying 3,5-dibromo-4-hydroxybenzoate catabolism via a new oxidative decarboxylation pathway. *Appl Env. Microbiol* **84**: e02467–17.

Chen, N.H., Djoko, K.Y., Veyrier, F.J., and McEwan, A.G. (2016) Formaldehyde stress responses in bacterial pathogens. *Front. Microbiol.* **7**:.

Chistoserdova, L. and Kalyuzhnaya, M.G. (2018) Current trends in methylotrophy. *Trends Microbiol.* **26**: 703–714.

Choudhary, P.K., Duret, A., Rohrbach-Brandt, E., Holliger, C., Sigel, R.K.O., and Maillard, J. (2013) Diversity of cobalamin riboswitches in the corrinoid-producing organohalide respirer *Desulfitobacterium hafniense*. *J. Bacteriol.* **195**: 5186–5195.

Chubiz, L.M., Purswani, J., Carroll, S.M., and Marx, C.J. (2013) A novel pair of inducible expression vectors for use in *Methylobacterium extorquens*. *BMC Res. Notes* **6**: 183.

Clausen, L.P.W. and Trapp, S. (2017) Toxicity of 56 substances to trees. *Environ. Sci. Pollut. Res.* **24**: 18035–18047.

Coco, W.M., Parsek, M.R., and Chakrabarty, A.M. (1994) Purification of the LysR family regulator, ClcR, and its interaction with the *Pseudomonas putida clcABD* chlorocatechol operon promoter. *J Bacteriol* **176**: 5530–5533.

Collier, H. (1936) Methylene dichloride intoxication in industry. *The Lancet* **227**: 594–595.

Colomb, A., Yassaa, N., Williams, J., Peeken, I., and Lochte, K. (2008) Screening volatile organic compounds (VOCs) emissions from five marine phytoplankton species by head space gas chromatography/mass spectrometry (HS-GC/MS). *J. Environ. Monit.* **10**: 325–330.

Correia, M.A.S., Otrelo-Cardoso, A.R., Schwuchow, V., Sigfridsson Clauss, K.G.V., Haumann, M., Romão, M.J., *et al*. (2016) The *Escherichia coli* periplasmic aldehyde oxidoreductase is an exceptional member of the xanthine oxidase family of molybdoenzymes. *ACS Chem. Biol.* **11**: 2923–2935.

Couñago, R.M., Chen, N.H., Chang, C.W., Djoko, K.Y., McEwan, A.G., and Kobe, B. (2016) Structural basis of thiol-based regulation of formaldehyde detoxification in *H. influenzae* by a MerR regulator with no sensor region. *Nucleic Acids Res.* **44**: 6981–6993.

Cunningham, J.J., Kinner, N.E., and Lewis, M. (2009) Protistan predation affects trichloroethene biodegradation in a bedrock aquifer. *Appl. Environ. Microbiol.* **75**: 7588–7593.

Daniel, L., Buryska, T., Prokop, Z., Damborsky, J., and Brezovsky, J. (2015) Mechanism-based discovery of novel substrates of haloalkane dehalogenases using *in silico* screening. *J. Chem. Inf. Model.* **55**: 54–62.

Dar, D. and Sorek, R. (2018) High-resolution RNA 3′-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res.* **46**: 6797–6805.

De Marco, P. (2004) Methylotrophy versus heterotrophy: a misconception. *Microbiology* **150**: 1606–1607.

Deana, A., Celesnik, H., and Belasco, J.G. (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal. *Nature* **451**: 355–358.

DeMarini, D.M., Shelton, M.L., Warren, S.H., Ross, T.M., Shim, J.Y., Richard, A.M., and Pegram, R.A. (1997) Glutathione S-transferase-mediated induction of GC → AT transitions by halomethanes in *Salmonella*. *Environ. Mol. Mutagen.* **30**: 440–447.

Deponte, M. (2013) Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1830**: 3217–3266.

Di Martino, M.L., Romilly, C., Wagner, E.G.H., Colonna, B., and Prosseda, G. (2016) One gene and two proteins: a leaderless mRNA supports the translation of a shorter form of the *Shigella* VirF regulator. *mBio* **7**:.

Díaz-Marrero, A.R., Rovirosa, J., Darias, J., San-Martín, A., and Cueto, M. (2002) Plocamenols A−C, novel linear polyhalohydroxylated monoterpenes from *Plocamium cartilagineum*. *J. Nat. Prod.* **65**: 585–588.

Dorman, C.J. and Dorman, M.J. (2016) DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophys. Rev.* **8**: 209–220.

Doronina, N.V., Trotsenko, Y.A., Tourova, T.P., Kuznetsov, B.B., and Leisinger, T. (2000) *Methylopila helvetica* sp. nov. and *Methylobacterium dichloromethanicum* sp. nov. — novel aerobic facultatively methylotrophic bacteria utilizing dichloromethane. *Syst. Appl.*

Microbiol. **23**: 210–218.

Dvořák, P., Nikel, P.I., Damborský, J., and de Lorenzo, V. (2017) Bioremediation 3.0: Engineering pollutant-removing bacteria in the times of systemic biology. *Biotechnol. Adv.* **35**: 845–866.

Eckweiler, D., Dudek, C.-A., Hartlich, J., Brötje, D., and Jahn, D. (2018) PRODORIC2: the bacterial gene regulation database in 2018. *Nucleic Acids Res.* **46**: D320–D326.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.

El Mouali, Y. and Balsalobre, C. (2018) 3′untranslated regions: regulation at the end of the road. *Curr. Genet.*

Elvidge, L.C.E., Oram, D.E., Baker, A.K., Montzka, S.A., Humphrey, S., O'Sullivan, D.A., and Brenninkmeijer, C. a. M. (2015) Increasing concentrations of dichloromethane, CH_2Cl_2, inferred from CARIBIC air samples collected 1998–2012. *Atmospheric Chem. Phys.* **15**: 1939–1958.

Emanuelsson, M.A.E., Osuna, M.B., Ferreira Jorge, R.M., and Castro, P.M.L. (2009) Isolation of a *Xanthobacter* sp. degrading dichloromethane and characterization of the gene involved in the degradation. *Biodegradation* **20**: 235–244.

Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* **42**: 2646–2659.

Ettwiller, L., Buswell, J., Yigit, E., and Schildkraut, I. (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* **17**:.

Evans, G.J., Ferguson, G.P., Booth, I.R., and Vuilleumier, S. (2000) Growth inhibition of *Escherichia coli* by dichloromethane in cells expressing dichloromethane dehalogenase/glutathione S-transferase. *Microbiology* **146**: 2967–2975.

Farhan Ul Haque, M., Besaury, L., Nadalig, T., Bringel, F., Mutterer, J., Schaller, H., and Vuilleumier, S. (2017) Correlated production and consumption of chloromethane in the *Arabidopsis thaliana* phyllosphere. *Sci. Rep.* **7**: 17589.

Feng, Y., Bie, P., Wang, Z., Wang, L., and Zhang, J. (2018) Bottom-up anthropogenic dichloromethane emission estimates from China for the period 2005–2016 and predictions of future emissions. *Atmos. Environ.* **186**: 241–247.

Ferreira, M.I.M., Iida, T., Hasan, S.A., Nakamura, K., Fraaije, M.W., Janssen, D.B., and Kudo, T. (2009) Analysis of two gene clusters involved in the degradation of 4-fluorophenol by *Arthrobacter* sp. Strain IF1. *Appl. Environ. Microbiol.* **75**: 7767–7773.

Figurski, D.H. and Helinski, D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. *Proc. Natl. Acad. Sci. U. S. A.* **76**:

1648–1652.

Fincker, M. and Spormann, A.M. (2017) Biochemistry of catabolic reductive dehalogenation. *Annu. Rev. Biochem.* **86**: 357–386.

Firsova, J., Doronina, N., Lang, E., Spröer, C., Vuilleumier, S., and Trotsenko, Y. (2009) *Ancylobacter dichloromethanicus* sp. nov.--a new aerobic facultatively methylotrophic bacterium utilizing dichloromethane. *Syst. Appl. Microbiol.* **32**: 227–232.

Firsova, Y.E., Doronina, N.V., and Trotsenko, Y.A. (2010) Analysis of the key functional genes in new aerobic degraders of dichloromethane. *Microbiology* **79**: 66–72.

Firsova, Y.E., Doronina, N.V., and Trotsenko, Y.A. (2004) Physiological and biochemical analysis of the transformants of aerobic methylobacteria expressing the *dcmA* gene of dichloromethane dehalogenase. *Microbiology* **73**: 24–29.

Firsova, Y.E., Torgonskaya, M.L., and Trotsenko, Y.A. (2017) Functionality of *Metdi5511* gene in *Methylobacterium dichloromethanicum* DM4. *Appl. Biochem. Microbiol.* **53**: 194–200.

Frank, S.A. (2009) The common patterns of nature. *J. Evol. Biol.* **22**: 1563–1585.

Frascari, D., Zanaroli, G., and Danko, A.S. (2015) *In situ* aerobic cometabolism of chlorinated solvents: A review. *J. Hazard. Mater.* **283**: 382–399.

Fulton, G.L., Nunn, D.N., and Lidstrom, M.E. (1984) Molecular cloning of a malyl coenzyme A lyase gene from *Pseudomonas* sp. strain AM1, a facultative methylotroph. *J. Bacteriol.* **160**: 718–723.

Gaillard, M., Pernet, N., Vogne, C., Hagenbuchle, O., and van der Meer, J.R. (2008) Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. *Proc Natl Acad Sci U A* **105**: 7058–7063.

Gälli, R. and Leisinger, T. (1988) Plasmid analysis and cloning of the dichloromethane-utilization genes of *Methylobacterium* sp. DM4. *J. Gen. Microbiol.* **134**: 943–952.

Gälli, R. and Leisinger, T. (1985) Specialized bacterial strains for the removal of dichloromethane from industrial waste. *Conserv. Recycl.* **8**: 91–100.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2017) Phyletic distribution and lineage-specific domain architectures of archaeal two-component signal transduction systems. *J. Bacteriol.* **200**:.

Geldart, K., Borrero, J., and Kaznessis, Y.N. (2015) Chloride-inducible expression vector for delivery of antimicrobial peptides targeting antibiotic-resistant *Enterococcus faecium*. *Appl Env. Microbiol* **81**: 3889–3897.

Georg, J. and Hess, W.R. (2011) cis-Antisense RNA, another level of gene regulation in bacteria. *Microbiol. Mol. Biol. Rev.* **75**: 286–300.

Georg, J. and Hess, W.R. (2018) Widespread antisense transcription in *Prokaryotes*. *Microbiol. Spectr.* **6**:.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**: 343–345.

Gisi, D., Willi, L., Traber, H., Leisinger, T., and Vuilleumier, S. (1998) Effects of bacterial host and dichloromethane dehalogenase on the competitiveness of methylotrophic bacteria growing with dichloromethane. *Appl. Environ. Microbiol.* **64**: 1194–1202.

Gong, T., Xu, X., Che, Y., Liu, R., Gao, W., Zhao, F., *et al*. (2017) Combinatorial metabolic engineering of *Pseudomonas putida* KT2440 for efficient mineralization of 1,2,3-trichloropropane. *Sci. Rep.* **7**:.

Goodman, M.F. and Woodgate, R. (2013) Translesion DNA polymerases. *Cold Spring Harb. Perspect. Biol.* **5**:.

Goris, T., Schiffmann, C.L., Gadkari, J., Schubert, T., Seifert, J., Jehmlich, N., *et al*. (2015) Proteomics of the organohalide-respiring *Epsilonproteobacterium Sulfurospirillum multivorans* adapted to tetrachloroethene and other energy substrates. *Sci. Rep.* **5**:.

Govantes, F., Garcia-Gonzalez, V., Porrua, O., Platero, A.I., Jimenez-Fernandez, A., and Santero, E. (2010) Regulation of the atrazine-degradative genes in *Pseudomonas* sp. strain ADP. *FEMS Microbiol Lett* **310**: 1–8.

Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. U. S. A.* **87**: 4645–4649.

Green, P.N. and Ardley, J.K. (2018) Review of the genus *Methylobacterium* and closely related organisms: a proposal that some *Methylobacterium* species be reclassified into a new genus, *Methylorubrum* gen. nov. *Int. J. Syst. Evol. Microbiol.* **68**: 2727–2748.

Gribble, G.W. (2010) Naturally occurring organohalogen compounds - A comprehensive update Springer-Verlag, Wien.

Gruffaz, C., Muller, E.E.L., Louhichi-Jelail, Y., Nelli, Y.R., Guichard, G., and Bringel, F. (2014) Genes of the N-methylglutamate pathway are essential for growth of *Methylobacterium extorquens* DM4 with monomethylamine. *Appl Env. Microbiol* **80**: 3541–3550.

Gualerzi, C.O. and Pon, C.L. (2015) Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell. Mol. Life Sci.* **72**: 4341–4367.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**: 307–321.

Gvakharia, B.O., Permina, E.A., Gelfand, M.S., Bottomley, P.J., Sayavedra-Soto, L.A., and Arp,

D.J. (2007) Global transcriptional response of *Nitrosomonas* europaea to chloroform and chloromethane. *Appl Env. Microbiol* **73**: 3440–3445.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999) From molecular to modular cell biology. *Nature* **402**: C47–C52.

Hayes, R.P., Moural, T.W., Lewis, K.M., Onofrei, D., Xun, L., and Kang, C. (2014) Structures of the inducer-binding domain of pentachlorophenol-degrading gene regulator PcpR from *Sphingobium chlorophenolicum*. *Int J Mol Sci* **15**: 20736–20752.

Henrysson, T. and Mattiasson, B. (1993) A microbial biosensor system for dihalomethanes. *Biodegradation* **4**: 101–105.

Hermon, L., Denonfoux, J., Hellal, J., Joulian, C., Ferreira, S., Vuilleumier, S., and Imfeld, G. (2018) Dichloromethane biodegradation in multi-contaminated groundwater: Insights from biomolecular and compound-specific isotope analyses. *Water Res.* **142**: 217–226.

Hobman, J.L., Wilkie, J., and Brown, N.L. (2005) A design for life: Prokaryotic metal-binding MerR family regulators. *Biometals* **18**: 429–436.

Höfer, K. and Jäschke, A. (2018) Epitranscriptomics: RNA modifications in bacteria and archaea. *Microbiol. Spectr.* **6**:.

Hoffmann, D. and Muller, R.H. (2006) 2,4-Dichlorophenoxyacetic acid (2,4-D) utilization by *Delftia acidovorans* MC1 at alkaline pH and in the presence of dichlorprop is improved by introduction of the *tfdK* gene. *Biodegradation* **17**: 263–273.

Hosoda, A., Isomura, Y., Takeo, S., Onai, T., Takeuchi, K., Toda, M., and Tamura, H. (2017) Aerobic dechlorination of dichloromethane using biostimulation agent BD-C in continuous and batch cultures of *Xanthobacter autotrophicus* GJ10. *J. Oleo Sci.* **66**: 1247–1256.

Hossaini, R., Chipperfield, M.P., Montzka, S.A., Leeson, A.A., Dhomse, S.S., and Pyle, J.A. (2017) The increasing threat to stratospheric ozone from dichloromethane. *Nat. Commun.* **8**: 15962.

Hovasse, A., Bruneel, O., Casiot, C., Desoeuvre, A., Farasin, J., Hery, M., *et al.* (2016) Spatio-temporal detection of the *Thiomonas* population and the *Thiomonas* arsenite oxidase involved in natural arsenite attenuation processes in the Carnoulès acid mine drainage. *Front. Cell Dev. Biol.* **4**:.

Hu, Y., Kabler, S.L., Tennant, A.H., Townsend, A.J., and Kligerman, A.D. (2006) Induction of DNA–protein crosslinks by dichloromethane in a V79 cell line transfected with the murine glutathione-S-transferase theta 1 gene. *Mutat. Res. Toxicol. Environ. Mutagen.* **607**: 231–239.

Hua, A., Gueuné, H., Cregut, M., Thouand, G., and Durand, M.J. (2015) Development of a bacterial bioassay for atrazine and cyanuric acid detection. *Front Microbiol* **6**:.

Hug, L.A., Maphosa, F., Leys, D., Loffler, F.E., Smidt, H., Edwards, E.A., and Adrian, L. (2013) Overview of organohalide-respiring bacteria and a proposal for a classification system for

reductive dehalogenases. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120322–20120322.

Ibrahim, E.S., Kashef, M.T., Essam, T.M., and Ramadan, M.A. (2017) A degradome-based polymerase chain reaction to resolve the potential of environmental samples for 2,4-dichlorophenol biodegradation. *Curr. Microbiol.*

Islam, M.A., Waller, A.S., Hug, L.A., Provart, N.J., Edwards, E.A., and Mahadevan, R. (2014) New insights into *Dehalococcoides mccartyi* metabolism from a reconstructed metabolic network-based systems-level analysis of *D. mccartyi* transcriptomes. *PLoS ONE* **9**: e94808.

Ismaeil, M., Yoshida, N., and Katayama, A. (2017) Identification of multiple dehalogenase genes involved in tetrachloroethene-to-ethene dechlorination in a *Dehalococcoides*-dominated enrichment culture. *BioMed Res. Int.*

Jatsenko, T., Sidorenko, J., Saumaa, S., and Kivisaar, M. (2017) DNA polymerases ImuC and DinB are involved in DNA alkylation damage tolerance in *Pseudomonas aeruginosa* and *Pseudomonas putida*. *PLoS ONE* **12**: e0170719.

Jayaraj, R., Megha, P., and Sreedev, P. (2016) Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. *Interdiscip Toxicol* **9**: 90–100.

Jennings, L.K., Chartrand, M.M.G., Lacrampe-Couloume, G., Lollar, B.S., Spain, J.C., and Gossett, J.M. (2009) Proteomic and transcriptomic analyses reveal genes upregulated by cis-dichloroethene in *Polaromonas* sp. strain JS666. *Appl Env. Microbiol* **75**: 3733–3744.

Jeong, Y., Kim, J.-N., Kim, M.W., Bucca, G., Cho, S., Yoon, Y.J., *et al*. (2016) The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nat. Commun.* **7**: 11605.

Jianming, Y., Wei, L., Zhuowei, C., Yifeng, J., Wenji, C., and Jianmeng, C. (2014) Dichloromethane removal and microbial variations in a combination of UV pretreatment and biotrickling filtration. *J. Hazard. Mater.* **268**: 14–22.

Joyce, M.G., Levy, C., Gabor, K., Pop, S.M., Biehl, B.D., Doukov, T.I., *et al.* (2006) CprK crystal structures reveal mechanism for transcriptional control of halorespiration. *J. Biol. Chem.* **281**: 28318–28325.

Jugder, B.E., Ertan, H., Lee, M., Manefield, M., and Marquis, C.P. (2015) Reductive dehalogenases come of age in biological destruction of organohalides. *Trends Biotechnol.* **33**: 595–610.

Kaczmarczyk, A., Campagne, S., Danza, F., Metzger, L.C., Vorholt, J.A., and Francez-Charlot, A. (2011) Role of *Sphingomonas* sp. strain Fr1 PhyR-NepR-σEcfG cascade in general stress response and identification of a negative regulator of PhyR. *J. Bacteriol. J. Bacteriol.* **193**, **193**: 6629, 6629–6638.

Kahng, H.Y., Byrne, A.M., Olsen, R.H., and Kukor, J.J. (2000) Characterization and role of *tbuX* in utilization of toluene by *Ralstonia pickettii* PKO1. *J. Bacteriol.* **182**: 1232–1242.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., *et al.* (2018) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**: D335–D342.

Kanazawa, S. and Filip, Z. (1986) Effects of trichloroethylene, tetrachloroethylene and dichloromethane on enzymatic activities in soil. *Appl. Microbiol. Biotechnol.* **25**: 76–81.

Kappler, U. and Nouwens, A.S. (2013) The molybdoproteome of *Starkeya novella* – insights into the diversity and functions of molybdenum containing proteins in response to changing growth conditions. *Metallomics* **5**: 325.

Kato, Y., Asahara, M., Arai, D., Goto, K., and Yokota, A. (2005) Reclassification of *Methylobacterium chloromethanicum* and *Methylobacterium dichloromethanicum* as later subjective synonyms of *Methylobacterium extorquens* and of *Methylobacterium lusitanum* as a later subjective synonym of *Methylobacterium rhodesianum*. *J. Gen. Appl. Microbiol.* **51**: 287–299.

Kawanishi, M., Matsuda, T., and Yagi, T. (2014) Genotoxicity of formaldehyde: molecular basis of DNA damage and mutation. *Front. Environ. Sci.* **2**:.

Kayser, M.F. (2001) Genes and proteins associated with dichloromethane metabolism in *Methylobacterium dichloromethanicum* DM4, Thèse de l'université Swiss federal institute of technology Zürich.

Kayser, M.F., Stumpp, M.T., and Vuilleumier, S. (2000) DNA Polymerase I is essential for growth of *Methylobacterium dichloromethanicum* DM4 with dichloromethane. *J. Bacteriol.* **182**: 5433–5439.

Kayser, M.F., Ucurum, Z., and Vuilleumier, S. (2002) Dichloromethane metabolism and C1 utilization genes in *Methylobacterium* strains. *Microbiology* **148**: 1915–1922.

Kayser, M.F. and Vuilleumier, S. (2001) Dehalogenation of dichloromethane by dichloromethane dehalogenase/glutathione S-transferase leads to formation of DNA adducts. *J. Bacteriol.* **183**: 5209–5212.

Kim, H.H., Kim, S.R., Park, A., Han, T., and Choi, E.M. (2013) The role of glutathione in dichloromethane toxicity to *Lemna paucicostata*. *Toxicol. Environ. Health Sci.* **5**: 1–8.

Kim, S.H., Harzman, C., Davis, J.K., Hutcheson, R., Broderick, J.B., Marsh, T.L., and Tiedje, J.M. (2012) Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiol* **12**: 21.

King, N.P., Lee, T.M., Sawaya, M.R., Cascio, D., and Yeates, T.O. (2008) Structures and functional implications of an AMP-binding cystathionine beta-synthase domain protein from a hyperthermophilic archaeon. *J. Mol. Biol.* **380**: 181–192.

Kleindienst, S., Higgins, S.A., Tsementzi, D., Konstantinidis, K.T., Mack, E.E., and Löffler, F.E. (2016) Draft genome sequence of a strictly anaerobic dichloromethane-degrading bacterium. *Genome Announc.* **4**:.

Klucar, L., Stano, M., and Hajduk, M. (2010) phiSITE: database of gene regulation in bacteriophages. *Nucleic Acids Res.* **38**: D366–D370.

Koenig, J.C., Groissmeier, K.D., and Manefield, M.J. (2014) Tolerance of anaerobic bacteria to chlorinated solvents. *Microbes Env.* **29**: 23–30.

Kohler-Staub, D., Hartmans, S., Galli, R., Suter, F., and Leisinger, T. (1986) Evidence for identical dichloromethane dehalogenases in different methylotrophic bacteria. *J Gen Microbiol 132 1986 2837-2843* **132**:.

Kohler-Staub, D. and Leisinger, T. (1985) Dichloromethane dehalogenase of *Hyphomicrobium* sp. strain DM2. *J. Bacteriol.* **162**: 676–681.

Kolusu, S.R., Schlünzen, K.H., Grawe, D., and Seifert, R. (2018) Determination of chloromethane and dichloromethane in a tropical terrestrial mangrove forest in Brazil by measurements and modelling. *Atmos. Environ.* **173**: 185–197.

Korkmaz, G., Holm, M., Wiens, T., and Sanyal, S. (2014) Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J. Biol. Chem.* jbc.M114.606632.

Kozhevnikova, E.N., van der Knaap, J.A., Pindyurin, A.V., Ozgur, Z., van Ijcken, W.F.J., Moshkin, Y.M., and Verrijzer, C.P. (2012) Metabolic enzyme IMPDH is also a transcription factor regulated by cellular state. *Mol. Cell* **47**: 133–139.

Kranzioch, I., Ganz, S., and Tiehm, A. (2015) Chloroethene degradation and expression of *Dehalococcoides dehalogenase* genes in cultures originating from Yangtze sediments. *Environ. Sci. Pollut. Res.* **22**: 3138–3148.

Krasper, L., Lilie, H., Kublik, A., Adrian, L., Golbik, R., and Lechner, U. (2016) The MarR-type regulator Rdh2R regulates *rdh* gene transcription in *Dehalococcoides mccartyi* strain CBDB1. *J. Bacteriol.* **198**: 3130–3141.

Kröger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., Sivasankaran, S.K., Hokamp, K., *et al.* (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc. Natl. Acad. Sci. U. S. A.* **109**: E1277–E1286.

Kruse, T., Smidt, H., and Lechner, U. (2016) Comparative genomics and transcriptomics of organohalide-respiring bacteria and regulation of *rdh* gene transcription. In, *Organohalide-Respiring Bacteria*. Springer Berlin Heidelberg, pp. 345–376.

Kumar, A., Trefault, N., and Olaniran, A.O. (2014) Microbial degradation of 2,4-dichlorophenoxyacetic acid: Insight into the enzymes and catabolic genes involved, their regulation and biotechnological implications. *Crit. Rev. Microbiol.* 1–15.

Kuo, C. and Genthner, B. (1996) Effect of added heavy metal ions on biotransformation and biodegradation of 2-chlorophenol and 3-chlorobenzoate in anaerobic bacterial consortia. *Appl. Environ. Microbiol.* **62**: 2317–2323.

La Roche, S.D. and Leisinger, T. (1991) Identification of *dcmR*, the regulatory gene governing expression of dichloromethane dehalogenase in *Methylobacterium* sp. strain DM4. *J. Bacteriol.* **173**: 6714–6721.

Lacal, J., Reyes-Darias, J.A., Garcia-Fontana, C., Ramos, J. l., and Krell, T. (2013) Tactic responses to pollutants and their potential to increase biodegradation efficiency. *J. Appl. Microbiol.* **114**: 923–933.

Lang, G. and Ogawa, N. (2009) Mutational analysis of the inducer recognition sites of the LysR-type transcriptional regulator TfdT of *Burkholderia* sp. NK8. *Appl. Microbiol. Biotechnol.* **83**: 1085–1094.

Lang, K.S. and Johnson, T.J. (2015) Transcriptome modulations due to A/C2 plasmid acquisition. *Plasmid* **80**: 83–89.

Laturnus, F., Svensson, T., and Wiencke, C. (2010) Release of reactive organic halogens by the brown macroalga *Saccharina latissima* after exposure to ultraviolet radiation: Reactive organic halogens. *Polar Res.* **29**: 379–384.

Laturnus, F., Svensson, T., Wiencke, C., and Oberg, G. (2004) Ultraviolet radiation affects emission of ozone-depleting substances by marine macroalgae: results from a laboratory incubation study. *Environ. Sci. Technol.* **38**: 6605–6609.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., *et al*. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**: e1003118.

Lee, A.J. and Wallace, S.S. (2017) Hide and seek: How do DNA glycosylases locate oxidatively damaged DNA bases amidst a sea of undamaged bases? *Free Radic. Biol. Med.* **107**: 170–178.

Lee, P.K.H., Johnson, D.R., Holmes, V.F., He, J., and Alvarez-Cohen, L. (2006) Reductive dehalogenase gene expression as a biomarker for physiological activity of *Dehalococcoides* spp. *Appl. Environ. Microbiol.* **72**: 6161–6168.

Lee, P.K.H., Macbeth, T.W., Sorenson, K.S., Deeb, R.A., and Alvarez-Cohen, L. (2008) Quantifying genes and transcripts to assess the in situ physiology of *Dehalococcoides* spp. in a trichloroethene-contaminated groundwater site. *Appl. Environ. Microbiol.* **74**: 2728–2739.

Li, N., Tong, R.-L., Yao, L., Chen, Q., Yan, X., Ding, D.-R., *et al*. (2018) Identification of two glutathione-dependent 3,6-dichlorogentisate dehalogenases and their roles in the catabolism of the herbicide dicamba in *Rhizorhabdus dicambivorans* Ndbn-20. *Appl. Environ. Microbiol.* AEM.00623-18.

Li, N., Yao, L., He, Q., Qiu, J., Cheng, D., Ding, D., *et al.* (2017) 3,6-Dichlorosalicylate catabolism Is initiated by the DsmABC cytochrome P450 monooxygenase system in *Rhizorhabdus* dicambivorans Ndbn-20. *Appl. Environ. Microbiol.* **84**: e02133–17.

Liang, yi, Liu, X., Singletary, M.A., Wang, K., and Mattes, T.E. (2017) Relationships between the

abundance and expression of functional genes from vinyl chloride (VC)-degrading bacteria and geochemical parameters at VC-contaminated sites. *Environ. Sci. Technol.*

Liang, B., Jiang, J., Zhang, J., Zhao, Y., and Li, S. (2012) Horizontal transfer of dehalogenase genes involved in the catalysis of chlorinated compounds: Evidence and ecological role. *Crit Rev Microbiol* **38**: 95–110.

Liang, J., Woodward, C., and Edelsbrunner, H. (1998) Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**: 1884–1897.

Lidstrom, M.E. and Marx, C.J. (2001) Development of improved versatile broad-host-range vectors for use in methylotrophs and other Gram-negative bacteria. *Microbiology* **147**: 2065–2075.

Lillis, L., Clipson, N., and Doyle, E. (2010) Quantification of catechol dioxygenase gene expression in soil during degradation of 2,4-dichlorophenol. *FEMS Microbiol Ecol* **73**: 363–369.

Lin, S. and Cronan, J.E. (2012) The BioC O-methyltransferase catalyzes methyl esterification of malonyl-acyl carrier protein, an essential step in biotin synthesis. *J. Biol. Chem.* **287**: 37010–37020.

Lin, Y., A, D.R., Guan, S., Mamanova, L., and McDowall, K.J. (2013) A combination of improved differential and global RNA-seq reveals pervasive transcription initiation and events in all stages of the life-cycle of functional RNAs in *Propionibacterium acnes*, a major contributor to wide-spread human disease. *BMC Genomics* **14**: 620.

Liu, X., Germaine, K.J., Ryan, D., and Dowling, D.N. (2007) Development of a GFP-based biosensor for detecting the bioavailability and biodegradation of polychlorinated biphenyls (PCBs). *J. Environ. Eng. Landsc. Manag.* **15**: 261–268.

Liu, X., Germaine, K.J., Ryan, D., and Dowling, D.N. (2010) Whole-cell fluorescent biosensors for bioavailability and biodegradation of polychlorinated biphenyls. *Sensors* **10**: 1377–1398.

Lobert, J.M., Keene, W.C., Logan, J.A., and Yevich, R. (1999) Global chlorine emissions from biomass burning: reactive chlorine emissions inventory. *J. Geophys. Res. Atmospheres* **104**: 8373–8389.

Lopes, N., Hawkins, S.A., Jegier, P., Menn, F.M., Sayler, G.S., and Ripp, S. (2012) Detection of dichloromethane with a bioluminescent (*lux*) bacterial bioreporter. *J. Ind. Microbiol. Biotechnol.* **39**: 45–53.

de Lorenzo, V. (2009) Recombinant bacteria for environmental release: what went wrong and what we have learnt from it. *Clin. Microbiol. Infect.* **15**: 63–65.

de Lorenzo, V., Silva-Rocha, R., Carbajosa, G., C Galvao, T., and Cases, I. (2010) Sensing xenobiotic compounds: Lessons from bacteria that face pollutants in the environment., p. 81.

Lovley, D.R. (2003) Cleaning up with genomics: applying molecular biology to bioremediation.

Nat. Rev. Microbiol. **1**: 35–44.

Lybecker, M., Bilusic, I., and Raghavan, R. (2014) Pervasive transcription: Detecting functional RNAs in bacteria. *Transcription* **5**: e944039.

MacMillan, D.S., Murray, J., Sneddon, H.F., Jamieson, C., and Watson, A.J.B. (2012) Replacement of dichloromethane within chromatographic purification: a guide to alternative solvents. *Green Chem.* **14**: 3016.

Mägli, A., Messmer, M., and Leisinger, T. (1998) Metabolism of dichloromethane by the strict anaerobe *Dehalobacterium formicoaceticum*. *Appl. Environ. Microbiol.* **64**: 646–650.

Mao, X., Stenuit, B., Polasko, A., and Alvarez-Cohen, L. (2015) Efficient metabolic exchange and electron transfer within a syntrophic trichloroethene-degrading coculture of *Dehalococcoides mccartyi* 195 and *Syntrophomonas wolfei*. *Appl. Environ. Microbiol.* **81**: 2015–2024.

Marqués, S., Aranda-Olmedo, I., and Ramos, J.L. (2006) Controlling bacterial physiology for optimal expression of gene reporter constructs. *Curr. Opin. Biotechnol.* **17**: 50–56.

Marsch, G.A., Mundkowski, R.G., Morris, B.J., Manier, M.L., Hartman, M.K., and Guengerich, F.P. (2001) Characterization of nucleoside and DNA adducts formed by *S* -(1-Acetoxymethyl)glutathione and implications for dihalomethane−glutathione conjugates. *Chem. Res. Toxicol.* **14**: 600–608.

Marx, C.J. (2008) Development of a broad-host-range *sacB*-based vector for unmarked allelic exchange. *BMC Res. Notes* **1**: 1.

Marzorati, M., de Ferra, F., Van Raemdonck, H., Borin, S., Allifranchini, E., Carpani, G., *et al.* (2007) A novel reductive dehalogenase, identified in a contaminated groundwater enrichment culture and in *Desulfitobacterium dichloroeliminans* strain DCA1, is linked to dehalogenation of 1,2-dichloroethane. *Appl. Environ. Microbiol.* **73**: 2990–2999.

Mashiyama, S.T., Malabanan, M.M., Akiva, E., Bhosle, R., Branch, M.C., Hillerich, B., *et al.* (2014) Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere. *PLoS Biol.* **12**: e1001843.

Masulis, I.S., Babaeva, Z.S., Chernyshov, S.V., and Ozoline, O.N. (2015) Visualizing the activity of *Escherichia coli* divergent promoters and probing their dependence on superhelical density using dual-colour fluorescent reporter vector. *Sci. Rep.* **5**:.

Mattes, T.E., Jin, Y.O., Livermore, J., Pearl, M., and Liu, X. (2015) Abundance and activity of vinyl chloride (VC)-oxidizing bacteria in a dilute groundwater VC plume biostimulated with oxygen and ethene. *Appl. Microbiol. Biotechnol.* **99**: 9267–9276.

Matturro, B. and Rossetti, S. (2015) GeneCARD-FISH: Detection of *tceA* and *vcrA* reductive dehalogenase genes in *Dehalococcoides mccartyi* by fluorescence *in situ* hybridization. *J. Microbiol. Methods* **110**: 27–32.

McDermott, C. and Heffron, J.J.A. (2013) Toxicity of industrially relevant chlorinated organic solvents *in vitro*. *Int. J. Toxicol.* **32**: 136–145.

McFall, S.M., Abraham, B., Narsolis, C.G., and Chakrabarty, A.M. (1997) A tricarboxylic acid cycle intermediate regulating transcription of a chloroaromatic biodegradative pathway: fumarate-mediated repression of the *clcABD* operon. *J Bacteriol* **179**: 6729–6735.

McFall, S.M., Chugani, S.A., and Chakrabarty, A.M. (1998) Transcriptional activation of the catechol and chlorocatechol operons: variations on a theme. *Gene* **223**: 257–267.

McLLean, J.E., Hamaguchi, N., Belenky, P., Mortimer, S.E., Stanton, M., and Hedstrom, L. (2004) Inosine 5′-monophosphate dehydrogenase binds nucleic acids *in vitro* and *in vivo*. *Biochem. J.* **379**: 243–251.

Men, Y., Feil, H., Verberkmoes, N.C., Shah, M.B., Johnson, D.R., Lee, P.K.H., *et al*. (2012) Sustainable syntrophic growth of *Dehalococcoides ethenogenes* strain 195 with *Desulfovibrio vulgaris* Hildenborough and *Methanobacterium congolense*: global transcriptomic and proteomic analyses. *ISME J* **6**: 410–421.

Men, Y., Yu, K., Baelum, J., Gao, Y., Tremblay, J., Prestat, E., *et al.* (2017) Metagenomic and metatranscriptomic analyses reveal the structure and dynamics of a dechlorinating community containing *Dehalococcoides mccartyi* and corrinoid-providing microorganisms under cobalamin-limited conditions. *Appl. Environ. Microbiol.* **83**: e03508–16.

Mena-Benitez, G.L., Gandia-Herrero, F., Graham, S., Larson, T.R., McQueen-Mason, S.J., French, C.E., *et al.* (2008) Engineering a catabolic pathway in plants for the degradation of 1,2-dichloroethane. *PLANT Physiol.* **147**: 1192–1198.

Miazek, K., Kratky, L., Sulc, R., Jirout, T., Aguedo, M., Richel, A., *et al.* (2017) Effect of organic solvents on microalgae growth, metabolism and industrial bioproduct extraction: a review. *Int. J. Mol. Sci.* **18**: 1429.

Michener, J.K., Camargo Neves, A.A., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014) Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post-transfer refinement. *eLife* **3**:.

Michener, J.K., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014) Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. *J. Bacteriol.* **196**: 2101–2107.

Michener, J.K., Vuilleumier, S., Bringel, F., and Marx, C.J. (2016) Transfer of a catabolic pathway for chloromethane in *Methylobacterium* strains highlights different limitations for growth with chloromethane or with dichloromethane. *Front. Microbiol.* **7**:.

Mielecki, D. and Grzesiuk, E. (2014) Ada response – a strategy for repair of alkylated DNA in bacteria. *FEMS Microbiol. Lett.* **355**: 1–11.

Miyauchi, K., Lee, H.S., Fukuda, M., Takagi, M., and Nagata, Y. (2002) Cloning and characterization of *linR*, involved in regulation of the downstream pathway for

hexachlorocyclohexane degradation in *Sphingomonas paucimobilis* UT26. *Appl. Environ. Microbiol.* **68**: 1803–1807.

Moglich, A., Ayers, R.A., and Moffat, K. (2009) Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* **17**: 1282–1294.

Mohn, W.W., Garmendia, J., Galvao, T.C., and De Lorenzo, V. (2006) Surveying biotransformations with *à la carte* genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into *lacZ*-based phenotypes. *Environ. Microbiol.* **8**: 546–555.

Monnet, J., Grange, W., Strick, T.R., and Joly, N. (2013) Mfd as a central partner of transcription coupled repair. *Transcription* **4**: 109–113.

Mukhopadhyay, A. (2015) Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. *Trends Microbiol.* **23**: 498–508.

Muller, E.E.L. (2011) Le métabolisme du dichlorométhane chez *Methylobacterium extorquens* DM4 : génomique fonctionnelle et physiologie, thèse de l'université de Strasbourg.

Muller, E.E.L., Bringel, F., and Vuilleumier, S. (2011) Dichloromethane-degrading bacteria in the genomic age. *Res. Microbiol.* **162**: 869–876.

Muller, E.E.L., Hourcade, E., Louhichi-Jelail, Y., Hammann, P., Vuilleumier, S., and Bringel, F. (2011) Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4. *Environ. Microbiol.* **13**: 2518–2535.

Muller, J.A., Rosner, B.M., von Abendroth, G., Meshulam-Simon, G., McCarty, P.L., and Spormann, A.M. (2004) Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. strain VS and its environmental distribution. *Appl. Environ. Microbiol.* **70**: 4880–4888.

Münch, R., Hiller, K., Barg, H., Heldt, D., Linz, S., Wingender, E., and Jahn, D. (2003) PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.* **31**: 266–269.

Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., and Senda, T. (2003) Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *J. Mol. Biol.* **328**: 555–566.

Mustoe, A.M., Busan, S., Rice, G.M., Hajdin, C.E., Peterson, B.K., Ruda, V.M., *et al.* (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* **173**: 181-195.e18.

Nesvizhskii, A.I. (2014) Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* **11**: 1114–1125.

Nielsen, T.K., Rasmussen, M., Demanèche, S., Cecillon, S., Vogel, T.M., and Hansen, L.H. (2017) Evolution of *Sphingomonad* gene clusters related to pesticide catabolism revealed by genome sequence and mobilomics of *Sphingobium herbicidovorans* MH. *Genome Biol.*

Evol. **9**: 2477–2490.

Nikel, P.I. and de Lorenzo, V. (2013) Engineering an anaerobic metabolic regime in *Pseudomonas putida* KT2440 for the anoxic biodegradation of 1,3-dichloroprop-1-ene. *Metab. Eng.* **15**: 98–112.

Nikolausz, M., Kappelmeyer, U., Nijenhuis, I., Ziller, K., and Kästner, M. (2005) Molecular characterization of dichloromethane-degrading *Hyphomicrobium* strains using 16S rDNA and DCM dehalogenase gene sequences. *Syst. Appl. Microbiol.* **28**: 582–587.

Nirmalakhandan, N., Xu, S., Trevizo, C., Brennan, R., and Peace, J. (1997) Additivity in microbial toxicity of nonuniform mixtures of organic chemicals. *Ecotoxicol. Environ. Saf.* **37**: 97–102.

Norman, A., Hansen, L.H., and Sørensen, S.J. (2006) A flow cytometry-optimized assay using an SOS–green fluorescent protein (SOS–GFP) whole-cell biosensor for the detection of genotoxins in complex environments. *Mutat. Res. Toxicol. Environ. Mutagen.* **603**: 164–172.

Novotna, G.B., Kwun, M.J., and Hong, H.-J. (2016) *In Vivo* characterization of the activation and interaction of the VanR-VanS two-component regulatory system controlling glycopeptide antibiotic resistance in two related *Streptomyces* Species. *Antimicrob. Agents Chemother.* **60**: 1627–1637.

Ochsner, A.M., Sonntag, F., Buchhaupt, M., Schrader, J., and Vorholt, J.A. (2015) *Methylobacterium extorquens*: methylotrophy and biotechnological applications. *Appl. Microbiol. Biotechnol.* **99**: 517–534.

Ohtsubo, Y., Goto, H., Nagata, Y., Kudo, T., and Tsuda, M. (2006) Identification of a response regulator gene for catabolite control from a PCB-degrading *beta-proteobacteria, Acidovorax* sp. KKS102. *Mol. Microbiol.* **60**: 1563–1575.

Olaniran, A.O., Balgobind, A., and Pillay, B. (2013) Bioavailability of heavy metals in soil: impact on microbial biodegradation of organic compounds and possible improvement strategies. *Int. J. Mol. Sci.* **14**: 10197–10228.

Oliva, G., Sahr, T., and Buchrieser, C. (2015) Small RNAs, 5′ UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS Microbiol. Rev.* **39**: 331–349.

Oram, D.E., Ashfold, M.J., Laube, J.C., Gooch, L.J., Humphrey, S., Sturges, W.T., *et al.* (2017) A growing threat to the ozone layer from short-lived anthropogenic chlorocarbons. *Atmospheric Chem. Phys.* **17**: 11929–11941.

Orosz, A., Boros, I., and Venetianer, P. (1991) Analysis of the complex transcription termination region of the *Escherichia coli rrnB* gene. *Eur. J. Biochem.* **201**: 653–659.

Ottengraf, S.P.P., Meesters, J.J.P., Van Den Oever, A.H.C., and Rozema, H.R. (1986) Biological elimination of volatile xenobiotic compounds in biofilters. *Bioprocess Eng.* **1**: 61–69.

Pachkov, M., Erb, I., Molina, N., and van Nimwegen, E. (2007) SwissRegulon: a database of

genome-wide annotations of regulatory sites. *Nucleic Acids Res.* **35**: D127–D131.

Pardue, J.H., Kongara, S., and Jones, J.W. (1996) Effect of cadmium on reductive dechlorination of trichloroaniline. *Environ. Toxicol. Chem.* **15**: 1083–1088.

Parnell, J.J., Park, J., Denef, V., Tsoi, T., Hashsham, S., Quensen, J., and Tiedje, J.M. (2006) Coping with polychlorinated biphenyl (PCB) toxicity: Physiological and genome-wide responses of *Burkholderia xenovorans* LB400 to PCB-mediated stress. *Appl Env. Microbiol* **72**: 6607–6614.

Pasteur, L. (1848) Mémoire sur la relation qui peut exister entre la forme cristalline et la composition chimique, et sur la cause de la polarisation rotatoire. *Comptes Rendus Hebd. Séances Académie Sci.* **XXVII**: 535–538.

Paulin, M.M., Nicolaisen, M.H., and Sorensen, J. (2010) Abundance and expression of enantioselective *rdpA* and *sdpA* dioxygenase genes during degradation of the racemic herbicide (R,S)-2-(2,4-dichlorophenoxy)propionate in soil. *Appl. Environ. Microbiol.* **76**: 2873–2883.

Pavlova, M., Klvana, M., Prokop, Z., Chaloupkova, R., Banas, P., Otyepka, M., *et al.* (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.* **5**: 727–733.

Peng, X., Yamamoto, S., Vertès, A.A., Keresztes, G., Inatomi, K., Inui, M., and Yukawa, H. (2012) Global transcriptome analysis of the tetrachloroethene-dechlorinating bacterium *Desulfitobacterium hafniense* Y51 in the presence of various electron donors and terminal electron acceptors. *J. Ind. Microbiol. Biotechnol.* **39**: 255–268.

Perez, J.C. and Groisman, E.A. (2009) Evolution of transcriptional regulatory circuits in bacteria. *Cell* **138**: 233–244.

Perez-Pantoja, D., De la Iglesia, R., Pieper, D.H., and Gonzalez, B. (2008) Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus* necator JMP134. *FEMS Microbiol. Rev.* **32**: 736–794.

Perez-Rueda, E., Hernandez-Guerrero, R., Martinez-Nunez, M.A., Armenta-Medina, D., Sanchez, I., and Ibarra, J.A. (2018) Abundance, diversity and domain architecture variability in prokaryotic DNA-binding transcription factors. *PLoS ONE* **13**: e0195332.

Petersen, T.N., Brunak, S., Heijne, G. von, and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**: 785–786.

Pietrosanto, M., Adinolfi, M., Casula, R., Ausiello, G., Ferrè, F., and Helmer-Citterich, M. (2018) BEAM web server: a tool for structural RNA motif discovery. *Bioinformatics* **34**: 1058–1060.

Platero, A.I., Garcia-Jaramillo, M., Santero, E., and Govantes, F. (2012) Transcriptional organization and regulatory elements of a *Pseudomonas* sp. strain ADP operon encoding a

LysR-type regulator and a putative solute transport system. *J. Bacteriol.* **194**: 6560–6573.

Poelarends, G.J., Kulakov, L.A., Larkin, M.J., Vlieg, J.E.T. van H., and Janssen, D.B. (2000) Roles of horizontal gene transfer and gene integration in evolution of 1,3-dichloropropene- and 1,2-dibromoethane-degradative pathways. *J Bacteriol* **182**: 2191–2199.

Puglisi, E., Cahill, M.J., Lessard, P.A., Capri, E., Sinskey, A.J., Archer, J.A.C., and Boccazzi, P. (2010) Transcriptional response of *Rhodococcus* aetherivorans I24 to polychlorinated biphenyl-contaminated sediments. *Microb. Ecol.* **60**: 505–515.

Putnam, C.D. (2016) Evolution of the methyl directed mismatch repair system in *Escherichia coli*. *DNA Repair* **38**: 32–41.

Quintero, N., Stashenko, E.E., and Fuentes, J.L. (2012) The influence of organic solvents on estimates of genotoxicity and antigenotoxicity in the SOS chromotest. *Genet. Mol. Biol.* **35**: 503–514.

R Core Team (2017) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: URL http://www. R-project.org/.

Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* **403**: 646–649.

Rahm, B.G., Morris, R.M., and Richardson, R.E. (2006) Temporal expression of respiratory genes in an enrichment culture containing *Dehalococcoides ethenogenes*. *Appl. Environ. Microbiol.* **72**: 5486–5491.

Ravikumar, S., Baylon, M.G., Park, S.J., and Choi, J. (2017) Engineered microbial biosensors based on bacterial two-component systems as synthetic biotechnology platforms in bioremediation and biorefinery. *Microb. Cell Factories* **16**:.

Ray, S., Gunzburg, M.J., Wilce, M., Panjikar, S., and Anand, R. (2016) Structural basis of selective aromatic pollutant sensing by the effector binding domain of MopR, an NtrC family transcriptional regulator. *ACS Chem. Biol.* **11**: 2357–2365.

Regnault, V. (1839) De l'action du chlore sur les éthers hydrochloriques de l'alcool et de l'esprit de bois, et de plusieurs points de la théorie des éthers. *Ann. Chim. Phys.* **79**: 353–430.

Ren, G.-X., Guo, X.-P., and Sun, Y.-C. (2017) Regulatory 3′ untranslated regions of bacterial mRNAs. *Front. Microbiol.* **8**:.

Richardson, E.J. and Watson, M. (2013) The automatic annotation of bacterial genomes. *Brief. Bioinform.* **14**: 1–12.

Ritz, C. and Spiess, A.-N. (2008) qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics* **24**: 1549–1551.

Rivera-Gomez, N., Segovia, L., and Perez-Rueda, E. (2011) Diversity and distribution of

transcription factors: their partner domains play an important role in regulatory plasticity in bacteria. *Microbiology* **157**: 2308–2318.

Roberts, M.F. (2006) Inositol in bacteria and archaea. In, Majumder,A.L. and Biswas,B.B. (eds), *Biology of inositols and phosphoinositides*. Springer US, pp. 103–133.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011) Integrative Genomics Viewer. *Nat. Biotechnol.* **29**: 24–26.

Roche, S.D.L. and Leisinger, T. (1990) Sequence analysis and expression of the bacterial dichloromethane dehalogenase structural gene, a member of the glutathione S-transferase supergene family. *J. Bacteriol.* **172**: 164–171.

Roselli, S., Nadalig, T., Vuilleumier, S., and Bringel, F. (2013) The 380 kb pCMU01 plasmid encodes chloromethane utilization genes and redundant genes for vitamin B12- and tetrahydrofolate-dependent chloromethane metabolism in *Methylobacterium extorquens* CM4: A proteomic and bioinformatics study. *PLoS ONE* **8**: e56598.

Ruhal, R., Kataria, R., and Choudhury, B. (2013) Trends in bacterial trehalose metabolism and significant nodes of metabolic pathway in the direction of trehalose accumulation: Trehalose metabolism in bacteria. *Microb. Biotechnol.* **6**: 493–502.

Rupakula, A., Kruse, T., Boeren, S., Holliger, C., Smidt, H., and Maillard, J. (2013) The restricted metabolism of the obligate organohalide respiring bacterium *Dehalobacter* restrictus: lessons from tiered functional genomics. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120325–20120325.

Sallabhan, R., Kerdwong, J., Dubbs, J.M., Somsongkul, K., Whangsuk, W., Piewtongon, P., *et al*. (2013) The *hdhA* gene encodes a haloacid dehalogenase that is regulated by the LysR-type regulator, HdhR, in *Sinorhizobium meliloti*. *Mol. Biotechnol.* **54**: 148–157.

Sambrook, J. (2001) Molecular Cloning: A Laboratory Manual, Third Edition 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Samoiloff, M.R., Schulz, S., Jordan, Y., Denich, K., and Arnott, E. (1980) A rapid simple long-term toxicity assay for aquatic contaminants using the nematode *Panagrellus redivivus*. *Can. J. Fish. Aquat. Sci.* **37**: 1167–1174.

San Millan, A., Toll-Riera, M., Qi, Q., and MacLean, R.C. (2015) Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nat Commun* **6**: 6845.

Sanchez, M.A. and Gonzalez, B. (2007) Genetic characterization of 2,4,6-trichlorophenol degradation in *Cupriavidus necator* JMP134. *Appl. Environ. Microbiol.* **73**: 2769–2776.

Saraogi, I. and Shan, S. (2014) Co-translational protein targeting to the bacterial membrane. *Biochim. Biophys. Acta* **1843**: 1433–1441.

Schada von Borzyskowski, L., Remus-Emsermann, M., Weishaupt, R., Vorholt, J.A., and Erb, T.J. (2015) A set of versatile brick vectors and promoters for the assembly, expression, and

integration of synthetic operons in *Methylobacterium extorquens* AM1 and other *Alphaproteobacteria*. *ACS Synth. Biol.* **4**: 430–443.

Scheublin, T.R., Deusch, S., Moreno-Forero, S.K., Muller, J.A., van der Meer, J.R., and Leveau, J.H.J. (2014) Transcriptional profiling of Gram-positive *Arthrobacter* in the phyllosphere: induction of pollutant degradation genes by natural plant phenolic compounds. *Env. Microbiol* **16**: 2212–2225.

Schirmer, F., Ehrt, S., and Hillen, W. (1997) Expression, inducer spectrum, domain structure, and function of MopR, the regulator of phenol degradation in *Acinetobacter calcoaceticus* NCIB8250. *J. Bacteriol.* **179**: 1329–1336.

Schirmer, T. (2016) C-di-GMP synthesis: Structural aspects of evolution, catalysis and regulation. *J. Mol. Biol.* **428**: 3683–3701.

Schlosser, P.M., Bale, A.S., Gibbons, C.F., Wilkins, A., and Cooper, G.S. (2015) Human health effects of dichloromethane: Key findings and scientific issues. *Environ. Health Perspect.* **123**: 114–119.

Schlüter, J.-P., Reinkensmeier, J., Barnett, M.J., Lang, C., Krol, E., Giegerich, R., *et al*. (2013) Global mapping of transcription start sites and promoter motifs in the symbiotic α-*proteobacterium Sinorhizobium meliloti* 1021. *BMC Genomics* **14**: 156.

Schmid, J., Sieber, V., and Rehm, B. (2015) Bacterial exopolysaccharides: Biosynthesis pathways and engineering strategies. *Front. Microbiol.* **6**:.

Schmid-Appert, M. (1996) Untersuchungen zur regulation des dichlormethan-dehalogenase gens aus *Methylobacterium* sp. stamm DM4 und struktur des angrenzenden DNA-region, thèse de l'École polytechnique fédérale de Zurich.

Schmid-Appert, M., Zoller, K., Traber, H., Vuilleumier, S., and Leisinger, T. (1997) Association of newly discovered IS elements with the dichloromethane utilization genes of methylotrophic bacteria. *Microbiol. Read. Engl.* **143 (Pt 8)**: 2557–2567.

Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**: 1101–1108.

Scholtz, R., Wackett, L.P., Egli, C., Cook, A.M., and Leisinger, T. (1988) Dichloromethane dehalogenase with improved catalytic activity isolated from a fast-growing dichloromethane-utilizing bacterium. *J. Bacteriol.* **170**: 5698–5704.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012) PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* **113**: 1014–1026.

Schwandner, F.M., Seward, T.M., Giże, A.P., Hall, K., and Dietrich, V.J. (2013) Halocarbons and other trace heteroatomic organic compounds in volcanic gases from Vulcano (Aeolian Islands, Italy). *Geochim. Cosmochim. Acta* **101**: 191–221.

Segura, A., Molina, L., Fillet, S., Krell, T., Bernal, P., Muñoz-Rojas, J., and Ramos, J.-L. (2012)

Solvent tolerance in Gram-negative bacteria. *Curr. Opin. Biotechnol.* **23**: 415–421.

Seibold, G.M. and Eikmanns, B.J. (2007) The *glgX* gene product of *Corynebacterium glutamicum* is required for glycogen degradation and for fast adaptation to hyperosmotic stress. *Microbiology* **153**: 2212–2220.

Selifonova, O.V. and Eaton, R.W. (1996) Use of an *ipb-lux* fusion to study regulation of the isopropylbenzene catabolism operon of *Pseudomonas putida* RE204 and to detect hydrophobic pollutants in the environment. *Appl. Environ. Microbiol.* **62**: 778–783.

Sentchilo, V., Ravatn, R., Werlen, C., Zehnder, A.J.B., and van der Meer, J.R. (2003) Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. Strain B13. *J. Bacteriol.* **185**: 4530–4538.

Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**: 776–790.

Seyedabbasi, M.A., Newell, C.J., Adamson, D.T., and Sale, T.C. (2012) Relative contribution of DNAPL dissolution and matrix diffusion to the long-term persistence of chlorinated solvent source zones. *J. Contam. Hydrol.* **134**–**135**: 69–81.

Sharma, A.K., Rigby, A.C., and Alper, S.L. (2011) STAS domain structure and function. *Cell. Physiol. Biochem.* **28**: 407–422.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., *et al*. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250–255.

Shee, C., Cox, B.D., Gu, F., Luengas, E.M., Joshi, M.C., Chiu, L.Y., *et al*. (2013) Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. *eLife* **2**:.

Shestakova, M. and Sillanpää, M. (2013) Removal of dichloromethane from ground and wastewater: A review. *Chemosphere* **93**: 1258–1267.

Shrivastav, N., Li, D., and Essigmann, J.M. (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis* **31**: 59–70.

Siedenburg, G. and Jendrossek, D. (2011) Squalene-hopene cyclases. *Appl Env. Microbiol* **77**: 3905–3915.

Sipilä, T.P., Väisänen, P., Paulin, L., and Yrjälä, K. (2010) *Sphingobium* sp. HV3 degrades both herbicides and polyaromatic hydrocarbons using ortho- and meta-pathways with differential expression shown by RT-PCR. *Biodegradation* **21**: 771–784.

Škulj, M., Okršlar, V., Jalen, Š., Jevševar, S., Slanc, P., Štrukelj, B., and Menart, V. (2008) Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. *Microb. Cell Factories* **7**: 6.

Smidt, H., van Leest, M., van der Oost, J., and de Vos, W.M. (2000) Transcriptional regulation of the *cpr* gene cluster in ortho-chlorophenol-respiring *Desulfitobacterium dehalogenans*. *J Bacteriol* **182**: 5683–5691.

Sneath, P.H.A. (1989) Analysis and interpretation of sequence data for bacterial systematics: The view of a numerical taxonomist. *Syst. Appl. Microbiol.* **12**: 15–31.

Sowa, S.W., Gelderman, G., Leistra, A.N., Buvanendiran, A., Lipp, S., Pitaktong, A., *et al.* (2017) Integrative FourD omics approach profiles the target network of the carbon storage regulatory system. *Nucleic Acids Res.* **45**: 1673–1686.

Speed, M.C., Burkhart, B.W., Picking, J.W., and Santangelo, T.J. (2018) An archaeal fluoride-responsive riboswitch provides an inducible expression system for hyperthermophiles. *Appl Env. Microbiol* **84**: e02306-17.

Spiess, A.-N., Feig, C., and Ritz, C. (2008) Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. *BMC Bioinformatics* **9**: 221.

Sporer, A.J., Kahl, L.J., Price-Whelan, A., and Dietrich, L.E.P. (2017) Redox-based regulation of bacterial development and behavior. *Annu. Rev. Biochem.* **86**: 777–797.

Springer, A.L., Morris, C.J., and Lidstrom, M.E. (1997) Molecular analysis of *mxbD* and *mxbM*, a putative sensor-regulator pair required for oxidation of methanol in *Methylobacterium extorquens* AM1. *Microbiology* **143**: 1737–1744.

Starr, L.M., Fruci, M., and Poole, K. (2012) Pentachlorophenol induction of the *Pseudomonas aeruginosa mexAB-oprM* efflux operon: involvement of repressors NalC and MexR and the antirepressor ArmR. *PLoS ONE* **7**: e32684.

Stierand, K. and Rarey, M. (2010) Drawing the PDB: Protein−ligand complexes in two dimensions. *ACS Med. Chem. Lett.* **1**: 540–545.

Stourman, N.V., Rose, J.H., Vuilleumier, S., and Armstrong, R.N. (2003) Catalytic mechanism of dichloromethane dehalogenase from *Methylophilus* sp. Strain DM11. *Biochemistry* **42**: 11048–11056.

Su, X. and Tsang, J.S.H. (2013) Existence of a robust haloacid transport system in a *Burkholderia* species bacterium. *Biochim. Biophys. Acta BBA - Biomembr.* **1828**: 187–192.

Tabata, M., Ohhata, S., Nikawadori, Y., Kishida, K., Sato, T., Kawasumi, T., *et al.* (2016) Comparison of the complete genome sequences of four γ-hexachlorocyclohexane-degrading bacterial strains: insights into the evolution of bacteria able to degrade a recalcitrant man-made pesticide. *DNA Res* **23**: 581–599.

Takeda, H., Shimodaira, J., Yukawa, K., Hara, N., Kasai, D., Miyauchi, K., *et al.* (2010) Dual two-component regulatory systems are involved in aromatic compound degradation in a polychlorinated-biphenyl degrader, *Rhodococcus* jostii RHA1. *J. Bacteriol.* **192**: 4741–4751.

Tas, N., van Eekert, M.H.A., Wagner, A., Schraa, G., de Vos, W.M., and Smidt, H. (2011) Role of

"*Dehalococcoides*" spp. in the aaerobic transformation of hexachlorobenzene in European rivers. *Appl. Environ. Microbiol.* **77**: 4437–4445.

The Merck index: An encyclopedia of chemicals, drugs and biologicals: Edition 15 (2013).

Tittlemier, S.A., Blank, D.H., Gribble, G.W., and Norstrom, R.J. (2002) Structure elucidation of four possible biogenic organohalogens using isotope exchange mass spectrometry. *Chemosphere* **46**: 511–517.

Tobajas, M., Verdugo, V., Polo, A.M., Rodriguez, J.J., and Mohedano, A.F. (2016) Assessment of toxicity and biodegradability on activated sludge of priority and emerging pollutants. *Environ. Technol.* **37**: 713–721.

Tolba, M.E.M., Kobayashi, S., Imada, M., Suzuki, Y., and Sugano, S. (2013) *Giardia lamblia* transcriptome analysis using TSS-Seq and RNA-Seq. *PloS One* **8**: e76184.

Torgonskaya, M.L., Doronina, N.V., Hourcade, E., Trotsenko, Y.A., and Vuilleumier, S. (2011) Chloride-associated adaptive response in aerobic methylotrophic dichloromethane-utilising bacteria. *J. Basic Microbiol.* **51**: 296–303.

Torgonskaya, M.L., Zyakun, A.M., Trotsenko, Y.A., Laurinavichius, K.S., Kümmel, S., Vuilleumier, S., and Richnow, H.H. (2018) Individual stages of bacterial dichloromethane degradation mapped by stable carbon and chlorine isotope analysis. *J. Environ. Sci.*

Torii, H., Machida, A., Hara, H., Hatta, T., and Takizawa, N. (2013) The regulatory mechanism of 2,4,6-trichlorophenol catabolic operon expression by HadR in *Ralstonia pickettii* DTP0602. *Microbiology* **159**: 665–677.

Trefault, N., Guzmán, L., Pérez, H., Godoy, M., González, B., and others (2009) Involvement of several transcriptional regulators in the differential expression of *tfd* genes in *Cupriavidus necator* JMP134. *Int. Microbiol.* **12**: 97.

Trudinger, C.M., Etheridge, D.M., Sturrock, G.A., Fraser, P.J., Krummel, P.B., and McCulloch, A. (2004) Atmospheric histories of halocarbons from analysis of Antarctic firn air: Methyl bromide, methyl chloride, chloroform, and dichloromethane. *J. Geophys. Res. Atmospheres* **109**:.

Tsai, W.-T. (2017) Fate of chloromethanes in the atmospheric environment: Implications for human health, ozone formation and depletion, and global warming impacts. *Toxics* **5**:.

T'Syen, J., Tassoni, R., Hansen, L., Sorensen, S.J., Leroy, B., Sekhar, A., *et al.* (2015) Identification of the amidase BbdA that initiates biodegradation of the groundwater micropollutant 2,6-dichlorobenzamide (BAM) in *Aminobacter* sp. MSH1. *Environ. Sci. Technol.* **49**: 11703–11713.

Türkowsky, D., Esken, J., Goris, T., Schubert, T., Diekert, G., Jehmlich, N., and von Bergen, M. (2018) A retentive memory of tetrachloroethene respiration in *Sulfurospirillum halorespirans* - involved proteins and a possible link to acetylation of a two-component regulatory system. *J. Proteomics*.

Türkowsky, D., Lohmann, P., Mühlenbrink, M., Schubert, T., Adrian, L., Goris, T., *et al.* (2018) Thermal proteome profiling allows quantitative assessment of interactions between tetrachloroethene reductive dehalogenase and trichloroethene. *J. Proteomics*.

Uchida, H., Mikami, B., Yamane-Tanabe, A., Ito, A., Hirano, K., and Oki, M. (2018) Crystal structure of an aldehyde oxidase from *Methylobacillus* sp. KY4400. *J. Biochem. (Tokyo)* **163**: 321–328.

Ul Haque, M.F., Nadalig, T., Bringel, F., Schaller, H., and Vuilleumier, S. (2013) A fluorescence-based bacterial bioreporter for the specific detection of methyl halide emissions in the environment. *Appl. Environ. Microbiol.* AEM.01738–13.

Van Dien, S.J. (2003) Reconstruction of C3 and C4 metabolism in *Methylobacterium extorquens AM1* using transposon mutagenesis. *Microbiology* **149**: 601–609.

Vannelli, T., Logan, M., Arciero, D.M., and Hooper, A.B. (1990) Degradation of halogenated aliphatic compounds by the ammonia-oxidizing bacterium *Nitrosomonas europaea*. *Appl. Environ. Microbiol.* **56**: 1169–1171.

Villacieros, M., Whelan, C., Mackova, M., Molgaard, J., Sanchez-Contreras, M., Lloret, J., *et al.* (2005) Polychlorinated biphenyl rhizoremediation by *Pseudomonas fluorescens* F113 derivatives, using a *Sinorhizobium meliloti nod* system to drive *bph* gene expression. *Appl. Environ. Microbiol.* **71**: 2687–2694.

Vuilleumier, S., Chistoserdova, L., Lee, M.C., Bringel, F., Lajus, A., Zhou, Y., *et al.* (2009) *Methylobacterium* genome sequences: a reference blueprint to investigate microbial metabolism of C1 compounds from natural and industrial sources. *PLoS ONE* **4**: e5584.

Vuilleumier, S., Ivoš, N., Dean, M., and Leisinger, T. (2001) Sequence variation in dichloromethane dehalogenases/glutathione S-transferases. *Microbiology* **147**: 611–619.

Vuilleumier, S. and Leisinger, T. (1996) Protein engineering studies of dichloromethane dehalogenase/glutathione S-transferase from *Methylophilus* sp. strain DM11. Ser12 but not Tyr6 is required for enzyme activity. *Eur. J. Biochem.* **239**: 410–417.

Wagner, A., Segler, L., Kleinsteuber, S., Sawers, G., Smidt, H., and Lechner, U. (2013) Regulation of reductive dehalogenase gene transcription in *Dehalococcoides mccartyi*. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120317–20120317.

Waller, A.S., Hug, L.A., Mo, K., Radford, D.R., Maxwell, K.L., and Edwards, E.A. (2012) Transcriptional analysis of a *Dehalococcoides*-containing microbial consortium reveals prophage activation. *Appl. Environ. Microbiol.* **78**: 1178–1186.

Wang, C.-H., Zhang, C., and Xing, X.-H. (2016) Xanthine dehydrogenase: An old enzyme with new knowledge and prospects. *Bioengineered* **7**: 395–405.

Wang, H., Hu, J., Xu, K., Tang, X., Xu, X., and Shen, C. (2018) Biodegradation and chemotaxis of polychlorinated biphenyls, biphenyls, and their metabolites by *Rhodococcus* spp. *Biodegradation* **29**: 1–10.

Wang, M.S. and Reed, S.M. (2012) Direct visualization of electrophoretic mobility shift assays using nanoparticle-aptamer conjugates. *Electrophoresis* **33**: 348–351.

Wang, X., Teng, Y., Tu, C., Luo, Y., Greening, C., Zhang, N., *et al.* (2018) Coupling between nitrogen fixation and tetrachlorobiphenyl dechlorination in a rhizobium–legume symbiosis. *Environ. Sci. Technol.* **52**: 2217–2224.

Ward, W.O., Swartz, C.D., Porwollik, S., Warren, S.H., Hanley, N.M., Knapp, G.W., *et al.* (2007) Toxicogenomic analysis incorporating operon-transcriptional coupling and toxicant concentration-expression response: analysis of MX-treated *Salmonella*. *BMC Bioinformatics* **8**: 378.

Wei, Y., Wang, J., and Xia, X. (2016) Coevolution between stop codon usage and release factors in bacterial species. *Mol. Biol. Evol.* **33**: 2357–2367.

Weir, K.M., Sutherland, T.D., Horne, I., Russell, R.J., and Oakeshott, J.G. (2006) A single monooxygenase, Ese, is involved in the metabolism of the organochlorides endosulfan and endosulfate in an *Arthrobacter* sp. *Appl. Environ. Microbiol.* **72**: 3524–3530.

Weng, X. and Xiao, J. (2014) Spatial organization of transcription in bacterial cells. *Trends Genet. TIG* **30**: 287–297.

Werner, J.J., Ptak, A.C., Rahm, B.G., Zhang, S., and Richardson, R.E. (2009) Absolute quantification of *Dehalococcoides* proteins: enzyme bioindicators of chlorinated ethene dehalorespiration. *Environ. Microbiol.* **11**: 2687–2697.

Whangsuk, W., Dubbs, J.M., Sallabhan, R., Somsongkul, K., Mongkolsuk, S., and Loprasert, S. (2010) ChpR Is a chlorpyrifos-responsive transcription regulator in *Sinorhizobium meliloti*. *J. Mol. Microbiol. Biotechnol.* **18**: 141–147.

White, P.M., Potter, T.L., and Culbreath, A.K. (2010) Fungicide dissipation and impact on metolachlor aerobic soil degradation and soil microbial dynamics. *Sci. Total Environ.* **408**: 1393–1402.

van den Wijngaard, A.J., van der Kamp, K.W., van der Ploeg, J., Pries, F., Kazemier, B., and Janssen, D.B. (1992) Degradation of 1,2-dichloroethane by *Ancylobacter aquaticus* and other facultative methylotrophs. *Appl. Environ. Microbiol.* **58**: 976–983.

van den Wijngaard, A.J., Reuvekamp, P.T., and Janssen, D.B. (1991) Purification and characterization of haloalcohol dehalogenase from *Arthrobacter* sp. strain AD2. *J. Bacteriol.* **173**: 124–129.

Wise, A.A. and Kuske, C.R. (2000) Generation of novel bacterial regulatory proteins that detect

priority pollutant phenols. *Appl. Environ. Microbiol.* **66**: 163–169.

Wöhrnschimmel, H., Scheringer, M., Bogdal, C., Hung, H., Salamova, A., Venier, M., *et al.* (2016) Ten years after entry into force of the Stockholm convention: What do air monitoring data tell about its effectiveness? *Environ. Pollut. Barking Essex 1987* **217**: 149–158.

Wu, S.J., Hu, Z.H., Zhang, L.L., Yu, X., and Chen, J.M. (2009) A novel dichloromethane-degrading Lysinibacillus sphaericus strain wh22 and its degradative plasmid. *Appl. Microbiol. Biotechnol.* **82**: 731–740.

Wu, S.J., Zhang, L.L., Wang, J.D., and Chen, J.M. (2007) *Bacillus circulans* WZ-12—a newly discovered aerobic dichloromethane-degrading methylotrophic bacterium. *Appl. Microbiol. Biotechnol.* **76**: 1289–1296.

Xiu, Z., Gregory, K.B., Lowry, G.V., and Alvarez, P.J.J. (2010) Effect of bare and coated nanoscale zerovalent iron on *tceA* and *vcrA* gene expression in *Dehalococcoides* spp. *Environ. Sci. Technol.* **44**: 7647–7651.

Xu, H.H., Viebahn, M., and Hanson, R.S. (1993) Identification of methanol-regulated promoter sequences from the facultative methylotrophic bacterium *Methylobacterium organophilum* XX. *J. Gen. Microbiol.* **139**: 743–752.

Yadid, I., Rudolph, J., Hlouchova, K., and Copley, S.D. (2013) Sequestration of a highly reactive intermediate in an evolving pathway for degradation of pentachlorophenol. *Proc. Natl. Acad. Sci.* **110**: E2182–E2190.

Yalkowsky, S.H., He, Y., and Jain, P. (2010) Handbook of aqueous solubility data 2 edition. CRC Press, Boca Raton, FL.

Yan, Q., Philmus, B., Chang, J.H., and Loper, J.E. (2017) Novel mechanism of metabolic co-regulation coordinates the biosynthesis of secondary metabolites in *Pseudomonas protegens*. *eLife* **6**: e22835.

Yoshida, S., Ogawa, N., Fujii, T., and Tsushima, S. (2009) Enhanced biofilm formation and 3-chlorobenzoate degrading activity by the bacterial consortium of *Burkholderia* sp. NK8 and *Pseudomonas aeruginosa* PAO1. *J. Appl. Microbiol.* **106**: 790–800.

Articles et communications scientifiques

Articles dans des journaux à comité de relecture

Bibi-Triki, S., Husson, G., **Maucourt, B.**, Vuilleumier, S., Carapito, C., and Bringel, F. (2018) N-terminome and proteogenomic analysis of the *Methylobacterium extorquens* DM4 reference strain for dichloromethane utilization. *Journal of Proteomics* **179**: 131–139.

Bringel, F., Postema, C.P., Mangenot, S., Bibi-Triki, S., Chaignaud, P., Haque, M.F.U., Gruffaz, C., Hermon, L., Louhichi, Y., **Maucourt, B.**, Muller, E.E.L., Nadalig, T., Lajus, A., Rouy, Z., Médigue, C., Barbe, V., Janssen, D.B., and Vuilleumier, S. (2017) Genome sequence of the dichloromethane-degrading bacterium *Hyphomicrobium* sp. strain GJ21. *Genome Announc.* **5**: e00622-17.

Chaignaud, P., **Maucourt, B.**, Weiman, M., Alberti, A., Kolb, S., Cruveiller, S., Vuilleumier, S., and Bringel, F. (2017) Genomic and transcriptomic analysis of growth-supporting dehalogenation of chlorinated methanes in *Methylobacterium*. *Front. Microbiol.* **8**:.

Imfeld, G., Beseaury, L., **Maucourt, B.**, Donadello, S., Baran, N., and Vuilleumier, S. Towards integrative bacterial monitoring of metolachlor toxicity in groundwater. *Front. Microbiol.* **9**:.
(Travail de stage dans le même laboratoire ; indépendant de la thèse)

Articles en cours d'écriture

Maucourt, B., Vuilleumier, S., and Bringel, F., Dehalogenation and growth with toxic organochlorine pollutants: a review on bacterial transcriptional responses
Article utilisé en tant que chapitre n°1

Maucourt, B., Bibi-Triki, S., Husson, G., Louhichi, Y., Vuilleumier, S., Carapito, C., and Bringel, F. Transcriptional regulation by DcmR of genes involved in dichloromethane metabolism
Article utilisé en tant que chapitre n°4

Communications orales et posters

Maucourt, B., Bibi-Triki, S., Husson, G., Chaignaud, P., Louhichi, Y., Vuilleumier, S., Carapito, C., and Bringel, F. (2016) Transcriptionnal and translational regulation of the *dcm* islet involved in bacterial degradation of dichloromethane, a major pollutant
Présentation orale au Séminaire de Microbiologie de Strasbourg, le 3 mars 2016
(Prix de la meilleure présentation)

Maucourt, B., Bibi-Triki, S., Husson, G., Chaignaud, P., Louhichi, Y., Boyon, A., Vuilleumier, S., Carapito, C., and Bringel, F. (2018) Bacterial growth with the pollutant dichloromethane: How does the transcriptional factor DcmR operate?
Poster présenté au Séminaire de Microbiologie de Strasbourg, le 29 mars 2018
(Prix du meilleur poster)
Poster présenté au European C1 meeting à Nijmegen (Pays-Bas), le 17 et 18 mai 2018

Vuilleumier, S., Besaury, L., Baran, N., Gruffaz, C., Donadello, S., Maalej, M., **Maucourt, B**., and Imfeld, G. (2015) Towards integrative monitoring of metolachlor contamination, biodegradation and toxicity in groundwater aquifers
Poster présenté à Applied & Environmental Microbiology Gordon Research Conference à South Hadley (États-Unis), le 11 et 12 juillet 2015

Bibi-Triki, S., **Maucourt, B.**, Husson, G., Vuilleumier, S., Carapito, C., and Bringel, F. (2016) Regulation of the dcm islet involved in degradation of dichloromethane by *Methylobacterium extorquens* DM4.
Poster et présentation orale au Gordon Research Seminar on Molecular basis of microbial one-carbon à Waterville Valley (États-Unis), du 31 juillet au 05 août 2016

Bibi-Triki, S., **Maucourt, B.**, Husson, G., Chaignaud, P., Louhichi, Y., Masquida, B., Vuilleumier, S., Carapito, C., and Bringel, F. (2017) Regulation of dichloromethane dehalogenation: a role for *dcmB* and *dcmC* alongside transcriptional regulator *dcmR*?
Présentation orale Séminaire de Microbiologie de Strasbourg, le 01 mars 2017

Marguier, A., Bardon, J., Delmée, M., Hajjar-Garreau, S., Kunemann, P., **Maucourt, B.**, Mertz, G.,

Roucoules, V., Vuilleumier, S., Ruch, D., and Ploux, L. (2017) Prevention of bacterial colonization by superhydrophobic coating in liquid surrounding: bacterial behaviour and the associated gene expression Présentation orale au XXVI International Materials Research Congress (IMRC 2017), à Cancun (Mexique), le 22 août 2017

Le nom de la personne ayant présentée la communication est souligné.

Activités annexes à la thèse

Enseignement

Chargé d'enseignement dans le cadre d'un contrat doctoral 64 heures par an.

- Année 2016/2017 (TD/TP licence 1 : UE Histologie et Cytologie)
- Année 2017/2018 (TD/TP licence 1 : UE Histologie et Cytologie)

Engagement à l'école doctorale

Représentant des doctorants à l'école doctorale des Sciences de la Vie et de la Santé ED414

Février 2016 - janvier 2019

- Membre du conseil pédagogique et du conseil scientifique
- Observateur au concours d'attribution des contrats doctoraux

Encadrement de stagiaires

Alicia Boyon

- Stage de 1er année BTS Biotechnologies (mai 2017 – juillet 2017)
- Stage de 2nd année BTS Biotechnologies (janvier 2018 – mars 2018)

Emmanuelle Chevalier

- Stage de 1er année BTS Biotechnologies (mai 2018 – juillet 2018)

Engagement associatif

Association des Doctorants et Docteurs d'Alsace ADDAL (octobre 2015 - mars 2017)

- Membre du bureau (commission poursuite de carrière)

Pint of Science 2018 (septembre 2017 - mai 2018)

- Membre de l'équipe d'organisation (Planète Terre) à Strasbourg
- Présentation le 16 mai 2018 « Bactéries et plantes : amies ou ennemies ? »

Université de Strasbourg

Bruno MAUCOURT

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Régulation *in labo* et *in situ* de l'expression de génomes de souches bactériennes méthylotrophes dégradant le dichlorométhane

Le dichlorométhane (DCM ; CH_2Cl_2) est un polluant chloré toxique émis dans l'environnement principalement par les activités industrielles. Ce polluant peut être dégradé par des bactéries méthylotrophes qui utilisent des composés en C_1 réduits comme seule source de carbone et d'énergie. La protéobactérie *Methylorubrum extorquens* DM4 porte quatre gènes *dcm* au sein du transposon catabolique *dcm* très conservé chez les bactéries dégradant le DCM. Le gène *dcmA* code la DCM déshalogénase de la famille des glutathion-S transférases essentielle à la croissance avec le DCM. Son activité est modulée par DcmR, un facteur de transcription qui régule l'expression de *dcmA* ainsi que son propre gène, orienté de manière divergente par rapport à *dcmA*. DcmR porte un domaine hélice-tour-hélice de fixation à l'ADN et un second domaine appelé *methanogen / methylotroph, DcmR sensory* (MEDS) potentiellement impliqué dans la fixation d'un composé hydrocarboné ligand. Les objectifs de ma thèse étaient de répondre à plusieurs questions : i) quel est le niveau d'expression des transcrits *dcm* et des protéines correspondantes par rapport à d'autres gènes et protéines dont l'abondance est modifiée en réponse à la croissance sur le DCM ? ii) Comment le facteur de transcription DcmR intervient-il dans la régulation des gènes *dcm* ? iii) Quelle est la variabilité du gène *dcmR* et de son environnement génétique *in situ* ?
Des méthodes globales « -omiques » de transcriptomique et de protéomique ont permis d'inventorier les ARN et les protéines dont l'abondance varie chez la souche sauvage DM4 cultivée soit avec le DCM ou le méthanol, le substrat de référence de la méthylotrophie. Les gènes *dcm* sont parmi les plus exprimés en présence de DCM, ce qui confirme leur régulation en présence de DCM. Deux approches complémentaires ciblant la détermination des sites d'initiation de la transcription (TSS-Seq) et de la traduction (N-terminome) ont permis la recherche de motifs de régulation dans les régions 5'UTR (5'-*untranslated terminal region*) et les promoteurs de gènes régulés en réponse à la croissance avec le DCM. Le rôle régulateur de dcmR a été étudié en comparant les phénotypes de croissance, l'activité promotrice par fusion transcriptionnelle, la quantification des ARN et des protéines des gènes *dcm* chez la souche sauvage par rapport à des mutants du gène *dcmR* seul ou combiné à d'autres gènes *dcm* mutés. Ces travaux ont permis de confirmer qu'en absence de DCM, DcmR inhibe la transcription de son propre gène ainsi que celle de *dcmA*. Outre DcmR, la répression nécessite aussi l'expression d'au moins un des autres gènes *dcm* et ceci par un mécanisme indépendant des boîtes de 12 pb conservées dans les promoteurs des gènes *dcmR* et *dcmA*. Lors de la croissance en condition DCM, l'absence du gène *dcmR* confère une vitesse de croissance ralentie, qui ne résulte pas d'une différence de production des ARN et des protéines codées par le transposon *dcm*. Pour que l'expression du gène *dcmA* soit activée, l'ensemble de la région intergénique entre *dcmR* et *dcmA* doit être présente, ce qui suggère la présence de sites de régulation pour la fixation d'un facteur de transcription indépendant de DcmR. L'ensemble de ces résultats a permis de proposer un nouveau modèle de régulation des gènes *dcm*. Alors que le gène *dcmR* a été détecté en quantité similaire à celle du gène *dcmA* par qPCR dans des échantillons de sites contaminés par le DCM, une analyse bioinformatique à partir des données de séquences indique que des gènes *dcmR-like* sont trouvés dans d'autres contextes génétiques que celui du transposon catabolique *dcm*. Ainsi, DcmR pourrait exercer un rôle de régulateur dans d'autres contextes ouvrant de nouvelles pistes pour l'identification des ligands du domaine MEDS.
Mots clés : Déshalogénase ; Dichlorométhane ; Facteur de transcription ; Méthylotrophie ; *Methylorubrum extorquens* DM4 ; Solvant chloré

Regulation of *in labo* and *in situ* genome expression of dichloromethane-degrading methylotrophic bacteria

Dichloromethane (DCM; CH_2Cl_2) is a toxic chlorinated pollutant mainly emitted in the environment through industrial activities. Some methylotrophic bacteria that utilize reduced one carbon compounds as sole source of carbon and energy can degrade DCM. The four *dcm* genes found in the catabolic *dcm* transposon are highly conserved among DCM-degrading bacteria, including in the *Proteobacterium Methylorubrum extorquens* DM4. The *dcmA* gene encodes a DCM dehalogenase of the glutathione-S transferase family which is essential for growth with DCM. The transcriptional factor DcmR regulates the expression of *dcmA* and its own gene, two genes which are divergently oriented. DcmR consists of a helix-turn-helix domain for DNA fixation and of a domain, called MEDS for 'methanogen / methylotroph, DcmR sensory', proposed to bind a hydrocarbon ligand. The aim of my PhD was to address the following questions: i) What is the expression level of *dcm* transcripts and their corresponding proteins compared to other genes and proteins whose abundance is modulated in response to growth with DCM? ii) How does DcmR regulate the expression of the *dcm* genes? iii) How variable is *dcmR* and its genetic environment *in situ*?
Global transcriptomic and proteomic approaches were followed to inventory transcripts and proteins whose abundance varied in the wild-type DM4 strain grown with DCM compared to with methanol, the reference substrate for methylotrophy. The *dcm* genes were among the most expressed in cultures grown with DCM, confirming their regulation in response to DCM. The identification of transcription start sites (TSS-Seq) and translation starts (N-terminome) allowed the subsequent search for regulation motifs in the promoter and 5'-untranslated terminal regions (5'UTR) of regulated genes in response to DCM utilization. The role of *dcmR* was studied using growth phenotypes, promoter activities in transcription fusion assays, *dcm* transcript and corresponding protein product quantification, comparing the wild-type strain with strains mutated only in *dcmR*, or also in other *dcm* genes. In absence of DCM, DcmR inhibited the transcription of its own gene and of *dcmA*. In addition to DcmR, repression requires the expression of one of the other *dcm* genes by a mechanism that does not involve previously predicted DcmR-binding sites (a 12 bp conserved box shared by *dcmR* and *dcmA* promoters). Mutants with impaired *dcmR* show slower growth with DCM, although no difference in transcript or protein abundance encoded by the *dcm* transposon was observed. The complete *dcmR-dcmA* intergenic region was shown to be required to activate *dcmA* expression. This suggests that regulatory sites present in the intergenic region may be involved in DcmR-independent transcription activation. Taken together, these data allowed to propose a new model of *dcm* gene regulation. The *dcmR* gene was detected and quantified in similar amounts as *dcmA* in DCM-contaminated environmental samples, despite the fact that bioinformatics analysis of sequence databases suggests that *dcmR*-like genes are not necessary associated with the catabolic *dcm* transposon. Thus, DcmR may play a role in other contexts. This may provide new leads for future investigations of potential ligands of the MEDS domain.

Keywords: Chlorinated solvent; Dehalogenase; Dichloromethane; Methylotrophy; *Methylorubrum extorquens* DM4; Transcription factor