

Thèse présentée pour obtenir le grade de
Docteur de l'Université de Strasbourg



Discipline Science du Vivant

Mention : Aspects moléculaires et cellulaires de la biologie

Par Jessica Cleiss-Arnold

Réponse bactérienne au stress arsénié

Soutenue publiquement le 9 Septembre 2010

Composition du jury :

- M. Daniel Muller (Université Claude Bernard Lyon) - Examineur
- M. Axel Hartke (Université de Caen Basse Normandie) - Rapporteur externe
- M. Pierre Peyret (Université Blaise Pascal Clermont-Ferrand) - Rapporteur externe
- M. Vincent Phalip (Université de Strasbourg) - Rapporteur interne
- Mme Florence Arsène-Ploetze (Université de Strasbourg) - Co-directrice de thèse
- M. Philippe Bertin (Université de Strasbourg) - Directeur de thèse

Thèse présentée pour obtenir le grade de
Docteur de l'Université de Strasbourg



Discipline Science du Vivant

Mention : Aspects moléculaires et cellulaires de la biologie

Par Jessica Cleiss-Arnold

Réponse bactérienne au stress arsénié

Soutenue publiquement le 9 Septembre 2010

Composition du jury :

- M. Daniel Muller (Université Claude Bernard Lyon) - Examineur
- M. Axel Hartke (Université de Caen Basse Normandie) - Rapporteur externe
- M. Pierre Peyret (Université Blaise Pascal Clermont-Ferrand) - Rapporteur externe
- M. Vincent Phalip (Université de Strasbourg) - Rapporteur interne
- Mme Florence Arsène-Ploetze (Université de Strasbourg) - Co-directrice de thèse
- M. Philippe Bertin (Université de Strasbourg) - Directeur de thèse

« La rigueur vient toujours au bout de l'obstacle »
Leonard de Vinci

REMERCIEMENTS

A l'issue de ces trois années de Thèse, le moment est venu, pour moi, de remercier toutes les personnes qui m'ont entouré, pour mener à bien ce projet. La liste est longue et je vais essayer de n'oublier personne.

Je tiens tout d'abord à remercier les membres du Jury d'avoir accepté de juger ce travail.

Je remercie Philippe Bertin mon directeur et Florence Arsène-Ploetze ma co-directrice de Thèse, pour avoir dirigé ce travail de manière complémentaire, pour toute la confiance qu'ils m'ont accordé pendant ces trois années.

Je tiens aussi à remercier Florence Goulhen-Chollet, pour son enthousiasme, ses compétences et son aide précieuse. Un grand merci à Sandrine Koechler pour son aide, ses nombreux conseils, j'ai beaucoup appris à tes côtés, merci pour ton soutien et ton amitié.

Je remercie également tous ceux que j'ai eus le plaisir de côtoyer durant ces quelques années. Merci évidemment à toute l'équipe « Arsenic » : Didier, Marie-Claire, Marie, David et François, j'ai été heureuse de faire un bout de chemin avec vous.

Je n'aurais pas pu mener cette Thèse jusqu'au bout sans le soutien de ceux qui me sont le plus proches. J'adresse une tendresse toute particulière à mon mari, mes parents et à toute ma famille pour leurs encouragements tout au long de mes études et vis-à-vis de mes choix.

A toutes et à tous, un grand
merci !

SOMMAIRE

Liste des illustrations	13
Liste des abréviations	15

1^{ère} Partie : Introduction **19**

I. LES REPONSES AU STRESS CHEZ LES BACTERIES **21**

A/ Définition du stress **23**

B/ La réponse générale au stress **24**

1- Biogenèse, intégrité de l'enveloppe cellulaire et morphologie de la cellule bactérienne	25
1.1- Exemple de la motilité et du biofilm	25
<i>a) Mode planctonique : une motilité active</i>	27
<i>b) Mode biofilm : une structuration en communauté</i>	30
2- Protection et réparation des constituants macromoléculaires	31
2.1- Protection et réparation des protéines	32
<i>a) Le repliement des protéines : les chaperons</i>	32
<i>b) La protéolyse : les protéases</i>	33
2.2- Réparation de l'ADN : le système SOS	34
3- Homéostasie et détoxification du cytoplasme	35
3.1- Les pompes d'efflux : « Multidrug Resistance »	35

II. LA REGULATION DE L'INITIATION DE LA TRANSCRIPTION : UNE COMPOSANTE MAJEURE DE LA REPONSE AU STRESS CHEZ LES BACTERIES **41**

A/ L'ARN polymérase bactérienne: enzyme centrale de la transcription **44**

1- Structure de l'ARN polymérase bactérienne	44
2- Le promoteur bactérien	45

B/ Diversité des mécanismes de régulation de l'initiation de la transcription **46**

1- Les facteurs sigma	46
1.1- La famille σ^{70}	46
<i>a) Les facteurs primaires</i>	47

<i>b) Les facteurs alternatifs</i>	47
(i) Exemple du facteur sigma E	48
1.2- La famille σ^{54}	49
1.3- Régulation des facteurs σ	51
2- Les facteurs de transcription	51
2.1- Modulation des facteurs	52
<i>a) Le système à deux composants : un pivot de la réponse transcriptionnelle</i>	
2.2- Rôle des facteurs de transcription	55
<i>a) L'activation 'simple'</i>	55
<i>b) La répression 'simple'</i>	58
<i>c) La co-dépendence des régulateurs</i>	59
C/ La régulation du flagelle :	62
une réponse transcriptionnelle hiérarchisée	

III. L'ARSENIC : UN STRESS CHEZ LES BACTERIES ?	65
--	-----------

A/ Propriétés générales de l'arsenic	67
B/ Toxicité de l'arsenic	67
C/ Mécanismes d'adaptation chez les bactéries	68
1- Séquestration de l'arsenic	69
2- Les métabolismes bactériens	69
2.1- Réduction de l'arséniate et arsénite	71
<i>a) ArsC : la réductase</i>	71
<i>b) ArsB/ArsA ou Acr3 : la pompe d'efflux</i>	72
<i>c) ArsR et ArsD : les régulateurs de l'opéron ars</i>	73
<i>d) ArsH : un gène essentiel ?</i>	74
2.2- Oxydation de l'arsénite en arséniate	74
<i>a) Métabolisme hétérotrophe vs chimioolithoautotrophe</i>	75
<i>b) L'arsénite oxydase</i>	76
(i) Structure de l'enzyme	76
(ii) Localisation de l'enzyme	76
(iii) Fonctionnement de l'enzyme	76
D/ <i>Herminiimonas arsenicoxydans</i> : un organisme de référence	78
1- Caractéristiques du génome	78

IV. OBJECTIFS	81
2^{ème} partie : Travaux personnels	85
Chap 1 : Régulation de l'activité arsénite oxydase chez <i>Herminiimonas arsenicoxydans</i>	87
A/ Introduction	89
B/ Manuscrit	90
C/ Discussion	104
Chap 2 : Mécanismes globaux du stress arsénié chez <i>Herminiimonas arsenicoxydans</i>	107
A/ Introduction	109
B/ Manuscrit	110
C/ Discussion	122
Chap 3 : Cinétique du stress arsénié chez <i>Herminiimonas arsenicoxydans</i>	123
A/ Introduction	125
B/ Manuscrit	125
C/ Discussion	151
Chap 4 : Caractérisation des souches de <i>Thiomonas</i> sp. par génomique descriptive	153
A/ Introduction	155
B/ Manuscrit	156
C/ Discussion	170

Chap 5 : Mécanismes globaux du stress arsénié chez <i>Rhizobium</i> sp. NT-26	173
A/ Introduction	175
B/ Résultats et Discussion	177
Chap 6 : Conclusion générale et perspectives	189
3^{ème} Partie : Annexes	195
Annexe 1 : Des bactéries modifient leur génome pour survivre à l'arsenic	199
Annexe 2 : Diversity of trophic interactions inside an arsenic-rich microbial ecosystem	203
4^{ème} Partie : Le monitorat	227
5^{ème} Partie : Bibliographie	231

LISTE DES ILLUSTRATIONS

- Figure 1** : Diversité des stress et des réponses aux stress chez les bactéries.
- Figure 2** : Structure, assemblage et fonctionnement du flagelle bactérien.
- Figure 3** : Représentation schématique du système chimiosenseur d'*Escherichia coli*.
- Figure 4** : Formation d'un biofilm « mature ».
- Figure 5** : Diversité des stress pouvant endommager l'ADN.
- Figure 6** : Représentation schématique des cinq familles de pompes d'efflux.
- Figure 7** : L'ARN polymérase bactérienne et son interaction avec le promoteur.
- Figure 8** : La réponse σ^E au stress de l'enveloppe.
- Figure 9** : Organisation en domaines du facteur σ^{54} .
- Figure 10** : Le système à deux composants.
- Figure 11** : Mécanismes de l'activation 'simple' au niveau de promoteur.
- Figure 12** : Représentation schématique des domaines fonctionnels d'un régulateur σ^{54} -dépendant.
- Figure 13** : Mécanismes de la répression 'simple' au niveau du promoteur.
- Figure 14** : Mécanismes de co-dépendance de deux activateurs au niveau du promoteur.
- Figure 15** : Cascade de régulation des gènes du flagelle.
- Figure 16** : Diagramme des différents métabolismes bactériens impliqués dans la biotransformation de l'arsenic dans l'environnement.
- Figure 17** : Les deux familles ArsC.
- Figure 18** : Photographie en microscopie électronique de *Herminiimonas arsenicoxydans*.
- Figure 19** : Organisation des gènes impliqués dans l'activité arsénite oxydase chez *Herminiimonas arsenicoxydans*.
- Figure 20** : Synthèse de polymères en réponse à l'arsenic chez *Herminiimonas arsenicoxydans*.
- Figure 21** : Régulation de l'opéron *aox* chez *Agrobacterium tumefaciens*.
- Table 1** : Protéines accumulées différenciellement en réponse à l'As(III) chez *Rhizobium* sp. NT-26.

Les figures des articles sont numérotées de manière indépendante.

LISTE DES ABREVIATIONS

aa	acides aminés
ABC	ATP Binding Casette
ADN	Acide DésoxyriboNucléique
ADNc	AND complémentaire
AMPC	Adénosine MonoPhosphate cyclique
ARN	Acide RiboNucléique
ARNm	ARN messenger
ARNt	ARN de transfert
As	Arsenic
As(III)	Arsénite
As(V)	Arséniate
BER	Base Excision Repair
CAI	Codon Adaptation Index
CCW	CounterClockWise
CDS	CoDing Sequences
CGH	Comparative Genomic Hybridization
CRP	cAMP Receptor Protein
CTD	Carboxy-Terminal-Domain
CW	CloCkwise
DMA	Drainage Minier Acide
PBP	Periplamic Binding Protein
ECF	ExtraCytoplasmic Factor
EPS	ExoPolySaccharide
GSH	Glutathion
GSSG	Glutathion disulfide
HAP	Hook Associated Protein
HPLC	High Performance Liquid Chromatography
HSK	Histidine Serine Kinase
Hsp	Heat shock proteins
HTH	Hélice-Tour-Hélice
ICP-AES	Inductively Coupled Plasma-Atomic Emission Spectrometry
MATE	Multidrug And Toxin Extrusion
MCP	Methyl-accepting Chemotaxis Protein

MDR	MultiDrug Resistance
MFS	Major Facilitator Superfamily
NER	Nucleotide Excision Repair
NIMS	Nanostructure-Initiator Mass Spectrometry
NTD	amino(N)-Terminal-Domain
pb	paires de bases
Pit	Phosphate inorganic transporter
PME	Protéines de la Membrane Externe
Pst	Phosphate specific transporter
RND	Resistance Nodulation cell Division
ROS	Reactive Oxygen Species
rpm	rotation par minutes
RR	Régulateur de Réponse
SMR	Small Drug Resistance
SOS	Save Our Souls
TAT	Twin Arginine Transporter
TCS	Two Component System
Trx	Thioredoxine
T3SS	Système de secretion de type III
2D-DIGE	Two dimensional fluorescence DIfference Gel Electrophoresis

Une étude bibliographique décrivant l'état des connaissances sur le stress chez les micro-organismes est présentée dans la première partie de ce manuscrit. Elle propose une brève description des réponses au stress chez les micro-organismes avec une distinction entre la réponse générale et la réponse spécifique. En sachant qu'une adaptation à un environnement changeant est la plus souvent gouvernée par un réseau de régulation d'un nombre important de gènes, les mécanismes de régulation de l'initiation de la transcription seront décrits dans un second point. Les propriétés de l'arsenic, sa toxicité et la diversité des métabolismes bactériens vis-à-vis de ce métalloïde font l'objet d'un troisième point, qui précède l'énoncé des objectifs de cette étude.

I. LES REPONSES AU STRESS CHEZ LES BACTERIES

L'influence de l'environnement sur la cellule bactérienne est énorme en raison de la petite taille des micro-organismes, qui sont donc en contact étroit avec le vaste milieu qui les entoure. Pourtant, les micro-organismes sont capables de coloniser des écosystèmes aux conditions environnementales parfois sévères en termes de pH, pression, salinité, disponibilité en eau, radiations ionisantes, métaux lourds... Dans un tel contexte, un micro-organisme survit et se développe lorsqu'il se trouve dans un état d'équilibre dynamique caractérisé par un cycle d'échange avec le milieu extérieur. Une perturbation biotique ou abiotique survenant dans son environnement peut constituer un facteur de stress et peut de ce fait induire un éloignement de cet équilibre.

A/ Définition du stress

On peut définir le stress comme une condition hostile qui tend à empêcher un système de fonctionner de façon optimale. Toutefois, si on se limite à cette assertion, il reste à définir l'état optimal. Aussi, le stress peut être considéré comme l'ensemble des conditions qui provoquent des changements de processus physiologiques conduisant éventuellement à une inhibition de croissance voire en dommages cellulaires. Ces changements imposent à la cellule de développer des mécanismes de résistance qui permettent la survie ou simplement une meilleure croissance (Booth, 2002).

La diversité des stress et les mécanismes de réponse développés par les micro-organismes sont présentés dans la Figure 1.

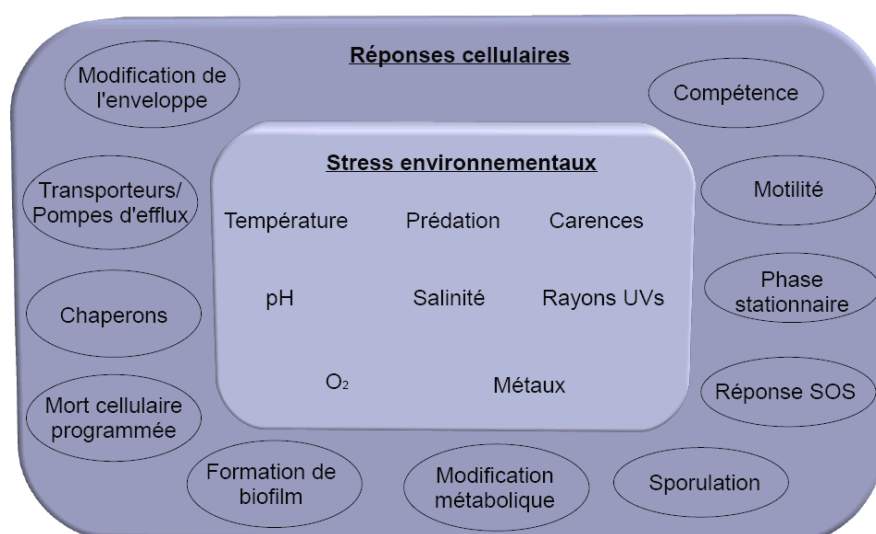


Figure 1 : Diversité des stress et des réponses aux stress chez les bactéries.

Ainsi, les stress peuvent être biotiques (concurrence) ou abiotiques (physico-chimique) et affectent principalement deux « compartiments » : l'enveloppe cellulaire et le cytoplasme. Les réponses chez les bactéries visent soit à rétablir l'homéostasie (réponse SOS, expression de chaperons, synthèse de transporteurs ...), soit à se protéger des effets du stress (modification de l'enveloppe, sporulation, formation de biofilm, motilité...). Par ailleurs, la réponse peut être spécifique et/ou générale c'est-à-dire dépendante ou indépendante de la nature du stress. La réponse générale au stress sera décrite dans un premier temps, avant d'aborder un exemple de réponse spécifique.

B/ La réponse générale au stress

La réponse générale regroupe l'ensemble des mécanismes de résistance décrits chez la plupart des bactéries étudiées à ce jour pour des stress variés. On considère que ce type de réponse joue un rôle essentiel pour la survie des bactéries. On distingue la réponse au niveau de l'enveloppe cellulaire, des constituants macromoléculaires et du cytoplasme.

1- Biogenèse, intégrité de l'enveloppe cellulaire et morphologie de la cellule bactérienne

L'enveloppe bactérienne confère à la cellule bactérienne sa forme et joue un rôle essentiel dans la physiologie de la bactérie. En effet, elle est le siège des transports, de la production d'énergie, d'échanges d'informations et de transduction du signal.

L'enveloppe, par sa localisation, constitue une interface entre l'environnement et le milieu intracellulaire et est donc la plus exposée aux stress extracellulaires. Ainsi, un stress environnemental peut par exemple induire un problème d'assemblage ou une dénaturation des protéines de la membrane externe (PME) et peut ainsi conduire à une accumulation de protéines mal repliées dans le périplasme. Cette accumulation peut devenir toxique pour la bactérie si elle n'est pas enrayée. Il n'est donc pas étonnant que la physiologie de l'enveloppe joue un rôle majeur dans l'adaptation aux changements environnementaux. L'architecture de l'enveloppe cellulaire est ainsi orchestrée par une régulation fine en réponse au stress. La réponse de la bactérie consiste notamment à modifier la production de nouvelles protéines de la membrane externe, d'activer la synthèse de protéases, de chaperons ou de protéines de maturation. Chez *Pseudomonas aeruginosa*, par exemple, la synthèse de protéines de la membrane externe est induite en réponse au chrome (VI) (Kılıç

et al., 2009). Beaucoup de ces réponses sont contrôlées par des facteurs sigma extra-cytoplasmiques (ECF). Le système σ^E sera présenté plus en détail dans la partie consacrée aux facteurs sigma alternatifs (II. B2) car sa compréhension a été nécessaire pour l'interprétation des travaux de thèse.

On assiste également à une modification de la morphologie de la cellule en réponse au stress chez de nombreuses bactéries (Young, 2006). Chez *Bacillus subtilis* par exemple des spores sont formées en réponse à une carence en nutriments (Errington, 1993). Chez *Vibrio cholerae* O1 il a été montré que les exopolysaccharides qui déterminent la morphologie des colonies, sont impliqués dans la résistance aux stress osmotique et oxydatif (Wai *et al.*, 1998). Chez *Escherichia coli*, la plupart des signaux environnementaux, notamment une forte concentration en métaux, induisent une réponse au stress et aboutissent à la formation d'un biofilm (Landini, 2009). Chez *Campylobacter jejuni*, la présence d'un flagelle rend la bactérie mobile et capable de fuir une acidité extrême (Reid *et al.*, 2008).

Ainsi, les principales modifications de la morphologie sont : la formation de spores de résistance, l'acquisition de structures extracellulaires utilisées pour la motilité des cellules ou pour donner la possibilité d'adhérer à une surface voire de former un biofilm. Le cas de la motilité et du biofilm seront décrits en détails ci-dessous car certaines observations effectuées lors de mes travaux de thèse reposent sur ce mécanisme.

1.1- Exemple de la motilité et du biofilm

La motilité permet aux bactéries de réagir à des stimuli, en étant attirée par des substances nutritives comme les sucres, les acides aminés, l'oxygène ou au contraire d'être repoussées par des substances nuisibles.

Il existe cinq catégories différentes de motilité : le 'gliding', le 'twitching', le 'sliding', le 'swimming' et le 'swarming'.

Le 'gliding' est un mouvement par glissement de la cellule sur des surfaces solides, sans structure extracellulaire particulière. Cette motilité est particulièrement observée dans trois groupes bactériens : *Mycobacteria*, *Cyanobacteria* et *Cytophaga-Flavobacterium* (Harshey, 2003).

Le 'twitching' est défini comme une motilité intermittente par contraction. Il nécessite un pilus (ou fimbriae) de type IV, tout comme certaines formes de 'gliding'. Les pili sont

des annexes cellulaires de 2 à 10 nm de diamètre et jusqu'à plusieurs μm de long. Ils sont répartis tout autour de la cellule (péritriche) ou à un pôle (polaire) et sont constitués par polymérisation de la piline. A plusieurs reprises le pilus va s'ancrer, se prolonger et se contracter, permettant ainsi à la bactérie de se déplacer. Ce mode de déplacement est retrouvé dans une large gamme de bactéries telles *Aeromonas*, *Comomonas*, *Ralstonia*, *Vibrio*... (Harshey, 2003).

Le 'sliding' se fait par translocation passive. Il a été observé chez beaucoup de bactéries, telles *Alcaligenes*, *Bacillus*, *Escherichia*, *Flavobacterium*, *Vibrio*... (Harshey, 2003).

Les deux derniers mécanismes ('swimming' et 'swarming') sont corrélés à la présence d'un flagelle (Jarrell & McBride, 2008). La nage ('swimming') est le type de motilité le plus fréquent dans les milieux liquides ou semi-solides. L'essaimage ou rampe ('swarming') est une alternative à la nage qui est observée lorsque les bactéries flagellées se retrouvent sur un milieu solide. Le passage de la nage à l'essaimage est souvent accompagné d'un changement de morphologie de la cellule. Chez *Proteus mirabilis*, un seul système flagellaire péritriche est responsable des deux types de déplacements ('swarming' et 'swimming'), alors que chez *Vibrio parahaemolyticus*, deux systèmes assurent chaque forme de motilité : le flagelle polaire est responsable de la nage, alors que des flagelles multiples assurent la rampe (Merino *et al.*, 2006). Les bactéries étudiées lors de mes travaux sont mobiles grâce à un flagelle. Cette motilité permet aux bactéries de se développer selon deux modes de vie que l'on appelle planctonique ou biofilm et qui sont présentés ci-dessous.

a) Mode planctonique : une motilité active

En mode planctonique, les bactéries se déplacent isolément à l'aide d'un (polaire) ou plusieurs flagelles (péritriche) qui émergent de la surface cellulaire et leur confèrent une capacité de déplacement actif (motilité). La longueur du flagelle est variable (typiquement de 5 à 10 μm) et le diamètre est de 20 nm. Chaque flagelle est composé d'un long filament hélicoïdal, d'un crochet et d'un corps basal constitué d'un domaine central, qui contient le moteur et d'une série d'anneaux (Fig. 2). Les différentes parties sont décrites ci-dessous :

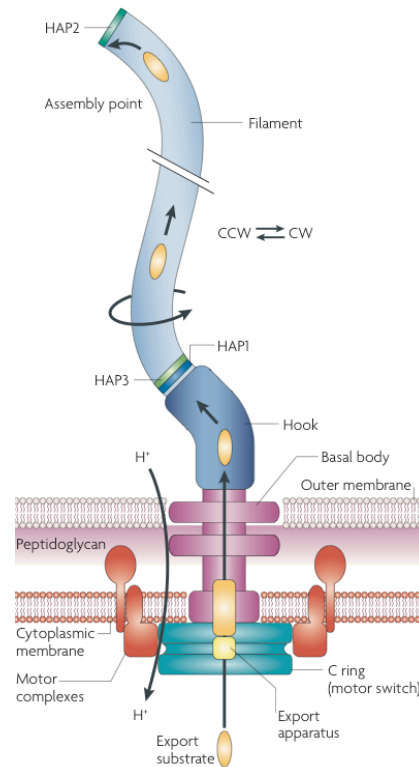


Figure 2 : Structure, assemblage et fonctionnement du flagelle bactérien

CCW, 'counterclockwise' (anti-horaire) ; CW, 'clockwise' (horaire) ; HAP, 'hook-associated protein' (protéines associées au crochet) (extrait de Jarrell & McBride, 2008).

Le corps basal est ancré dans la membrane alors que le crochet et le filament sont des éléments extracellulaires. Les flagelles des bactéries à Gram négatif ont la particularité de posséder des anneaux supplémentaires. En effet, on distingue un anneau L dans la couche lipopolysaccharidique de la membrane externe, un anneau P dans la couche de peptidoglycane et un anneau MS localisé dans la membrane interne. Les anneaux P et L ne sont pas présents chez les bactéries à Gram positif. Enfin, s'ajoute à cela un dernier anneau : l'anneau C (cytoplasmique) qui est composé de trois protéines : FliG, FliM et FliN qui sont nécessaires pour la rotation du flagelle mais aussi pour le contrôle du sens de rotation.

Le moteur agit comme un dispositif rotatif réversible et utilise, selon les organismes, la force proto-motrice ou un gradient de sodium comme source d'énergie. Le moteur flagellaire est constitué de cinq protéines : MotA et MotB qui fonctionnent comme « stator » et permettent le passage de protons, et trois protéines qui composent le « rotor », FliG, FliM et FliN. On distingue (i) les filaments activés par le gradient de proton (moteur à H^+) et (ii) ceux activés par une pompe Na^+ / H^+ (moteur à Na^+). Ces deux moteurs sont

incroyablement rapides : 18,000 rpm pour un moteur à H⁺ et 100,000 rpm pour un moteur à Na⁺ (Magariyama *et al.*, 1994).

L'appareil d'exportation est localisé au niveau du pore formé par l'anneau MS. Il est responsable de la sécrétion des composants du flagelle, c'est à dire aussi bien le crochet que la flagelline.

Le crochet est une structure cylindrique reliée au corps basal, sur laquelle vient se greffer la flagelline.

Le filament flagellaire est constitué de plusieurs filaments de flagelline et forme une structure hélicoïdale.

Des protéines de jonction sont situées entre le crochet et le filament et les **protéines de coiffe** terminent le filament.

L'assemblage du flagelle s'effectue de manière séquentielle, des structures proximales aux structures distales : d'abord la partie basale, puis le crochet et enfin le filament. Un système de sécrétion de type III spécifique au flagelle (T3SS) est localisé dans la membrane cytoplasmique. Celui-ci est indispensable pour l'assemblage des différents composants et est constitué de neuf protéines au début de la biosynthèse du flagelle. Cet assemblage nécessite l'intervention de chaperons pour chaque protéine exportée vers le filament : FliS (chaperon de FliC ; la flagelline), FliT (chaperon de FliD ; protéine qui recouvre le filament), FlgN (chaperon de FlgK et FlgL ; protéines de jonctions filament/crochet) et FliA (σ^{28}) (chaperon de FlgM ; un facteur anti-sigma). FliJ recycle les chaperons libres (sauf FliS).

Chez beaucoup de bactéries, le sens de rotation du flagelle est contrôlé par un système chimiotactique qui transmet les informations de l'environnement à l'intérieur de la bactérie. On distingue deux types de chimiotactisme selon la nature des substances : un chimiotactisme positif (substances attractives) et un chimiotactisme négatif (substances répulsives). Les chimiorécepteurs sont ancrés dans la membrane plasmique et peuvent détecter les substances attractives et/ou répulsives. Actuellement, 20 chimiorécepteurs de substances positives et 10 de substances négatives sont connus. Leurs mécanismes sont complexes, et un des plus étudiés est celui des MCPs (protéine chimiotactique accepteuse de méthyles) chez *E. coli*. Les MCPs peuvent reconnaître divers nutriments ou stress environnementaux. Du côté périplasmique ou extracellulaire, chaque MCP forme une boucle qui sert de site de fixation pour une ou plusieurs molécules attractives ou répulsives (Fig. 3) (Wadhams & Armitage, 2004).

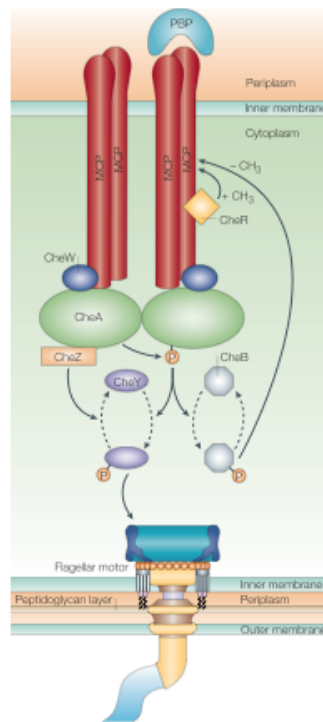


Figure 3 : Représentation schématique du système chimiosenseur d'*Escherichia coli*. Deux chimiorécepteurs dimériques –MCP- sont visualisés, l'un interagissant avec une protéine de liaison périplasmique (PBP). P représente le groupement phosphoryl (extrait de Wadhams & Armitage, 2004).

Le monomère CheW et le dimère CheA interagissent avec la MCP au niveau cytoplasmique. En absence de molécules attractives ou répulsives, la méthyltransférase CheB phosphorylée va déméthyle la MCP et le système maintient une concentration intermédiaire en CheA-phosphorylée et CheY-phosphorylée. Il en résulte un mode de déplacement normal, c'est-à-dire ponctué de courses et culbutes.

La méthylation des récepteurs MCP est favorisée en présence d'un substrat. Celle-ci est effectuée à l'aide de S-adenosylméthionine et est catalysée par la méthyltransférase CheR. La fixation de la substance attractive à la MCP conduit à une déphosphorylation de CheA et CheY. La déphosphorylation de CheA et CheY est catalysée par CheZ. L'absence de fixation de CheY sur FliM permet de tourner dans le sens antihoraire. Le mouvement résultant est une course (Wadhams & Armitage, 2004).

Une diminution de l'attractant induit l'autophosphorylation du dimère CheA, qui phosphoryle CheY et CheB. CheY phosphorylé se lie à la protéine interruptrice du flagelle FliM et induit un changement de direction (Wadhams *et al.*, 2003 ; Wadhams *et al.*, 2000).

Il s'agit ici d'un mécanisme proposé pour *E. coli* et *Salmonella enterica*, mais celui-ci ne représente qu'un exemple de mécanisme de transduction du chimiotactisme parmi d'autres. En particulier, beaucoup d'organismes comme les bactéries du sol ont plus de MCP que les cinq présentes chez *E. coli* (Armitage, 2006). D'autre part, chez *Rhodobacter sphaeroides*, la régulation du chimiotactisme nécessite une protéine CheA2 similaire à celle d'*E. coli* et deux protéines atypiques CheA3 et CheA4. Chez ce micro-organisme, il n'y a pas d'autophosphorylation de CheA3 et CheA4 comme cela a été décrit chez *E. coli*, mais une phosphorylation de CheA3 par CheA4 (Porter & Armitage, 2004).

b) Mode biofilm : une structuration en communauté

En mode biofilm, les bactéries se rassemblent pour former une communauté microbienne englobée dans une matrice de polymères et adhérente à une surface. La formation d'un biofilm « mature » sur une surface s'effectue généralement selon les cinq étapes successives suivantes (Busscher *et al.*, 1995) :

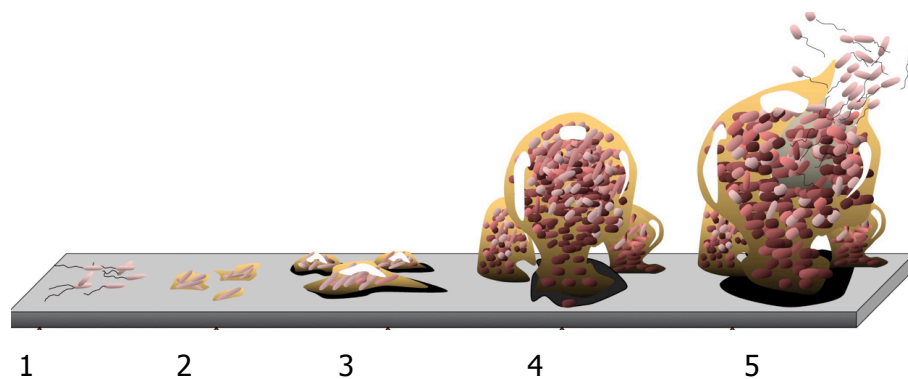


Figure 4 : Formation d'un biofilm « mature ». La description détaillée de la figure est donnée ci-dessous.

1- L'adhésion initiale réversible est établie par interactions physico-chimiques non covalentes (Van der Waals, Lewis...) entre les macromolécules présentes à la surface des micro-organismes et le support récepteur. **2- L'adhésion permanente** est possible grâce à des interactions moléculaires plus spécifiques entre les structures polymériques des micro-organismes (capsule, exopolymères, pili...) et le support. **3- La colonisation de la surface et 4- la maturation** sont établies par multiplication et spécialisation des micro-organismes. Au sein du biofilm des relations s'établissent entre les bactéries conduisant à une réponse cellulaire intégrée. **5- La dispersion** est induite par le vieillissement du

biofilm, certains stress ou carences. Celle-ci se fait par détachement d'un fragment ou libération des bactéries planctoniques ayant recouvré leur motilité.

Les bactéries qui vivent sous forme de communauté ont plusieurs avantages par rapport à une cellule bactérienne isolée. Elles peuvent optimiser i) leur croissance et leur survie en attribuant des fonctions spécialisées à des types de cellules distinctes, ii) l'accès aux substances nutritives (source de carbone, azote, phosphore ...), à une source d'énergie et à l'eau iii) les mécanismes de défense pour se protéger d'un stress.

Ainsi, la production de matrice extracellulaire peut souvent être associée à la séquestration de certaines molécules, notamment les métaux (Teitzel & Parsek, 2003). En effet, les polymères qui composent le biofilm possèdent des résidus chargés négativement à leur extrémité. La pénétration des ions métalliques toxiques dans un biofilm peut donc être modifiée ou perturbée. Les métaux lourds sous forme de cations peuvent réagir avec les groupements phosphodiester et carboxylates de certains biofilms, ce qui retarde leur diffusion dans le biofilm. Par ailleurs, ces cations peuvent également se lier aux thiolates, sulfates ou phosphates et sous cette forme être séquestrés (Harrison *et al.*, 2007). La séquestration des métaux a été mise en évidence chez de nombreuses souches et pour de nombreux métaux, on peut notamment citer l'exemple de *Pseudomonas aeruginosa* PAO1 qui est capable de séquestrer le fer, le cuivre ou l'or (Langley & Beveridge, 1999).

L'enveloppe cellulaire constitue la première ligne de défense et est impliquée dans de nombreuses réponses au stress. Le cytoplasme, lui, a un degré de complexité supérieur car il contient l'ADN et tous les autres constituants macromoléculaires. A l'intérieur de la cellule, on peut distinguer deux mécanismes de réponse aux stress : d'une part la protection et la réparation des constituants macromoléculaires et d'autre part, la détoxification du cytoplasme et le maintien de l'homéostasie.

2- Protection et réparation des constituants macromoléculaires

Les stress physico-chimiques altèrent très largement les protéines, l'ADN et l'ARN en créant des dommages majeurs ou *via* l'accumulation de produits toxiques. Les protéines peuvent être altérées directement par les stress hyperthermique et acide par exemple, qui déstabilisent les liaisons covalentes ou indirectement par le stress oxydatif par exemple qui génère des radicaux libres, intermédiaires réactifs qui peuvent oxyder les résidus cystéines (Storz & Imlay, 1999). L'ADN peut aussi être affecté par différents stress : les stress thermiques vont influencer l'état de super-enroulement, des cassures de brins peuvent être

causés par les radicaux libres générés lors d'un stress oxydatif, une dépurination et une dépyrimidation lors d'un stress acide ou encore une cassure double brin lors d'un stress ionisant (Fridovich, 1998).

Les protéines et les acides nucléiques sont des constituants essentiels au fonctionnement de la cellule, il apparaît donc évident que leur altération est très dommageable pour la cellule. Une des réponses majeures au stress chez les micro-organismes est l'induction de protéines impliquées dans leur protection ou leur réparation.

2.1- Protection et réparation des protéines

Les modifications des protéines face à un stress peuvent être structurales ou oxydatives. Trois processus sont impliqués dans la réponse aux dommages des protéines : (i) la réparation des dommages oxydatifs, (ii) le repliement des protéines endommagées, et (iii) la protéolyse. Le premier aspect sera développé dans la partie consacrée à la réponse spécifique. Les deux derniers points sont développés ci-dessous.

a) *Le repliement des protéines : les chaperons*

Un des effets majeurs du stress est la dénaturation des protéines. En effet, certaines protéines peuvent perdre leur configuration native, fonctionnelle et ont tendance à s'agréger. Un composant essentiel de la réponse au stress est l'activation d'un certain nombre de « gènes de stress » qui protègent la cellule des effets destructeurs. Il s'agit d'un mécanisme rapide et direct de réponse, qui favorise la renaturation des protéines partiellement dénaturées. Toutefois, la plupart de ces protéines dites « de stress » agissent également en absence de stress et jouent un rôle critique dans la biogenèse des protéines. Elles aident les autres protéines cellulaires (i) à se replier correctement pendant et après la traduction ; (ii) à migrer vers leur localisation finale ; (iii) à s'organiser lorsqu'elles fonctionnent en polymères ; (iv) à séquestrer les protéines à agréger ou (v) empêcher leur agrégation, c'est la raison pour laquelle la plupart de ces protéines sont appelées « chaperons ». Par ailleurs, certaines protéines participent à la dégradation de polypeptides. Dans la littérature, ces protéines sont souvent appelées « heat-shock proteins » (Hsp) pour des raisons historiques. En effet, ces chaperons ont été identifiés pour la première fois lors de l'étude de l'effet d'un choc thermique, mais on a pu montrer depuis qu'elles sont également impliquées dans la majorité des stress chez la plupart des micro-organismes. Chez *Mycobacterium bovis*, par exemple, les Hsp sont

impliquées dans la réponse au chlorure de sodium (Jang *et al.*, 2009). Chez *Desulfovibrio vulgaris*, ces mêmes protéines sont impliquées dans l'adaptation à la salinité (He *et al.*, 2010). Cinq chaperons moléculaires constituent le protéome minimal de réponse au stress : DnaK (Hsp70), DnaJ (Hsp40), GrpE, Hsp60 et la peptidyl-prolyl isomerase (PPIase) (Kültz, 2005). Cette forte conservation dans le monde bactérien souligne l'importance des chaperons moléculaires dans la réponse au stress. Beaucoup d'aspects fonctionnels de ces cinq chaperons moléculaires sont connus en détail, mais ils ne seront pas développés ici car ils ne sont pas nécessaires à la compréhension du travail.

b) La protéolyse : les protéases

Malgré l'intervention des chaperons, certaines protéines peuvent être altérées lors d'un stress. Ainsi, beaucoup de protéines sont dégradées par protéolyse et re-synthétisées *de novo*. Les protéases jouent un rôle prépondérant dans la dégradation de protéines altérées grâce à leur activité protéolytique. Parmi les mécanismes majeurs de réponse au stress, on distingue les protéines très conservées suivantes : prolyl endopeptidase, aminobutyrate aminotransferase, serine protease, FtsH, (HflB : sous-unité régulatrice du protéasome), Lon (Kültz, 2005). Toutes ces protéases font parties d'un complexe protéique constitué de sous-unités multiples qui ont : (i) un domaine core protéolytique, (ii) des domaines régulateurs avec une activité ATPase, et (iii) un domaine moléculaire chaperon (Hlaváček & Váchová, 2002). Les mécanismes de ces enzymes sont décrits ci-après en utilisant les protéases FtsH et Lon comme exemple, car ces deux protéases sont induites en réponse à de très nombreux stress. Par ailleurs, FtsH est la seule protéase ATP-dépendante intégrée à la membrane, qui est conservée de manière universelle chez tous les micro-organismes (Karata *et al.*, 1999). FtsH est essentielle, alors que Lon peut être substituée par des protéases comme Clp (caséinolytique). FtsH et Lon permettent l'insertion des protéines dans la membrane et le désassemblage ou l'oligomérisation de complexes protéiques. FtsH utilise l'état de repliement de ces protéines substrats comme critère de dégradation, ce qui constitue un haut degré de sélection des protéines à dégrader (Herman *et al.*, 2003). Chez les bactéries, les substrats de FtsH sont notamment SecY, la protéine FtsZ impliquée dans la division cellulaire et le facteur alternatif σ^{32} (Akiyama *et al.*, 1996 ; Anilkumar *et al.*, 2001 ; Bertani *et al.*, 2001). Lon reconnaît également des motifs spécifiques des protéines clefs, mais la reconnaissance de substrats moins spécifiques se fait en coopération avec des chaperons moléculaires, notamment DnaK et DnaJ (Jubete *et al.*, 1996 ; Savell'ev *et al.*, 1998).

2.2- Réparation de l'ADN : la réponse SOS

Le patrimoine génétique des organismes est constamment menacé par les conditions environnementales agressives qui endommagent l'ADN. Les stress induisant des dommages sont variés : le stress oxydatif, les radiations, le stress acide, les métaux, les antibiotiques... (Fig. 5).

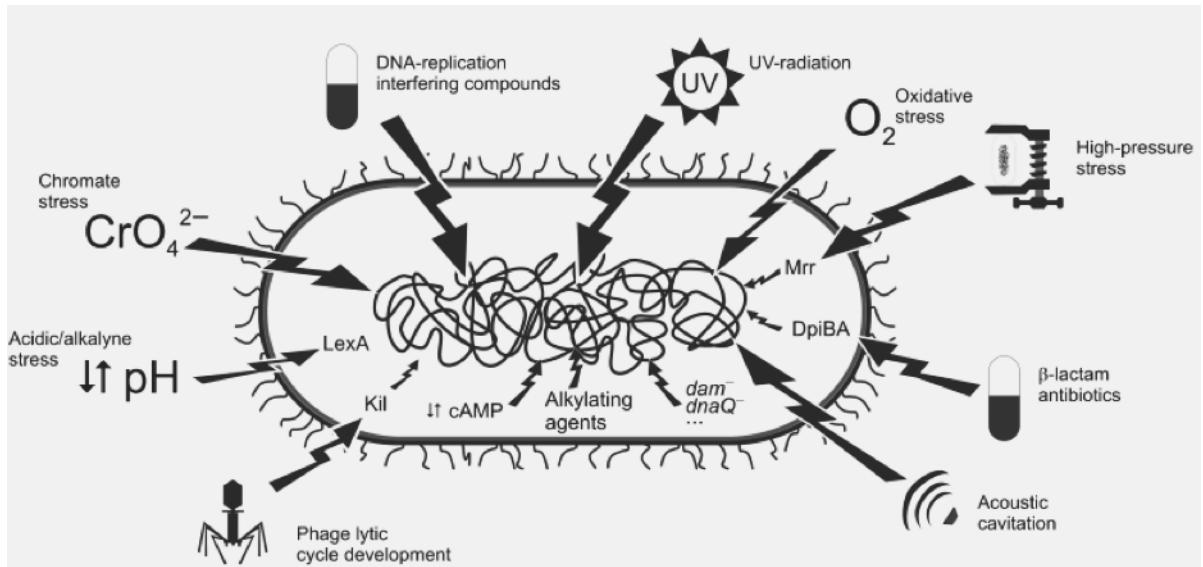


Figure 5 : Diversité des stress pouvant endommager l'ADN (extrait de Erill *et al.*, 2007).

Ces dommages subits peuvent entraîner la mort de la cellule. Aussi, les micro-organismes ont développé plusieurs mécanismes permettant de prévenir les dommages à l'ADN, de mettre en œuvre le système de réparation voire d'induire le suicide si l'atteinte est trop grave. La cellule dispose ainsi de plusieurs mécanismes enzymatiques de réparation tels que la réversion des lésions ponctuelles, l'excision de base (BER), l'excision de nucléotide (NER), la recombinaison ou la réparation de mésappariement ('mismatch repair'). Lorsque les dégâts sont trop importants et que les autres mécanismes ne sont pas capables de réparer l'ADN, le système SOS (Save Our Souls) se déclenche. Cette multiplicité des mécanismes de réparation souligne l'importance pour la cellule de gérer avec soin l'intégrité des chromosomes, la multiplication des cellules et leur survie. Le système SOS est développé ici, car ce mécanisme de régulation global est impliqué dans la majorité des réponses au stress.

Le système SOS a été identifié pour la première fois dans les cellules d'*E. coli* traitées par irradiation ou d'autres agents endommageant l'ADN tel que la mitomycine C (Costa De

Oliveira *et al.*, 1987). Depuis, il a été mis en évidence dans d'autres réponses au stress telles que la résistance au mercure chez *E. coli* K17 (Onnis-Hayden *et al.*, 2009) ou la réponse au stress thermique, oxydatif et acide, chez *Listeria monocytogenes* (van der Veen *et al.*, 2010). Il constitue un mécanisme prévenant les lésions d'ADN chez *E. coli* et est régulé par les produits des gènes *lexA* et *recA* qui agissent respectivement comme activateur et répresseur. Ces deux gènes appartiennent au système SOS (Little & Mount, 1982). Chez *E. coli*, le système SOS est constitué de 40 gènes environ régulés par LexA et RecA (Fernández De Henestrosa *et al.*, 2000). Lors d'une altération de l'ADN, la protéine RecA s'associe à l'ADN simple brin, ce qui lui confère une activité protéolytique qui détruit le répresseur LexA. Ainsi, s'il y a beaucoup de forme simple brin, le répresseur LexA sera inactivé. Des homologues de séquence au gène *lexA*, sont retrouvés dans la plupart des génomes bactériens séquencés. Ainsi, *lexA* semble être universel à toutes les réponses SOS chez les procaryotes (Erill *et al.*, 2007).

3- Homéostasie et détoxification du cytoplasme

En plus des modifications de structure des protéines et de l'ADN énoncées jusqu'ici, la composition du cytoplasme peut être altérée par diverses conditions de stress. On peut assister à une acidification, une modification de la composition ionique, une accumulation de produits toxiques... Les mécanismes de réponse passent généralement par l'induction de systèmes de transport ou de détoxification.

3.1- Les pompes d'efflux : « Multidrug Resistance »

Les bactéries peuvent résister à de nombreux composés par exportation active grâce à des transporteurs membranaires appelés pompes d'efflux. En effet, les pompes d'efflux peuvent intervenir dans la protection contre les effets de nombreux toxiques de l'environnement (métaux lourds, pesticides, solvants organiques, antiseptiques, sels biliaires...) et jouer un rôle dans la détoxification et le maintien de l'homéostasie qui sont importants dans la réponse au stress. Elles peuvent être spécifiques d'un substrat donné ou au contraire de plusieurs composés : MDR (**M**ultidrug **R**esistance). Pour fonctionner ces pompes utilisent l'énergie fournie par dissipation d'un gradient de protons (Familles MFS : **M**ajor **F**acilitator **S**uperfamily, RND : **R**esistance **N**odulation **C**ell division, SMR : **S**mall **M**ultidrug **R**esistance) ou d'ions sodium (Famille MATE : **M**ultidrug **a**nd **T**oxin **E**xtrusion)

ou encore par hydrolyse de l'ATP (Famille ABC : **A**T**P** **B**inding **C**asette) (Fig.6) (Martinez *et al.*, 2009).

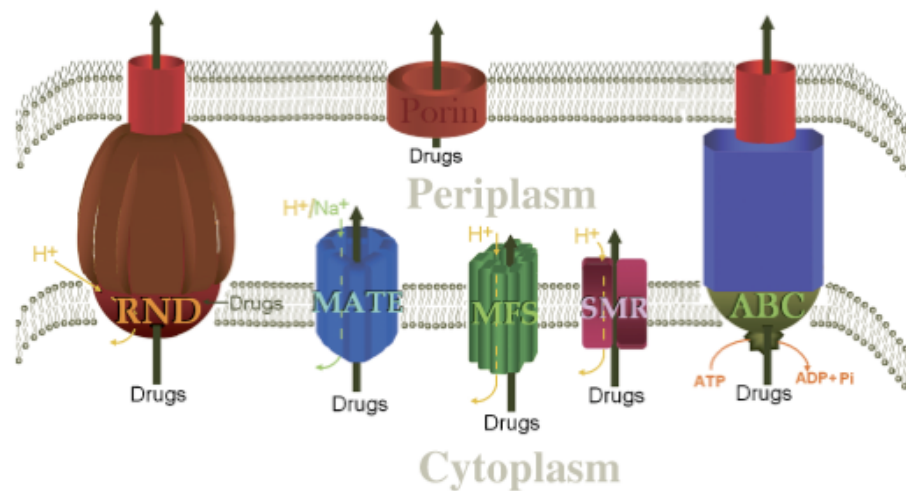


Figure 6 : Représentation schématique des cinq familles de pompes d'efflux (extrait de Martinez *et al.*, 2009)

La majorité des protéines appartenant à la superfamille des RND possèdent 12 segments transmembranaires (Tseng *et al.*, 1999). On retrouve notamment le système EmhABC qui est impliqué dans la résistance à des composés toxiques tels que l'ampicilline, le chloramphénicol, le cristal violet, chez *Pseudomonas fluorescens* (Tian *et al.*, 2010).

La famille MATE comprend des protéines à 12 segments transmembranaires (Brown *et al.*, 1999). NorM chez *Vibrio parahaemolyticus* ou son orthologue chez *E. coli* (Morita *et al.*, 1998) sont les deux pompes de type MATE qui sont caractérisées à ce jour. Elles permettent l'efflux de différents composés toxiques tels que la norfloxacine, la kanamycine, l'éthidium ou la streptomycine et constituent une protection contre les espèces réactives de l'oxygène (ROS) (Guelfo *et al.*, 2010 ; Morita *et al.*, 2000).

La superfamille des MFS est constituée de 30 sous-familles (Pao *et al.*, 1998). On retrouve notamment le système CsbX qui est impliqué dans l'efflux du catécholate chez *Azotobacter vinelandii* (Page *et al.*, 2003).

La famille SMR est constituée de petites protéines d'environ 100 à 110 acides aminés (aa) et quatre segments transmembranaires (Paulsen *et al.*, 1996). Le système AbeS chez *Acinetobacter baumannii* est un SMR qui est impliqué dans la résistance à divers antibiotiques, détergents et teintures (Srinivasan *et al.*, 2009).

La superfamille des ABC est constituée de 52 sous-familles. Elle inclut les systèmes d'efflux de nombreux substrats notamment les sucres, les aa, les ions métalliques, les

peptides... On peut citer l'exemple de LmrA qui confère une résistance à de nombreux antibiotiques et composés toxiques chez *Lactococcus lactis* (Poelarends *et al.*, 2002) ou SmdAB qui est impliqué dans la résistance à de fortes concentrations en tétracycline, DAPI (4'6-diamidino-2-phenylindole) ou Hoechst 3334, chez *Serratia marcescens* (Matsuo *et al.*, 2008).

Les réponses énoncées jusqu'ici peuvent être induites par de nombreux stress et ont été décrites chez de nombreux micro-organismes. On considère que ce type de réponse joue un rôle essentiel dans la survie des bactéries dans l'environnement. Toutefois, de nombreux mécanismes d'adaptation peuvent être spécifiques à un stress donné.

C/ La réponse spécifique à un stress

Les modalités de réponse au stress oxydatif seront abordées dans cette partie car l'identification des spécificités de cette réponse a été nécessaire à l'analyse des résultats de mes travaux.

L'oxygène moléculaire a un rôle fondamental dans les fonctions cellulaires et représente un net avantage énergétique dont les bactéries aérobies savent tirer profit pour leur métabolisme, mais l'oxygène est aussi un élément dangereux. Toutes les cellules possèdent des systèmes, tels que l'ascorbate, le glutathion, la thioredoxine et d'autres enzymes anti-oxydantes, qui permettent de minimiser ou de réparer les dommages oxydatifs.

La toxicité de l'oxygène vis-à-vis d'un micro-organisme est généralement attribuée aux espèces réactives de l'oxygène (ROS) comme le radical superoxide (O_2^-), l'ion superoxyde (OH^\cdot) ou le peroxide d'hydrogène (H_2O_2) qui détériorent les protéines, les lipides, les acides nucléique et constituent ainsi des causes de la mort cellulaire (Van de Guchte, 2002). Cette production de ROS est induite en condition de stress. Ainsi, chez *Desulfovibrio vulgaris* ou *Mycobacterium bovis* une induction des protéines impliquées dans la réponse au stress oxydatif (catalase, gluathion) a été observée respectivement, en réponse au stress salin ou au chlorure de sodium (Jang *et al.*, 2009 ; He *et al.*, 2010).

Les mécanismes de défense des bactéries consistent à (i) éliminer les espèces réactives par dégradation enzymatique par activité catalase ou superoxyde dismutase, (ii) protéger les protéines des méfaits du stress oxydatif par des thiols et (iii) réparer les dommages causés par des mécanismes généraux de réponse aux stress. Ce dernier point a été décrit dans la partie précédente. Aussi, ce ne sont que les deux premiers points qui seront détaillés ici.

1- Elimination des espèces réactives

La superoxyde dismutase SodA assure l'élimination du radical superoxyde O_2^- et conduit à la formation de peroxyde d'hydrogène H_2O_2 , qui est un composé moins toxique. Cette conversion de O_2^- permet donc de limiter les réactions d'oxydation qui pourraient affecter les constituants macromoléculaires de la cellule. SodA est impliquée dans la réponse au stress oxydatif (Sanders *et al.*, 1995) mais participe aussi à la réponse au stress acide (Frees *et al.*, 2003) et probablement à d'autres stress. La catalase permet ensuite la dismutation de $2H_2O_2$ en $2H_2O + O_2$. Ces deux enzymes doivent être co-régulées en réponse au stress car un déséquilibre du ratio superoxyde dismutase/catalase en présence d'ions métalliques peut conduire à une conversion de H_2O_2 en radicaux hydroxyles nocifs. L'accumulation de radicaux hydroxyles peut être délétère pour la cellule car ils peuvent endommager les macromolécules telles que l'ADN, les protéines ou les lipides (Imlay & Linn, 1988 ; Keyer & Imlay, 1996).

2- Protection des protéines par des thiols

Les dommages oxydatifs des protéines peuvent être de différentes formes, les plus répandus sont l'oxydation de résidu cystéine (où on assiste à la formation de ponts disulfures) ou l'oxydation de méthionine. Les thiol-transférases telles que le glutathion et la thiorédoxine sont impliquées dans la réparation de ce type de dommages, notamment la réduction des ponts disulfures.

Le glutathion (gamma-glutamyl-cysteinyl-glycine, GSH) est un groupement thiol de faible poids moléculaire synthétisé *de novo*. Lorsqu'il est présent à de forte concentration, il permet de maintenir un environnement intracellulaire réducteur. La voie principale du métabolisme du GSH est la réduction d'hydroperoxydes par la glutathione peroxidase et des peroxiredoxines qui génèrent le glutathion disulfide (GSSG). La glutathion-S-tranférance

catalyse la conjugaison du glutathion. La glutathion réductase catalyse la réduction NADH-dépendante de GSSG en GSH. L'activité glutathion réductase est essentielle dans la réponse au stress car lorsque les ROS augmente, elle permet le maintient de GSH dans la cellule (Dickinson & Forman, 2002).

La thioredoxine est une protéine de 12 kDa. Les enzymes impliquées dans la réparation des dommages aux protéines *via* le système thiorédoxine sont la thioredoxine peroxidase et la thioredoxine réductase. Cette dernière transfert les électrons du NADPH vers la thiorédoxine et assure ainsi sa ré-oxydation (Arnér & Holmgren, 2000).

La méthionine sulfoxide réductase permet quant à elle de réparer les dommages causés aux méthionines (Stadtman, 2004).

**II. LA REGULATION DE L'INITIATION DE LA
TRANSCRIPTION : UNE COMPOSANTE
MAJEURE DE LA REPONSE AU STRESS
CHEZ LES BACTERIES**

Dans la première partie nous avons vu que les bactéries disposent de différents systèmes pour protéger la cellule des contraintes extérieures. Ces activités sont modulées par les variations des paramètres environnementaux. Ainsi, la plupart du temps, il s'agira pour la bactérie de modifier la synthèse des protéines par un contrôle de l'expression des gènes au niveau transcriptionnel, post-transcriptionnel ou traductionnel. En effet, en activant l'expression de certaines enzymes ou en réprimant la synthèse d'autres, les micro-organismes modifient les capacités métaboliques permettant leur croissance. Des mécanismes divers et variés ont été mis en évidence et il apparaît que chaque étape de l'expression d'un gène (de l'ADN à la protéine) peut être soumise à régulation. La transcription est un point de contrôle très important de l'expression des gènes chez les bactéries. Elle est principalement régulée à deux niveaux : - au niveau de l'initiation et - au niveau de la terminaison.

Le contrôle de l'initiation de la transcription permet une adaptation adéquate aux conditions de l'environnement. En effet, la régulation par des facteurs sigma ou des facteurs de transcription repose sur l'utilisation de protéines instables ou facilement dégradées par les protéases. Ce type de régulation permet de moduler l'expression d'un ensemble de gènes facilement et rapidement. Chez *Burkholderia coenopacia*, par exemple, 315 (4,4%) ou 386 gènes (5,4%) sont induits et 185 (2,6%) ou 331 gènes (4,6%) sont réprimés, après une exposition respectivement à l'H₂O₂ ou au NaCl (Peeters *et al.*, 2010). De plus, différents régulateurs peuvent intervenir en réponse à un même stress environnemental et former un réseau de mécanismes complexes parfois organisé en une cascade de régulation. Ainsi, plusieurs types de réponses adaptatives peuvent être connectés entre eux, de telle manière que lorsqu'un mécanisme d'adaptation est déclenché, d'autres réponses sont associées.

Par ailleurs, plusieurs gènes peuvent être sous le contrôle d'un même régulateur, on parle alors d'un régulon. Ainsi, en modulant l'expression ou l'activité de ce régulateur, la cellule peut activer ou réprimer l'expression d'un groupe de gènes. Un autre mécanisme possible est l'induction de plusieurs gènes en réponse à un même stimulus mais par le biais de différents régulateurs, dans ce cas ci on parle de stimulon. Enfin, un même gène peut être soumis à deux ou plusieurs mécanismes de régulation. Ce type de gène code souvent des protéines nécessaires à la survie de la cellule à divers stress (Cases & de Lorenzo, 2005). Il n'est donc pas rare d'identifier plusieurs sites de fixation pour différents régulateurs dans les régions promotrices des gènes bactériens. Les composants du système de régulation de l'initiation de la transcription seront développés dans un premier temps.

L'exemple de la cascade de régulation du système flagellaire sera exposé dans un second point, car celui-ci paraît nécessaire à la compréhension du manuscrit.

A/ ARN polymérase bactérienne : enzyme centrale de la transcription

L'initiation de la transcription est un processus qui s'effectue en plusieurs étapes. Elle débute par la liaison de l'ARN polymérase à une séquence ADN qu'elle reconnaît, appelée promoteur. La structure de l'ARN polymérase bactérienne est explicitée dans un premier temps, puis la diversité des mécanismes de régulation sera décrite.

1- Structure de l'ARN polymérase bactérienne

Chez les bactéries une seule ARN polymérase ADN dépendante effectue la transcription de tous les types d'ARN. Cette enzyme est conservée dans l'évolution bactérienne. On distingue 1) le core de l'enzyme, qui est constitué des sous-unités $\beta\beta'\alpha_2\omega$ et 2) l'holoenzyme composée du core associé à une sous-unité σ (Fig. 7) (Ebright, 2000).

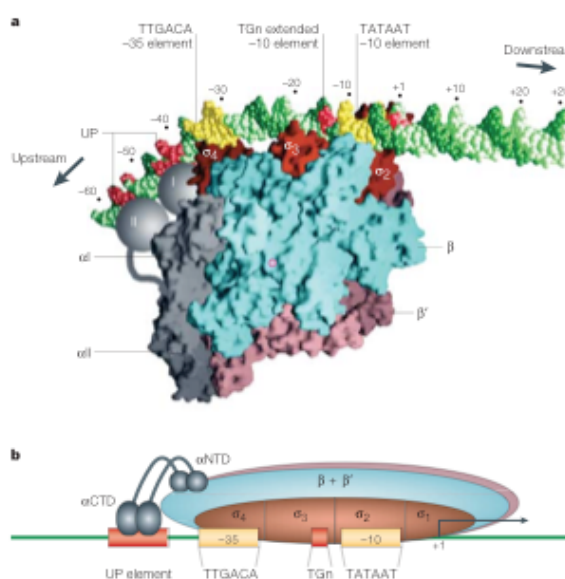


Figure 7 : L'ARN polymérase bactérienne et son interaction avec le promoteur (extrait de Browning & Busby, 2004).

Le site actif est formé par les sous-unités β (en bleu) et β' (en rose) qui assurent la liaison de l'ADN matrice (en vert) et de l'ARN néosynthétisé durant la transcription. Le

core de l'enzyme forme une structure en forme de « pince de crabe », tout comme l'ARN polymérase II de la levure (Zhang *et al.*, 1999 ; Fu *et al.*, 1999).

Les deux sous-unités α sont identiques et liées l'une à l'autre par un 'linker' d'environ 20 aa (Fig. 7 sphères grises I et II). Le domaine amino-terminal (α NTD) dimérise et intervient dans l'assemblage des sous-unités β et β' . Le domaine carboxy-terminal (α CTD) possède un module de liaison à l'ADN qui intervient dans la reconnaissance de certains promoteurs.

La sous-unité ω n'a pas de rôle direct dans la transcription, mais semble être un chaperon intervenant dans le repliement de la sous-unité β' .

Le core enzyme seul ne permet pas d'initier la transcription. En effet, une interaction du core enzyme avec un facteur σ est nécessaire pour que l'ARN polymérase s'associe à un promoteur particulier et débute la transcription.

La sous-unité σ (Fig. 7 en rouge) a trois fonctions principales : 1) assurer la reconnaissance de la séquence spécifique du promoteur, 2) positionner l'ARN polymérase sur un promoteur cible, et 3) faciliter le déroulement de l'ADN double brin au niveau du site d'initiation de la transcription et assurer le passage d'un complexe fermé à un complexe ouvert (Browning & Busby, 2004).

2- Le promoteur bactérien

Les promoteurs contrôlent l'expression de l'ensemble des gènes d'un organisme. Quatre régions ont été caractérisées au niveau des promoteurs bactériens.

Les deux éléments principaux sont les **hexamères -10 et -35** (Fig.7 en jaune), qui sont localisés 10 paires de bases (pb) et 35 pb en amont du site d'initiation de la transcription (+1). Les séquences consensus de ces deux éléments sont : TTGACA pour la boîte -35 et TATAAT pour la boîte -10. Les deux autres éléments du promoteur sont l'élément -10 'étendu' et l'élément 'UP'.

La région -10 'étendue' (Fig. 7 en rouge) est un motif de 3-4 pb (TGn) localisé immédiatement en amont de la boîte -10. Les séquences -10, -35 et -10 'étendue' sont reconnues par la sous-unité σ de l'ARN polymérase.

L'élément UP (Fig. 7 en rouge) est une séquence d'environ 20 pb localisée en amont de la boîte -35 et reconnu par le domaine C-terminal de la sous-unité α de l'ARN polymérase.

Ces quatre régions sont spécifiques de la liaison initiale de l'ARN polymérase à un promoteur. La contribution de chaque élément varie d'un promoteur à un autre. On sait que des promoteurs qui ont des séquences très proches des consensus ont une affinité pour l'ARN polymérase supérieure aux promoteurs qui ont des séquences éloignées. Par ailleurs, beaucoup de promoteurs forts possèdent des éléments UP qui assurent une liaison au domaine α CTD de l'ARN polymérase (Gourse *et al.*, 2000). La fréquence à laquelle l'holoenzyme initie la transcription est donc directement influencée par la séquence du promoteur et par la conformation de l'ADN dans cette région. La diversité des séquences promotrices est donc un puissant outil pour une distribution hétérogène des ARN polymérases et permet d'expliquer comment la cellule peut synthétiser une grande quantité d'un messager et un peu ou pas du tout d'un autre. Cependant, cette diversité ne permet qu'une « régulation » statique, qui ne peut pas être modulée par les conditions environnementales. Aussi, la plupart des régulations adaptatives sont dues à des facteurs agissant en *trans*, tels que les facteurs sigma ou les facteurs de transcription (Busby & Ebright, 1994).

B/ Diversité des mécanismes de régulation de l'expression des gènes

Deux mécanismes moléculaires majeurs participent à la régulation de l'initiation de la transcription : 1) les facteurs sigma (σ) et les facteurs de transcription.

1- Les facteurs sigma

La plupart des espèces bactériennes possèdent différents facteurs σ . Cette diversité permet de diriger l'ARN polymérase vers des classes de promoteurs ayant des séquences consensus distinctes. Ce mécanisme permet à la bactérie de maintenir un niveau d'expression basal de certains gènes et de répondre à un stress en régulant l'expression de gènes spécifiques. En se basant sur les homologies de séquences, on distingue deux familles de facteurs σ : la famille σ^{70} et la famille σ^{54} . Toutes les bactéries ont un ou plusieurs facteurs σ appartenant à la famille σ^{70} . Beaucoup de micro-organismes ont également un facteur appartenant à la famille σ^{54} . Le nombre de gènes codant des facteurs σ varie d'une seule copie chez *Mycoplasma genitalium* (Fraser *et al.*, 1995) à 65 gènes chez *Streptomyces coelicolor* (Bentley *et al.*, 2002).

1.1- La famille σ^{70}

La famille σ^{70} peut être divisée en deux sous-groupes structuraux et fonctionnels (Wösten, 1998), qui sont décrits ci-dessous.

a) *Les facteurs primaires*

Les facteurs primaires sont retrouvés chez tous les micro-organismes. Ils assurent la transcription de la majorité des gènes exprimés en phase exponentielle et sont essentiels pour la survie cellulaire. Chez *E. coli*, le facteur primaire est codé par *rpoD* et est connu sous le nom de σ^{70} . Les séquences des hexamères -10 et -35 sont celles décrites précédemment. Elles sont séparées par 16 à 18 nucléotides non conservés qui constituent l'espaceur. L'élément -10 'étendu' joue un rôle essentiel dans ce groupe de facteur σ car il permet d'avoir une reconnaissance fine de la sous-unité σ^{70} même si les séquences -10 et -35 s'éloignent du consensus. Toutefois, six autres facteurs σ^{70} ont été identifiés en réponse au stress chez *E. coli* (Ishihama, 2000). Lorsqu'ils s'accumulent, ils entrent en compétition avec le facteur primaire, se lient à un certain nombre d'ARN polymérases et initient la transcription de gènes ayant des promoteurs avec des séquences spécifiques. Ces facteurs sont dits alternatifs.

b) *Les facteurs alternatifs*

Les facteurs alternatifs contrôlent des régulons spécifiques en réponse à une condition donnée (Gross, 1992). Ce groupe comprend notamment les facteurs impliqués dans la régulation du flagelle, de fonctions extra-cytoplasmiques (ECF) et du stress thermique. Dans la régulation des gènes du flagelle, FliA ou σ^{28} est nécessaire pour l'expression des gènes de Classe III chez *E. coli* (Helmann, 1991). La fonction et la régulation de σ^D est similaire à FliA chez *B. subtilis* (Caramori *et al.*, 1996). Des analyses comparatives de promoteurs dépendants de ces facteurs ont montré que σ^{28} et σ^D reconnaissent les éléments -35, TAAA, et -10 GCCGATAA, séparées par un espaceur de 15 nucléotides. En condition de stress thermique, la réponse est contrôlée par le facteur σ^{32} codé par *rpoH* chez *E. coli* (Yura *et al.*, 1993) ou *Francisella tularensis* (Grall *et al.*, 2009) par exemple et chez *B. subtilis* la majorité des transcrits sont induits par le facteur σ^B (Hecker *et al.*, 1996). Ici aussi, des séquences spécifiques du promoteur ont été

déterminées : il s'agit de CTTGAAA en -35 et de CCCATnT en -10, espacés de 11 à 16 nucléotides pour le facteur σ^{32} (Gross, 1992). Chez les Entérobactéries et les *Pseudomonas*, on trouve un facteur alternatif spécifique à la phase stationnaire, σ^S ou RpoS (Mulvey & Loewen, 1989). Chez *Mycobacterium smegmatis* la phase stationnaire, la réponse au stress thermique et oxydatif, sont régulées par le même facteur alternatif SigF (Hümpel *et al.*, 2010). Enfin, des séquences spécifiques ont également été déterminées pour les facteurs ECF (par exemple σ^E , avec le consensus GAACTT en -35 et TCTRA en -10, tous deux espacés de 16 à 17 pb). Le système σ^E est décrit ci-dessous car l'étude de son réseau de régulation est nécessaire pour la compréhension de mes travaux.

(i) Exemple du facteur σ^E

Le facteur alternatif σ^E contrôle l'expression de nombreux gènes. Chez *E. coli*, il joue un rôle central dans le maintien de l'intégrité de l'enveloppe cellulaire quand celle-ci est endommagée, mais aussi en condition normale de croissance (Cf I.B1) (Raivio & Silhavy, 2001). Chez *Salmonella enterica*, il est également impliqué dans la pathogénicité de la souche (Muller *et al.*, 2009). Ce facteur alternatif est codé par le gène *rpoE* qui se trouve en opéron avec *rseABC* chez *E. coli*. Il est induit lors de la sur-expression de protéines de la membrane externe (PME) ou lorsqu'elles sont mal repliées et qu'elles s'accumulent dans le cytoplasme (Ades, 2004). Son régulon se compose notamment des gènes codant des protéines impliquées dans la maturation, la dégradation, l'assemblage des PME comme SurA, DegP, YaeT, YfiO, NlpB ou encore les gènes impliqués dans la biosynthèse des lipopolysaccharides (LPS) (Rhodius *et al.*, 2006, Skovierova *et al.*, 2006). σ^E intervient aussi dans sa propre régulation, en initiant la transcription du gène *rpoE*. Enfin, l'activation de σ^E conduit également à la production de petits ARN non codants qui vont inhiber l'expression des PME au niveau post-transcriptionnel (Vogel & Papenfort, 2006).

Les mécanismes de régulation de σ^E ont été bien étudiés chez *E. coli*. Brièvement, le facteur σ^E est séquestré au niveau de la membrane interne *via* RseA (ligne ondulée) qui possède une partie exposée dans le cytoplasme ainsi qu'une autre dans le périplasme (Ades, 2004). En absence de stress (Fig. 8 à droite), l'activité de σ^E est contrôlée par RseB (rectangle), une protéine périplasmique liée à RseA, qui empêche tout clivage de RseA par DegS. L'accumulation de protéines dénaturées active les protéases DegS et RseP (YaeL). Le clivage de RseA par ces deux protéases, permet de libérer σ^E dans le cytoplasme (Alba *et*

al., 2002). σ^E peut ensuite interagir avec l'ARN polymérase et activer la transcription des gènes impliqués dans le repliement des PME ou dans leur dégradation (Cezairliyan & Sauer, 2007).

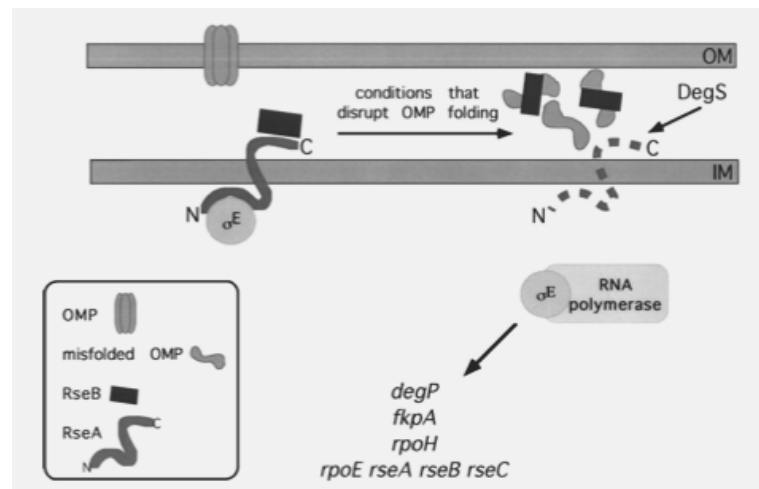


Figure 8 : La réponse σ^E au stress de l'enveloppe (extrait de Raivio & Silhavy, 2001).

1.2- La famille σ^{54}

Historiquement, le facteur σ^{54} (aussi appelé σ^N) a été mis en évidence pour la première fois dans le métabolisme de l'azote et les activateurs NtrC et NifA ont été identifiés. Depuis, son implication dans d'autres activités biologiques a été démontrée et d'autres activateurs ont été caractérisés aussi bien dans l'utilisation de différentes sources d'azote ou de carbone, le métabolisme énergétique, la dégradation de composés aromatiques, le chimiotactisme, le transport d'électrons, la modification d'ARN (Buck *et al.*, 2000).

La famille σ^{54} ne contient qu'un seul groupe de facteurs codé par *rpoN*. Ce facteur, structuralement et fonctionnellement différent des facteurs de type σ^{70} , n'est pas essentiel pour toutes les conditions de croissance. Dans la plupart des cas, son expression est constitutive, mais il n'est pas rare de voir qu'elle est régulée par des facteurs environnementaux (Iannino *et al.*, 2008). Habituellement il n'y a pas plus d'un gène codant le facteur dans un même organisme, à l'exception de *Bradyrhizobium japonicum*, *Rhodobacter sphaeroides* et *Rhizobium etli* qui possèdent deux gènes *rpoN* codant le facteur σ^{54} (Buck *et al.*, 2000).

Trois régions distinctes ont été caractérisées pour ce facteur (Fig. 6).

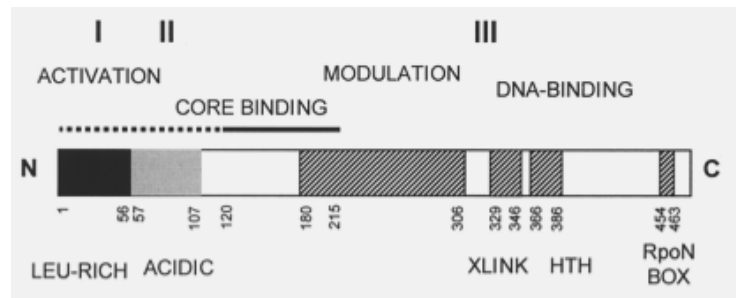


Figure 9 : Organisation en domaines du facteur σ^{54} (extrait de Buck et al., 2000).

La région I, constituée de 25 à 50 aa, est riche en glutamine. Ce domaine a deux fonctions i) inhiber l'isomérisation et l'initiation de la polymérase en absence d'activateur et ii) stimuler l'initiation en présence d'un activateur : il s'agit de la région régulatrice.

La région II est plus variable et contient généralement 60 à 110 aa avec un excès de résidus acides. Cette région II est impliquée dans la transition d'un complexe fermé à un complexe ouvert.

La région III C-terminale d'environ 400 aa contient la région-Xlink qui se lie à l'ADN (Cannon *et al.*, 1994) et deux motifs conservés : une hélice-tour-hélice (HTH) et la boîte RpoN. Ces deux motifs sont nécessaires pour la reconnaissance d'un promoteur σ^{54} -dépendant. La boîte RpoN est caractérisée par 10 aa conservés, ARRTVAKYRE (Taylor *et al.*, 1996). Les séquences adjacentes à cette région III C-term modulent son activité ('modulation' Fig. 9) ou constituent le déterminant de liaison principal (Fig. 9 'core binding' des aa 120 à 250).

Au niveau des promoteurs cibles, les éléments clés pour la reconnaissance d'une ARN polymérase contenant un facteur σ^{54} ($E\sigma^{54}$) sont localisés en -12 et -24, avec la séquence suivante : TGGCAC-N5-TTGC. Un changement d'une seule pb entre les motifs -24GG et -12GC, inactive le promoteur (Merrick, 1993). $E\sigma^{54}$ peut reconnaître le promoteur, mais la formation d'un complexe ouvert nécessite l'interaction d'un activateur avec l'holoenzyme (Morris *et al.*, 1994). En effet, les activateurs possèdent une activité nucléoside triphosphatase qui catalyse la formation d'un complexe ouvert (Rappas *et al.*, 2007). Ainsi, chez *Myxococcus xanthus* par exemple 53 séquences homologues à des activateurs de σ^{54} ont été identifiés (Jakobsen *et al.*, 2004). Ces activateurs se lient 100 pb ou plus en amont du site d'initiation de la transcription à une séquence UAS (upstream activating sequence) ou ELE (enhancer-like element).

1.3- Régulation des facteurs σ

Les différents facteurs σ peuvent intervenir dans la régulation des réponses cellulaires à différents stress. Il n'est donc pas surprenant que leur activité soit étroitement contrôlée. Dans la plupart des cas, l'activité des facteurs σ est contrôlée par des facteurs anti- σ qui le séquestrent à l'écart de l'ARN polymérase. Les facteurs anti- σ n'ont donc pas d'effet sur la transcription ou la traduction du facteur σ . L'activité des facteurs anti- σ est régulée par la liaison de ligands, de modifications covalentes ou par protéolyse. Dans la régulation des gènes du flagelle par exemple, FlgM est un facteur anti-sigma de σ^{28} (Helmann, 1999). Comme mentionné ci-dessus, pour le facteur σ^{54} , l'initiation de la transcription est dépendante de la liaison avec un co-activateur. Une description de son co-activateur sera donnée en II.B2.

2- Les facteurs de transcription

Une défaillance dans la séquence consensus de l'élément -10 ou plus fréquemment de l'élément -35 peut être l'indication d'une régulation facteur dépendante. En effet, si la séquence ne permet pas de recruter l'ARN polymérase, un activateur est nécessaire. Les protéines qui se lient à l'ADN pour réguler l'initiation de la transcription sont appelées des facteurs de transcription.

Chez *E. coli*, plus de 300 gènes codent des protéines qui sont connues pour se lier au promoteur. Celles-ci peuvent activer ou inhiber l'initiation de la transcription (Pérez-Rueda & Collado-Vides, 2000). Certains régulateurs de réponse ont été décrits comme répresseur dans un cas et activateur dans un autre. Chez *Streptococcus aureus* par exemple, SarA a été décrit comme activateur ou répresseur en fonction des conditions redox et du pH (Fujimoto *et al.*, 2009). Les répresseurs empêchent la fixation de l'ARN polymérase. Les activateurs quant à eux améliorent l'affinité du promoteur pour l'ARN polymérase. La plupart des activateurs se lient au promoteur avant d'agir sur l'ARN polymérase. On pensait que c'était le mécanisme qui s'applique à tous les activateurs de la cellule, mais un mécanisme alternatif a été proposé. En effet, les régulateurs MarA et SoxS par exemple ont été décrits comme interagissant avec l'ARN polymérase avant de se lier à leur ADN cible (Griffith & Wolf Jr, 2004 ; Martin *et al.*, 2002).

La présence et l'utilisation de facteurs de transcription est répandue dans l'ensemble du monde microbien. Certaines de ces protéines peuvent contrôler un ensemble de gènes, alors que d'autres ont des effets spécifiques sur un ou deux gènes. Chez *E. coli*, il a été établi que sept facteurs de transcription (CRP, FNR, IHF, Fis, ArcA, NarL et Lrp) contrôlent 50% des gènes régulés, alors qu'environ 60 facteurs de transcription contrôlent seulement un gène (Grainger *et al.*, 2006 ; Martinez-Antonio & Collado-Vides, 2003).

D'une manière générale, les facteurs de transcription régulent l'expression des gènes en réponse à des signaux environnementaux. Ils sont contrôlés par modulation de leur activité ou de leur expression.

2.1- Modulation des facteurs

Les différents mécanismes décrits sont les suivants :

1) L'affinité du facteur de transcription pour l'ADN cible peut être modulée par des petits ligands. La concentration de ces ligands varie en réponse au stress, ce qui augmente ou diminue l'affinité du régulateur. Un exemple de ce type de mécanisme est la diminution de l'affinité du répresseur LacI pour son ADN cible en présence d'allolactose en réponse au lactose contenu dans le milieu. On peut également citer l'exemple de l'interaction de l'AMP cyclique (AMPC) avec sa protéine réceptrice (CRP : cAMP Protein Receptor) qui permet d'augmenter l'affinité de l'ARN polymérase pour le promoteur de l'opéron *lac* (Müller *et al.*, 1996).

2) Certains facteurs de transcription ont une activité modulée par leur concentration dans la cellule. La concentration cellulaire est déterminée par régulation de l'expression du facteur de transcription ou par protéolyse. Par exemple, une des réponses cellulaires au fer est contrôlée par la concentration de la protéine FecA. La transcription du gène *fecA* est contrôlée par Fur (« iron-dependent ferric uptake regulator ») qui est lui-même régulé directement en réponse au fer (Danielli *et al.*, 2009).

3) Le facteur de transcription peut être séquestré par liaison à une protéine régulatrice (Plumbridge, 2002). Chez *E. coli* par exemple, Rob active la transcription de SoxRS/MarM. En absence d'inducteur (notamment les sels biliaires), Rob est séquestré par SoxS qui l'inactive (Griffith *et al.*, 2009)

4) L'activité de certains facteurs de transcription est modulée par des modifications de conformation. Ainsi certains régulateurs se lient à l'ADN cible uniquement s'ils sont phosphorylés par une protéine senseur kinase (Browning & Busby, 2004). Chez les

bactéries, de nombreux circuits de signalisation fonctionnent en utilisant ce type de « système à deux composants » (TCS). Une description du TCS est donnée ci-dessous car son fonctionnement est nécessaire pour la compréhension de mes travaux.

a) *Le système à deux composants : un pivot de la réponse transcriptionnelle*

Plus de 600 systèmes à deux composants ont été identifiés dans les génomes procaryotes annotés. Les éléments biochimiques de base ont été mis en évidence pour la première fois pour le système NR (Nitrogen Regulator). Chez *E. coli*, ce système contrôle l'expression des gènes en réponse à la disponibilité des sources azotées (Ninfa & Magasanik, 1989). Par la suite, d'autres systèmes bactériens de détection ont été caractérisés et la liste des TCS s'est étendue à des centaines de systèmes distincts. On peut citer l'exemple du régulateur ColRS qui est impliqué dans la résistance au phénol chez *Pseudomonas putida* (Putrinš *et al.*, 2010), les différents TCS tels que CiaHR, LevSR, LiaSR, ComD, impliqués dans la résistance au stress acide chez *Streptococcus* (Putrinš *et al.*, 2010 ; Gong *et al.*, 2009) ou encore CpxRA qui est un TCS qui répond au stress de la membrane et induit la formation de biofilm chez *E. coli* (Ma & Wood, 2009). Enfin, nous verrons dans cette partie que certains co-activateurs de σ^{54} font partie de ce groupe. Chez la plupart des micro-organismes, les gènes codant un système à deux composants sont localisés dans un même opéron, ce qui permet une expression coordonnée.

La membrane est une zone privilégiée pour percevoir tout changement de l'environnement car elle se situe à l'interface environnement-cellule. Les systèmes à deux composants assurent la double fonction de détecteur d'un changement environnemental et de transducteur de réponse. Le système sensoriel à deux composants à la base est constitué de deux protéines : un senseur histidine kinase (HSK) et un régulateur de réponse (RR) (Fig. 10). Ces deux classes de protéines constituent deux superfamilles caractérisées par des motifs d'acides aminés conservés. Une description de chacune d'entre elles est donnée ci-dessous.

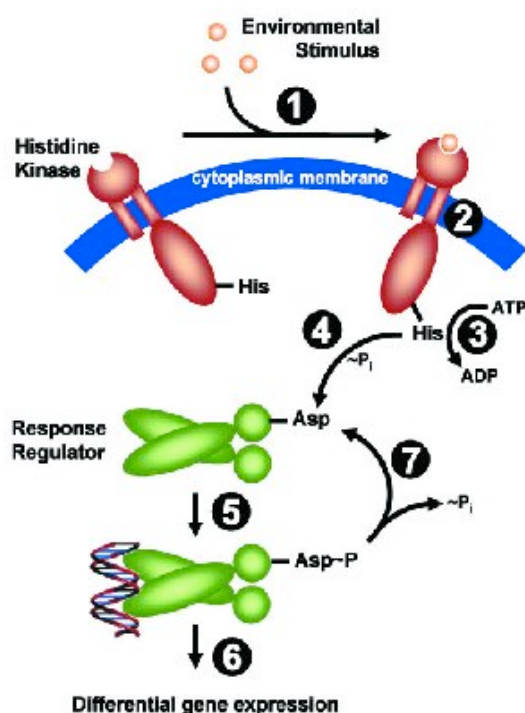


Figure 10 : Le système à deux composants (extrait de Parkinson, 1993).

La protéine HSK est composée d'un domaine senseur qui peut interagir avec les stimuli environnementaux (Fig. 10-1), et d'un domaine transmetteur capable de s'autophosphoryler sur un résidu histidine et de transférer son phosphate à un régulateur de réponse (Fig. 10-2 à 4). Les HSKs fonctionnent en dimères et sont le plus souvent ancrées *via* une ou deux séquences transmembranaires ou associées à la membrane plasmique (Mascher *et al.*, 2006). La majorité des HSKs sont des protéines homodimériques possédant un domaine périplasmique N-terminal senseur et un domaine kinase cytoplasmique C-terminal. Par ailleurs, les HSKs contiennent souvent des modules sensoriels extracellulaires fusionnés au module catalytique protéine kinase, ce qui facilite la détection des stimuli environnementaux (Bourret *et al.*, 1991).

Les domaines senseurs sont variables en terme de séquence, de manière à pouvoir répondre à une variété de signaux environnementaux. Le domaine kinase de 350 aa environ est beaucoup plus conservé, mais n'est pas apparenté par sa structure aux sérines/thréonines ou tyrosines kinases des eucaryotes. Néanmoins, un ensemble de motifs d'acides aminés conservés ont été identifiés dans le domaine kinase et sont appelés boîtes H, N, G1, F et G2 (Stock *et al.*, 1989, Parkinson & Kofoid, 1992).

Le RR possède généralement aussi deux domaines : un domaine récepteur N-terminal conservé de 120 aa environ et un domaine effecteur C-terminal variable, qui est

impliqué dans la réponse adaptative spécifique. Les résidus Asp-13, Asp-57 et Lys-109 appelés boîtes DD, D et K, ont été mis en évidence pour le RR CheY de *E. coli* (Stock *et al.*, 1989). Ces résidus sont particulièrement conservés dans tous les domaines receveurs des RR. Les domaines receveurs ont deux fonctions : 1) ils interagissent avec les HSKs phosphorylées qui catalysent le transfert de phosphate sur un de leur résidu aspartate (Fig. 10-4) et 2) ils régulent les activités de leurs domaines effecteurs (Fig. 10-5). Dans beaucoup de cas, les domaines effecteurs se lient à l'ADN et régulent la transcription de gènes spécifiques lorsque le régulateur est activé (Fig. 10-6). La phosphorylation du RR permet donc de contrôler sa capacité à se lier à une séquence d'ADN cible ou dans d'autres cas d'interagir avec d'autres composants de la machinerie transcriptionnelle (Hakenbeck & Stock, 1996). Un retour à un état antérieur au stress est possible par déphosphorylation du RR, par une activité phosphatase interne du senseur kinase ou externe (Fig. 10-7). Les RRs capables de se lier à l'ADN sont divisés en trois sous-familles appelées OmpR/PhoB, NtrC et FixJ (Stock *et al.*, 1990).

2.2- Rôles des facteurs de transcription

Pour beaucoup de promoteurs l'activation ou la répression est 'simple' c'est à dire qu'elle n'implique qu'un seul facteur de transcription. Toutefois, il n'est pas rare que certains régulateurs soient dépendants d'autres protéines. Dans un premier temps la simple activation et la simple répression seront abordées, puis dans un second temps, la co-dépendance des régulateurs sera présentée.

a) L'activation 'simple'

Trois mécanismes sont connus pour l'activation simple (Fig. 11).

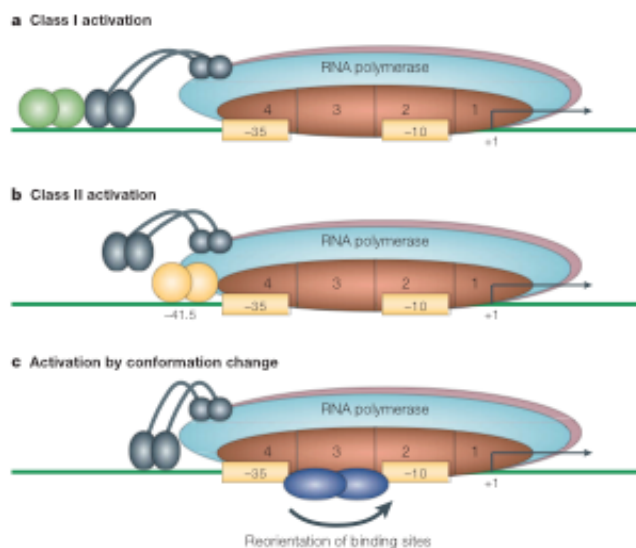


Figure 11 : Mécanismes de l'activation 'simple' au niveau de promoteur.

Beaucoup d'activateurs fonctionnent à l'état de dimère et c'est ainsi qu'ils sont représentés sur cette figure (extrait de Browning & Busby, 2004).

Dans l'activation de **Classe I** (Fig. 11a), l'activateur (en vert) se lie à une cible qui se situe en amont de l'élément -35 du promoteur et recrute l'ARN polymérase en interagissant directement avec la région α CTD de celle-ci. Un exemple d'activateur appartenant à cette classe est SoxS qui active le promoteur de *ompW* dont le produit est une porine impliquée dans la régulation de l'osmolarité (Gil *et al.*, 2009). On peut également citer les co-activateurs du facteur σ^{54} tel GabR par exemple, qui régule l'opéron *gab* impliqué dans la conversion de γ -aminobutyrate en succinate chez *Bacillus thuringiensis* (Zhu *et al.*, 2010). En effet, les co-activateurs de σ^{54} se lient 100 pb ou plus en amont du site d'initiation de la transcription. Ici, on assiste en plus à un changement de conformation de l'ADN. En effet, pour que l'ARN polymérase associée à σ^{54} puisse interagir avec l'activateur pour initier la transcription, l'ADN cible doit se replier sous forme de boucle. Ce type d'activateur est détaillé ici car la compréhension de son mécanisme a été nécessaire lors de mes travaux de thèse. Ces facteurs de transcription possèdent tous trois domaines conservés impliqués dans la réception du signal, l'activation de la transcription et la liaison à l'ADN (Fig. 12).

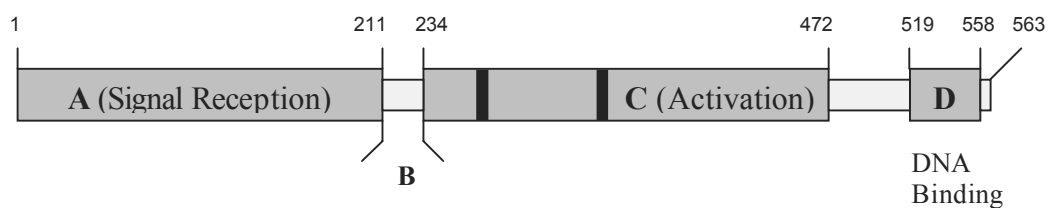


Figure 12 : Représentation schématique des domaines fonctionnels d'un régulateur σ^{54} -dépendant (d'après Shingler, 1996).

Le domaine A en N-terminal est associé au reste de la protéine par un inter-domaine flexible : **le domaine B** ou 'Q-linker'. Le domaine A varie de 12 (HrpS de *Pseudomonas syringae*) à 400 résidus (FlhA d'*E. coli*). Il n'y a pas d'homologie commune à tous les activateurs dans cette région. Toutefois, on peut constituer des sous-groupes en fonction du mécanisme par lequel l'activité du régulateur est modulée par un stimulus. On distingue : la régulation par phosphorylation, la modulation par interaction protéine-protéine, l'activation directe de l'effecteur et la dé-repression. La régulation par phosphorylation est un groupe qui est particulièrement intéressant car il comprend les activateurs de système de régulation à deux composants, c'est le cas de NtrC, CctD, HydG, PilR, AlgB, HoxA, Hup1, RocR et FlbD (Shingler, 1996). Chez *Pirelulla*, 16 activateurs de σ^{54} dont 10 TCS ont été identifiés dans son génome (Studholme & Dixon, 2004).

Le domaine central C (~240 aa) est impliqué dans l'activation de la transcription. C'est le domaine le plus conservé dans tous les activateurs ; sept régions (C1 à C7) ont été identifiées dans ce domaine. Les régions conservées C1 (G-G-GK) et C5 (Q-LLRVL) sont impliquées dans la liaison de l'ATP et son hydrolyse (Fig. 12 boîtes noires). La région C3 constituée de deux hélices alpha est impliquée dans l'interaction avec σ^{54} .

Le domaine D, en C-terminal est constitué de motifs de liaison à l'ADN (HTH), analogues à ceux trouvés dans de nombreux répresseurs et activateurs transcriptionnels.

Les activateurs de **Classe II** (en jaune, Fig. 11b) eux se lient à une cible qui chevauche l'élément -35 (généralement vers -41,5) et au domaine 4 du facteur σ (Dove *et al.*, 2003). Cette liaison permet de recruter l'ARN polymérase sur le promoteur cible (Busby

& Ebright, 1997). C'est notamment le cas de l'activation d'un certain nombre de gènes de virulence par SloR chez *Streptococcus* sp. (O'Rourke *et al.*, 2010) ou par PhhR qui peut se lier 40 pb en amont du site d'initiation du gène et ainsi activer l'expression de différents gènes dont *phhA* impliqué dans le catabolisme de la phénylalanine chez *Pseudomonas putida* (Herrera *et al.*, 2009).

Enfin, les activateurs de **Classe III** (en bleu) altèrent la conformation du promoteur cible, pour favoriser la fixation de l'ARN polymérase avec les éléments -10/-35 du promoteur (Fig. 11c). Ce mécanisme nécessite une liaison de l'activateur aux, ou près des, éléments -10 et -35 du promoteur, ce qui permet leur réalignement et leur interaction avec l'ARN polymérase, c'est pourquoi cette régulation est aussi appelée activation par modification. Par exemple, pour les promoteurs activés par des protéines de la famille MerR, l'espace entre les éléments -10 et -35 n'est pas optimal pour la fixation de l'ARN polymérase. Aussi, les activateurs de type MerR se lie à cet espaceur ce qui entraîne une torsion de l'ADN qui permet de réorienter les boîtes -10 et -35 (Brown *et al.*, 2003).

b) La répression 'simple'

Les répresseurs diminuent la transcription des promoteurs cibles. Pour la répression simple trois mécanismes sont également connus (Fig. 13).

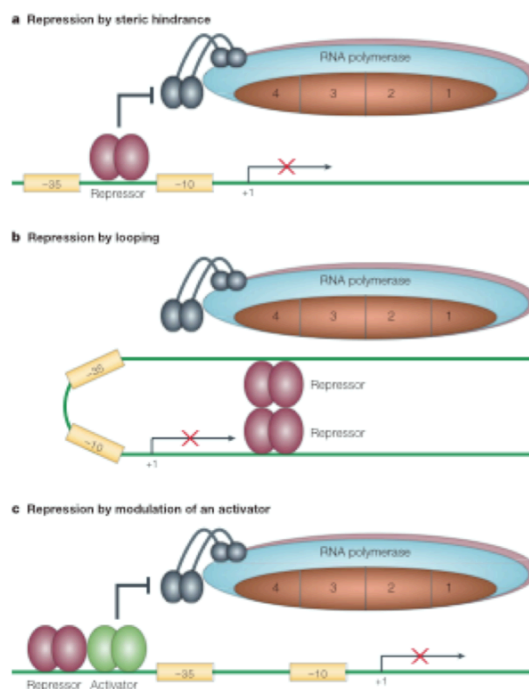


Figure 13 : Mécanismes de la répression 'simple' au niveau du promoteur (extrait de Browning & Busby, 2004).

La répression par encombrement stérique est probablement le mécanisme le plus courant. Celle-ci repose sur la fixation du répresseur aux éléments 'core' d'un promoteur (c'est-à-dire les boîtes -10 et -35) qui empêche la fixation de l'ARN polymérase (Fig. 13a). Dans ce cas de figure, le domaine de liaison du répresseur est situé au niveau, ou près, des éléments -10 et -35 du promoteur (Müller-Hill, 1998). C'est le cas notamment de SloR, qui réprime un certain nombre de gènes de virulence par un mécanisme de ce type (O'Rourke *et al.*, 2010). Pour d'autres promoteurs, plusieurs répresseurs se fixent en des régions distales des éléments -10 et -35. Ces répresseurs imposent la formation d'une boucle dans l'ADN qui contient les domaines d'initiation de la transcription, empêchant ainsi la fixation de l'ARN polymérase (Fig. 13b). A titre d'exemple, chez *E. coli* et *Salmonella* la liaison de plusieurs répresseurs GalR en des sites distants du promoteur de l'opéron galactose *gal* impose une courbure de l'ADN, ce qui réprime la transcription de ce dernier (Choy & Adhya, 1996). Enfin, il existe des mécanismes complexes où le répresseur fonctionne comme anti-activateur, c'est-à-dire qu'il module l'activateur en s'y liant et empêche sa reconnaissance par l'ARN polymérase (Fig. 13c). Un exemple éloquent est celui des promoteurs réprimés par CytR, qui est lui-même activé par CRP. L'interaction de CytR avec CRP confère une spécificité de répression de CytR au niveau du promoteur des gènes *udp* par exemple, qui codent une uridine phosphorylase (Shin *et al.*, 2001).

Récemment, un nouveau mécanisme de répression a été mis en évidence chez les bactéries. Il s'agit du répresseur de l'opéron mannitol, MtlR chez *Vibrio paratumeolyticus* et de son homologue YggD chez *Shigella flexneri* qui est dépourvu de motifs de liaison à l'ADN. La régulation de l'expression des gènes par ce type de régulateur se fait donc indirectement (Tan *et al.*, 2009).

c) La co-dépendance de régulateurs

L'activité de la majorité des promoteurs bactériens dépend de plusieurs facteurs environnementaux plutôt que d'un seul signal. Ainsi, beaucoup de promoteurs sont contrôlés par deux ou plusieurs facteurs de transcription, chaque facteur étant le témoin d'un signal environnemental. Dans de nombreux cas, un seul régulateur peut intégrer différents signaux. C'est le cas du système NifL-NifA de *Azotobacter vinelandii* par exemple, qui contrôle les gènes impliqués dans la fixation de l'azote en réponse à la concentration en oxygène, à la disponibilité en carbone et en azote (Dixon, 2002).

Pour les promoteurs qui sont co-dépendants de deux ou plusieurs activateurs le mécanisme est plus complexe. Quatre mécanismes généraux ont été décrits : le repositionnement de l'activateur, les contacts ARN polymérase-activateur indépendants, la liaison coopérative des activateurs et l'anti-répression par un activateur (Fig. 14)

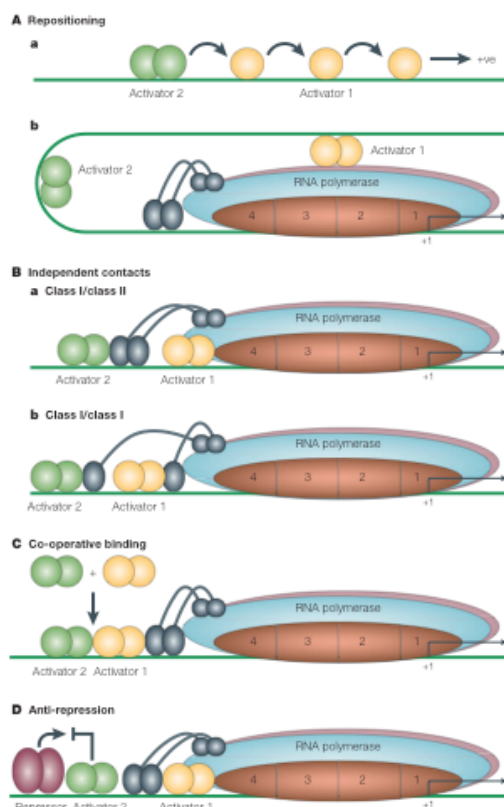


Figure 14 : Mécanismes de co-dépendance de deux activateurs au niveau du promoteur (extrait de Browning & Busby, 2004).

Dans le mécanisme de repositionnement (Fig. 14A), un premier activateur (en jaune) se fixe au niveau du promoteur, mais celui-ci est mal positionné pour initier la transcription des gènes en aval. Le rôle du second activateur (en vert) est donc de repositionner le premier en un endroit où il pourra activer la transcription. Ce repositionnement peut être réalisé par un déplacement de l'activateur d'un site de l'ADN à un autre, comme c'est le cas pour le repositionnement de MalT par CRP au niveau du promoteur de *malK* (Brown et al., 2003) (Fig. 13A,a). Une autre possibilité est que le second activateur modifie la conformation de l'ADN, en lui imposant une courbure, ce qui permet au premier activateur d'interagir avec l'ARN polymérase et ainsi d'activer l'initiation de la transcription (Schröder, 1993) (Fig. 14A,b). Chez *E. coli*, une multitude de protéines sont impliquées dans la compaction de l'ADN, comme Fis (factor for inversion stimulation), IHF (integration host

factor), H-NS et HU (histone-like nucleoid-structuring proteins), StpA (un homologue de H-NS) et Dps (DNA-protein from starved cells) (Azam & Ishihama, 1999). Certaines se fixent en des sites spécifiques tout le long du chromosome et peuvent interférer avec la distribution des ARN polymérase entre les promoteurs. Par exemple, dans la région régulatrice de *nir*, Fis, IHF et H-NS modifient l'ADN, formant ainsi un complexe nucléoprotéique ordonné qui réprime la transcription des gènes en aval (Browning *et al.*, 2000). Dans d'autres exemples, ces mêmes facteurs sont des activateurs de l'initiation de la transcription (McLeod & Johnson, 2001).

Un autre mécanisme possible est la liaison indépendante de deux activateurs. Ici, les deux activateurs établissent des liaisons avec l'ARN polymérase pour activer l'initiation de la transcription de manière optimale (Fig. 14B). Ces liaisons sont indépendantes l'une de l'autre. En effet, aucune interaction entre les deux activateurs n'est requise. Dans certains cas, l'un des activateurs est de Classe I et l'autre de Classe II (Fig. 14B,a) comme c'est le cas pour l'initiation de la transcription de *proP* où Fis est un activateur de Classe II et CRP son co-activateur de Classe I (McLeod *et al.*, 2002). Dans d'autres cas, les deux activateurs sont de Classe I (Fig. 13B,b), c'est le cas de l'activation de *acsP* par liaison en tandem de CRP (Tebbutt *et al.*, 2002). Dans un troisième cas, des activateurs qui coopèrent se lient au promoteur (Fig. 14C). Ici, la liaison du premier dépend de la liaison du second. Un exemple éloquent est celui de l'initiation de la transcription de *melAB* par CRP, qui n'est possible que si MelR est lié préalablement (Wade *et al.*, 2002). De la même manière, chez *E. coli* la transcription du gène *yhjA* dépend de la liaison des activateurs FNR et OxyR (Partridge *et al.*, 2007).

Enfin, dans un dernier mécanisme la liaison d'un deuxième activateur est nécessaire pour lever l'inhibition d'un répresseur et autoriser l'initiation de la transcription par le premier activateur. Ce mécanisme est dit anti-répression. Ici, on peut citer l'exemple très étudié des promoteurs *nir* et *nif*, qui sont codépendents de FNR, un régulateur induit en anaérobiose, et NarL ou NarP, des régulateurs activés par les ions nitrites ou nitrates. La liaison de FNR à l'ARN polymérase par un mécanisme de 'Classe II' active la transcription des deux promoteurs. Néanmoins, la liaison de IHF et Fis à des régions flanquantes annule cette activation. La liaison du second activateur NarL ou NarP est nécessaire pour supprimer l'effet de IHF et Fis (Browning *et al.*, 2002).

C/ La régulation du flagelle: une réponse transcriptionnelle hiérarchisée

Comme mentionné précédemment, différents régulateurs peuvent intervenir en réponse à un même stress environnemental et former un réseau de mécanismes complexes. L'exemple de la cascade de régulation du flagelle est donné ici car la description de son fonctionnement est nécessaire à la compréhension de ce manuscrit.

La structure et l'assemblage du flagelle a été donnée en I.B1. Ainsi, il s'agit d'un organelle extraordinaire en terme de complexité de structure, de mécanisme d'assemblage dont la fabrication représente un investissement énergétique non négligeable. Aussi, pour la plupart des bactéries, une régulation stricte et complexe, notamment au niveau de l'initiation de la transcription, est associée à l'assemblage du flagelle. Ainsi, chez *E. coli*, des études de génétique classique et *in vivo* (Kalir *et al.*, 2001), ont permis de mettre en évidence une cascade de régulation à trois niveaux. On distingue les gènes précoces, intermédiaires et tardifs et des promoteurs de classe I, II et III (Soutourina & Bertin, 2003).

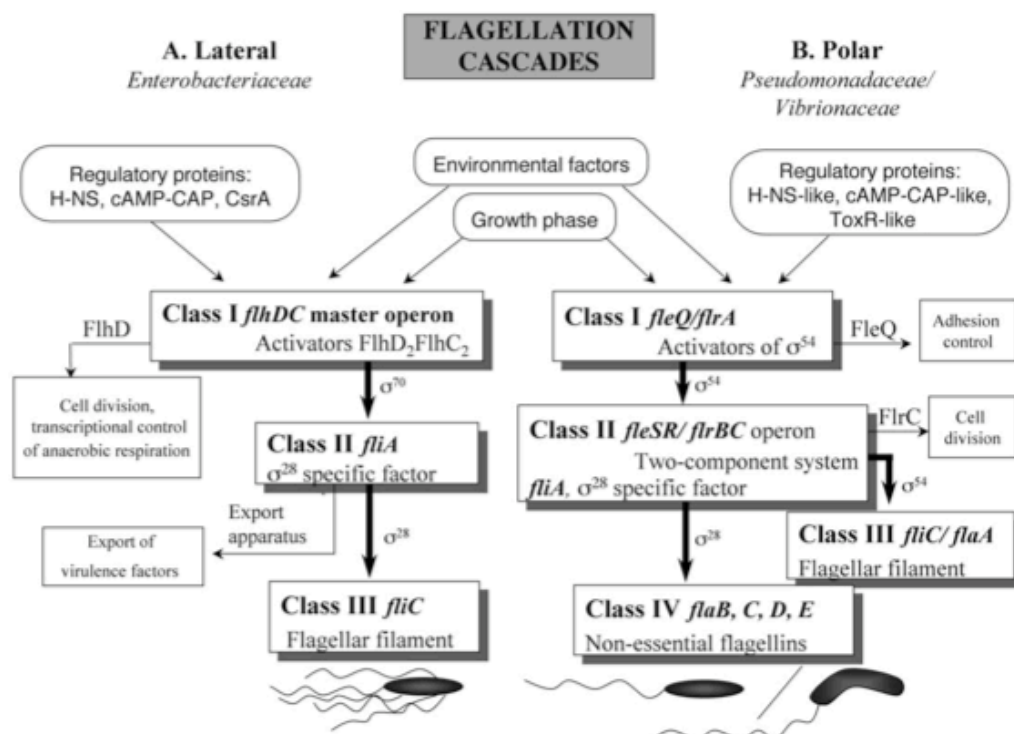


Figure 15 : Cascade de régulation des gènes du flagelle.

Une comparaison des cascades de régulation des flagelles péritriche (famille des *Enterobacteriaceae*) et polaire (familles des *Pseudomonadaceae* et *Vibrionaceae*) est donnée (extrait de Soutourina & Bertin, 2003).

E. coli est une bactérie dotée de flagelles péritriches (Fig.15 **A**). Chez cet organisme, le premier niveau de régulation de l'expression du flagelle est le contrôle de FlhD₂C₂ régulateur maître du flagelle (gènes de classe I). En effet, les protéines FlhC et FlhD sont nécessaires pour la transcription des gènes de classe II. Ces régulateurs sont eux-mêmes, sous le contrôle de régulateurs globaux, de facteurs environnementaux et des régulateurs du quorum-sensing. Les gènes de classe II assurent l'assemblage du T3SS, de la structure basale et du crochet. Le facteur sigma FliA (σ^{28}) est également codé par les gènes de classe II et est nécessaire pour la transcription des gènes de classe III. L'expression des gènes tardifs de classe III assure l'assemblage final du flagelle et son fonctionnement. La transcription des gènes tardifs est inhibée par le facteur anti-sigma FlgM tant que les structures du corps basal et du crochet ne sont pas complètes. Lorsque la transcription des gènes tardifs est autorisée par σ^{28} , deux autres protéines interviennent : FliK et FlhB. FliK régule la longueur du crochet et interagit avec FlhB, un composant clé du T3SS qui fonctionne comme porte d'exportation.

Les régulateurs qui interviennent dans la cascade de régulation sont différents d'un type flagellaire à un autre. Aussi, pour les bactéries à flagelle polaire (comme les *Pseudomonadaceae* et les *Vibrionaceae*), les gènes *fleQ* et *firA* se trouvent au sommet (gènes de classe I). Ces gènes codent des activateurs de type NtrC, associés au facteur σ^{54} . FleQ et FirA activent la transcription des gènes de Classe II, notamment *fleSR* et *firBC*. FleSR et FirBC sont des systèmes à deux composants qui agissent sur la transcription de gènes de Classe III, comme *fliC* ou *fliA*. Un autre gène de Classe II, *fliA*, code le facteur σ^{28} qui participe à la régulation des gènes de Classe IV comme les gènes non-essentiels de flagelline *fliB*, *C*, *D*, *E*, chez les *Vibrionaceae*.

Par ailleurs, certaines bactéries possèdent une cascade mixte c'est-à-dire des gènes caractéristiques d'une cascade polaire et d'autres appartenant à une ciliature péritriche dans une même cascade de régulation. La cascade à trois niveaux énoncée ici pour *E. coli*, n'est donc pas universelle. Par exemple, *Pseudomonas aeruginosa* et *Helicobacter pylori* possèdent une cascade à quatre niveaux (McCarter, 2006), certains spirochètes utilisent une régulation post-transcriptionnelle (Charon & Goldstein, 2002) et *Caulobacter crescentus* possède un système de régulation qui dépend du cycle cellulaire (England & Gober, 2001). Enfin, chez *Vibrio alginolyticus* le passage régulé des ions par un antiport Na^+/H^+ permet une différenciation du système flagellaire : polaire dans un cas (Na^+), péritriche dans l'autre (H^+) (Yorimitsu & Homma, 2001).

III.

**L'ARSENIC : UN STRESS CHEZ
LES BACTERIES ?**

Différents composants abiotiques ou biotiques peuvent constituer un stress chez les bactéries. Qu'en est-il de l'arsenic ? Par ailleurs, dans les parties précédentes nous avons vu que la réponse peut être générale ou spécifique en fonction de la nature du stress et celle-ci consiste notamment à réguler l'expression d'un ensemble de processus. Que sait-on de la réponse à l'arsenic chez les bactéries ?

A/ Propriétés générales de l'arsenic

L'arsenic, de symbole As, est un métalloïde qui peut se trouver sous quatre états d'oxydation : (i) l'arsine -III, (ii) l'arsenic 0, (iii) l'arsénite +III et (iv) l'arséniate +V. La distribution de ces différents composés est fonction de l'environnement. Ainsi, dans un milieu fortement réducteur on rencontre l'arsine et l'arsenic natif ; dans un milieu modérément réducteur la forme trivalente (arsénite) prédomine, alors que dans un environnement oxygéné c'est la forme pentavalente (arséniate) qui est la forme majeure. Par ailleurs, ce métalloïde a la capacité de se complexer à d'autres métaux et c'est sous cette forme insoluble qu'il est communément rencontré (Vaughan, 2006).

B/ Toxicité de l'arsenic

L'arsenic constitue un des polluants majeurs des eaux de consommation dans de nombreuses régions du globe et est à l'origine d'un des plus grands empoisonnements collectifs de notre histoire (Oremland & Stolz, 2003).

La toxicité de l'arsenic dépend principalement de sa spéciation c'est à dire de son degré d'oxydation et de sa forme chimique, mais aussi du taux d'absorption et du taux d'élimination de la cellule. L'As est ainsi classé par toxicité décroissante comme suit : Arsine > As^{III} > oxyde de monomethylarsine ($MMAO^{III}$) > diméthyl arsineux glutathion ($DMA^{III}GS$) > DMA^V > MMA^V > As^V > As (Mandal & Suzuki, 2002). L'arsenic est toxique pour tous les organismes vivants quand il est sous forme soluble : arséniate As(V) ou arsénite As(III). Il s'agit également des formes dominantes dans les milieux aqueux. Aussi, ce sont uniquement ces deux formes inorganiques qui seront détaillées dans ce manuscrit.

L'As(V) présente la particularité d'être un analogue structural du phosphate (Fig.16-1). Aussi, son entrée dans la cellule bactérienne se fait par un système de transport transmembranaire du phosphate (Pst ou Pit) (Ordonez *et al.*, 2005). En excès de

phosphate, l'As(V) emprunte le système Pit (Phosphate inorganic transporter) dont l'expression est constitutive et peu spécifique. Toutefois, en présence d'une concentration plus faible de phosphate, la cellule peut discriminer l'entrée de l'As(V) de celle du phosphate, en induisant la synthèse d'un système d'entrée Pst (Phosphate specific transporter) qui est cent fois plus spécifique au phosphate qu'à l'As(V) (Nies & Silver, 1995). L'As(V) inhibe la phosphorylation oxydative en formant un ester d'arséniate d'ADP instable qui court-circuite la production d'énergie. Il perturbe également le métabolisme cellulaire en produisant du glucose-6-arséniate à la place du glucose-6-phosphate. Enfin, il peut aussi remplacer le phosphate dans la molécule d'ADN et empêche ainsi les mécanismes de réparation (Mandal & Suzuki, 2002).

L'As(III) est chimiquement plus réactif que l'As(V). En effet, les composés d'As(III) sont considérés comme 60 à 80 fois plus toxiques que les composés d'As(V) (Villaescusa & Bollinger, 2008). Cette forme trivalente est un équivalent structural inorganique du glycérol (Fig.16-1). Aussi, son transport dans la cellule est rendu possible grâce à des aquaglycéroporines ou un système analogue chez les micro-organismes (Silver & Phung, 2005). Une fois dans la cellule, l'As(III) a une grande affinité pour les groupements thiols. Il peut donc se lier au glutathion (GSH), un tripeptide contenant une cystéine et aux autres protéines contenant des thiols pour former des complexes As(III)-SH. Ainsi, ces complexes As(III)-SH peuvent inhiber l'activité de nombreuses enzymes comme la glutathione réductase, la glutathione peroxydase, la thioredoxine réductase ou la thioredoxine peroxydase (Cf IC). Cette espèce d'arsenic empêche également la respiration en se fixant aux groupements thiols de la pyruvate deshydrogénase et à la 2-oxo-glutarate deshydrogénase. En réalité, l'As(III) inhibe l'activité d'une centaine d'enzymes (Gailer, 2007). Par ailleurs, l'As(III) contribue à la formation d'espèces réactives de l'oxygène (ROS) telles que l'anion superoxyde et le peroxyde d'hydrogène (Liu *et al.*, 2001). Aussi, l'As(III) a un effet mutagène sur l'ADN *via* la production de ROS (Pi *et al.*, 2002). La thèse de l'inactivation de la pyruvate deshydrogenase par l'As(III) est controversée, car mise en concurrence avec l'hypothèse des dommages oxydatifs causés par les ROS générés par l'As(III) (Aposhian & Aposhian, 2006).

C/ Les mécanismes de résistance chez les bactéries

Certaines bactéries ont développé des mécanismes d'adaptation extraordinaires à des milieux contaminés par l'arsenic et il n'est donc pas étonnant que de tels milieux soient

colonisés par des micro-organismes. Parmi ces mécanismes on retrouve (i) la colonisation du milieu sous forme de biofilm, qui permet de séquestrer l'arsenic et (ii) l'acquisition de métabolismes particuliers.

1- Séquestration de l'arsenic

Comme mentionné précédemment (I.B1), il est largement admis que les bactéries qui colonisent une surface sous forme de biofilm sont plus résistantes aux molécules toxiques, notamment à l'arsenic, que celles qui adoptent un mode de vie planctonique (Harrison *et al.*, 2007). Ainsi, certains micro-organismes montrent, en présence d'arsenic, une augmentation de la production de polymères qui peut jouer un rôle dans sa séquestration. Celle-ci a été mise en évidence pour la première fois chez *Desulfotomaculum auripigmentum*. Ainsi, en présence d'As(V) et de soufre, la souche produit des précipités d'As₂S₃ visible en microscopie électronique (Newman *et al.*, 1997).

Cette capacité de synthèse de biofilm et de séquestration peut être mise à profit dans des procédés de bioremédiation. Cette propriété a par exemple été utilisée dans la mise au point de réacteurs destinés à traiter de l'eau de surface contaminée par l'arsenic (Michel *et al.*, 2007).

2- Les métabolismes bactériens

Le métabolisme bactérien joue un rôle majeur dans la biotransformation de l'arsenic c'est à dire dans la mobilisation/immobilisation des différentes espèces d'arsenic et des différents états d'oxydation retrouvés dans un environnement. Ainsi, en absence de bactéries, les réactions d'oxydo-réduction de l'arsenic sont extrêmement lentes même dans des conditions thermodynamiquement favorables. Chez les micro-organismes, les réactions peuvent être utilisées comme mécanisme de détoxification ou comme un apport énergétique.

Les biotransformations de l'arsenic sont : la réduction de l'arséniate y compris lors de la respiration, l'oxydation de l'arsénite et les différentes réactions de méthylation (Paez-Espino *et al.*, 2009) (Fig. 16).

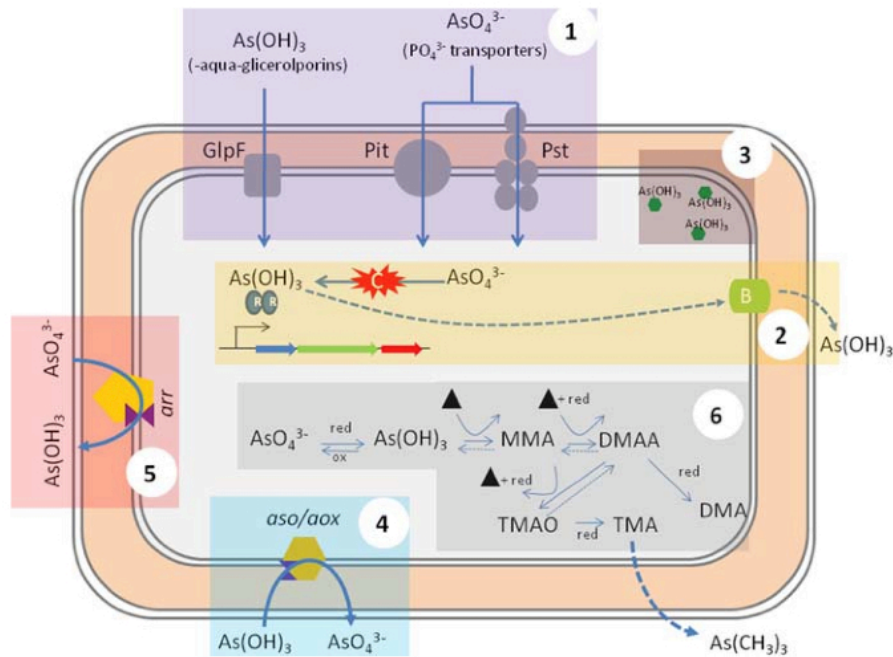


Figure 16 : Diagramme des différents métabolismes bactériens impliqués dans la biotransformation de l'arsenic dans l'environnement (extrait de Paez-Espino *et al.*, 2009).

Les formes inorganiques de l'arsenic peuvent être transformées en formes organiques *via* une cascade de méthylation (Fig.16-6) mais ce processus, chez les bactéries, reste encore non élucidé. Aussi, ce mécanisme ne sera pas détaillé ici puisqu'il n'est pas nécessaire à la compréhension de mon travail.

2.1- Réduction de l'arséniat en arsénite

Deux types de réductions de l'As(V) ont été décrits : la réduction qui a pour but de produire de l'As(III) et de le transporter hors de la cellule, *via* le système Ars (Fig.16-2), et la réduction dissimilatrice qui permet d'utiliser l'As(V) comme accepteur terminal d'électrons, *via* le système *arr* (Fig.16-5). Cette dernière ne sera pas abordée car elle ne fait pas l'objet de mon sujet d'étude.

D'une manière générale, on attribue aux protéines du système Ars les fonctions, de « percevoir » et de « détoxifier » l'arséniat. Les systèmes Ars ont été décrits pour la première fois chez *E. coli* comme des gènes inductibles situés sur des plasmides et présentant une organisation opéronique (Hedges & Baumberg, 1973). Ces gènes sont très

largement répandus dans le monde des micro-organismes, aussi bien les bactéries à Gram positif et à Gram négatif que les eucaryotes (Paez-Espino *et al.*, 2009). Par ailleurs, ces gènes peuvent être localisés sur un plasmide, associés ou non à des transposons (Neyt *et al.*, 1997) et/ou sur un chromosome bactérien (Silver & Phung, 1996).

Chez les micro-organismes, les opérons *ars* sont généralement constitués de trois (*arsRBC*) ou cinq gènes (*arsRDABC*), organisés en une seule unité transcriptionnelle, codant un régulateur ArsR, une pompe d'extrusion de l'As(III), ArsB et une arséniate réductase, ArsC.

Deux gènes additionnels *arsA* et *arsD*, codant respectivement une pompe d'efflux et un second répresseur de l'opéron *ars*, peuvent être retrouvés au niveau de ces opérons chez certains micro-organismes, comme par exemple celui localisé sur le plasmide R733 d'*E. coli* (Chen *et al.*, 1985, Chen *et al.*, 1986).

Un autre gène *arsH* a été retrouvé chez toutes les bactéries à Gram négatif ayant un opéron *ars* (Paez-Espino *et al.*, 2009), notamment *Yersinia enterocolitica* (Neyt *et al.*, 1997) et *Acidithiobacillus ferrooxidans* (Butcher *et al.*, 2000). Le rôle de ArsH est décrit plus loin (d).

Ainsi, suite à l'entrée de l'arséniate dans la cellule, sa réduction en As(III) est catalysée par l'arséniate réductase ArsC. L'As(III) est ensuite excrété hors de la cellule *via* la ou les protéine(s) transmembranaire(s), ArsB/ArsA ou Acr3. ArsC, ArsB et leurs régulateurs jouent un rôle essentiel dans le processus de réduction et d'efflux de l'arsenic, c'est pourquoi ils font l'objet d'une description plus détaillée dans les prochains paragraphes.

a) *ArsC* : la réductase

Il existe deux familles ArsC : l'une homologue à l'enzyme codée par *arsC* du plasmide R773 d'*E. coli*, et l'autre composée des enzymes proche de celle décrite pour le plasmide pI258 de *Staphylococcus aureus*. Dans les deux cas, l'arséniate réductase est une protéine monomérique qui fonctionne avec la chimie des thiols. Toutefois, le mode de couplage énergétique ainsi que leurs propriétés respectives montrent de grande différence.

(i) Réductase à glutarédoxine

La réductase ArsC homologue à R773 contient des résidus cystéines et nécessite une interaction avec le glutathion (GSH) et la glutarédoxine (Grx) pour être active (Shi *et al.*, 1999). Dans ce type de système, la réduction se fait en deux étapes : la première par la glutarédoxine et la deuxième par le glutathion (Rosen, 1999) (Fig. 17). Chez *E. coli* par exemple, trois glutarédoxines (Grx1, Grx2 et Grx3) peuvent servir de réducteur à ArsC même si l'enzyme utilise préférentiellement Grx2 (Shi *et al.*, 1999).

(ii) Réductase à thiorédoxine

La réductase ArsC homologue à pI258, utilise la thiorédoxine (Trx) comme donneur d'électrons (Rosen, 1999). L'enzyme pour être fonctionnelle *in vitro*, nécessite la présence de Trx, de thiorédoxine réductase (TR) et du NADPH (Messens *et al.*, 1999) (Fig. 17).

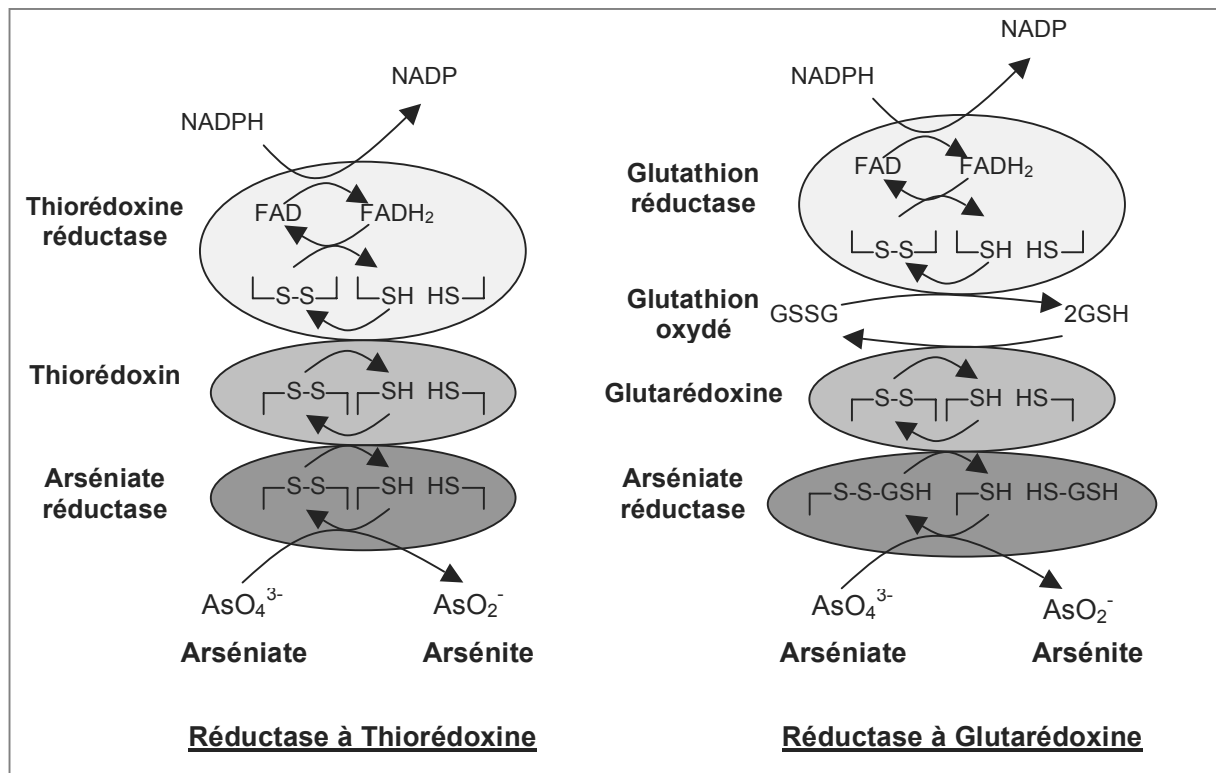


Figure 17 : Les deux familles ArsC (d'après Silver, 2002).

b) *ArsB/ArsA* ou *Acr3* : la pompe d'efflux

L'As(III) obtenu après réduction de l'As(V), est pris en charge par le système d'efflux. Deux familles codant ce système d'efflux ont été identifiées : *ArsB/ArsA* et *Acr3*.

- La famille des **ArsB** appartient à la famille des MDR (Multidrug Resistance-Associated protein) (Cf I.B3). Les gènes *arsB* sont principalement retrouvés chez les γ -protéobactéries et les Firmicutes (Paez-Espino *et al.*, 2009). La protéine *ArsB* est enchâssée dans la membrane interne et catalyse l'extrusion de l'arsénite et de l'antimoine (Rosen, 1999). *ArsB* seul fonctionne comme antiport métalloïde/proton, en utilisant le potentiel de membrane, ce qui permet d'exporter l'As(III) (Dey & Rosen, 1995 ; Meng *et al.*, 2004). Néanmoins, *ArsB* est souvent couplé à *ArsA* qui grâce à son activité ATPasique permet de pomper activement l'arsénite hors de la cellule. Ainsi, chez *E. coli* il a été montré que l'opéron *arsRBC* chromosomique ne confère qu'une résistance modérée à l'arsenic, mais lorsque la protéine *ArsA* produit de l'opéron *arsRDABC* est exprimée dans cette souche, l'efflux est meilleur et la cellule devient résistante à des concentrations plus élevées en arsenic (Dey & Rosen., 1995 ; Silver *et al.*, 2002). Le complexe *ArsB/ArsA* serait donc plus efficace que *ArsB* seul, pour exporter l'As(III) chez *E. coli*. Par conséquent, on peut envisager que ce complexe *ArsB/ArsA* soit issu d'une pression de sélection dans des environnements fortement contaminés par l'arsenic.

Récemment, un gène *arsP* présentant une faible homologie avec les gènes *arsB* déjà connus, a été mis en évidence chez *Campylobacter jejuni*. La protéine *ArsP* serait une perméase de la famille des MFS (Major-Facilitator-Superfamily) (Cf I.B3). *arsP* appartient à l'opéron *ars* et est régulé par *ArsR*. Néanmoins, à ce jour aucune étude n'a démontré l'implication de ce gène dans la résistance de la souche (Wang *et al.*, 2009).

- La famille **Acr3** constitue un second groupe. Les gènes *acr3* sont principalement retrouvés chez les α -protéobactéries et les actinobactéries (Paez-Espino *et al.*, 2009). Il s'agit d'un transporteur spécifique de l'arsénite, qui utilise le potentiel de membrane pour l'excréter hors de la cellule. Ce système est similaire à la pompe d'efflux de l'arsénite *Acr3p*, caractérisée chez *Saccharomyces cerevisiae* (Wysocki *et al.*, 1997).

c) *ArsR* et *ArsD* : les régulateurs de l'opéron *ars*

ArsR est un répresseur, il interfère donc de ce fait avec l'initiation de la transcription (Cf II. B2) (Wu & Rosen, 1993). Le site de fixation de ArsR au niveau de la région opératrice est un motif peu conservé de 12Nt-12Nt-12Nt, où il fonctionne généralement par un mécanisme de régulation de Classe I (Cf II. B2), c'est-à-dire qu'il se fixe au niveau des éléments -10/-35 ou à proximité (Busenlehner *et al.*, 2003). Chez *E. coli* par exemple, ArsR se fixe sous forme de dimère au niveau de la région -64 à -31 (Xu *et al.*, 1996).

En présence d'arsenic, le métalloïde se lie au répresseur et induit un changement de conformation du domaine de liaison à l'ADN, ce qui dissocie le répresseur ArsR du promoteur, autorisant ainsi la transcription de l'opéron *ars*. Chez *E. coli*, le site de fixation de l'arsenic sur ArsR est caractérisé par la séquence consensus suivante : TCAT-N₇-TTTG (Xu *et al.* 1996a ; Xu *et al.*, 1996b). Toutefois, même si les séquences de ArsR sont relativement proches, cette séquence consensus n'est pas retrouvée dans l'ensemble du monde bactérien : c'est notamment le cas des souches de *Pseudomonas* sp. (Patel *et al.*, 2007) ou *Acidithiobacillus ferrooxidans* (Butcher & Rawlings, 2002).

ArsD est un régulateur secondaire de l'expression de l'opéron *ars*, qui fixe la même région opératrice que ArsR. Il présente une affinité moindre à l'As et à l'ADN que ArsR, c'est pourquoi la régulation se fait à des concentrations plus fortes en arsenic (Chen & Rosen, 1997). ArsR est un homodimère qui contrôle le niveau basal de l'expression de l'opéron *ars* tandis que ArsD, lui contrôle le haut niveau de transcription, ceci probablement pour éviter une surexpression de ArsB qui est toxique à de fortes concentrations (Rosen, 1999). De récentes études menées chez *E. coli* ont montré que ArsD serait également un chaperon qui se lie à l'As(III) cytoplasmique et le livrerait directement au site de liaison de l'ATPase ArsA (Yang *et al.*, 2010).

d) *ArsH* : un gène essentiel ?

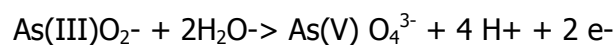
Le gène *arsH* a été décrit pour la première fois chez *Yersinia enterocolitica*, où il est essentiel pour la résistance de la souche (Neyt *et al.*, 1997). Ce gène est retrouvé chez toutes les bactéries à Gram négatif contenant un opéron *ars*, telles *Thiobacillus ferrooxidans* (Butcher *et al.*, 2000), *Pseudomonas putida* (Canovas *et al.*, 2003), *Pseudomonas* sp. TS44 (Cai *et al.*, 2009). Pour certaines de ces souches il a été montré que ce gène n'est pas nécessaire pour la résistance à l'arsenic. Certains auteurs suggèrent

qu'il s'agit d'un régulateur transcriptionnel (Neyt *et al.*, 1997), alors que d'autres ont montré qu'il s'agit d'une NADH-flavine mononucléotide oxydoréductase de 24,6 kDa (Ye *et al.*, 2007). Son rôle dans le système Ars reste donc toujours inconnu.

2.2- L'oxydation de l'arsénite en arséniate

La première bactérie arsénite-oxydante décrite est *Bacillus arsenoxydans* en 1918 (Green, 1918). Depuis, une centaine de procaryotes arsénite-oxydants ont été isolés (Oremland & Stolz, 2003). Ces organismes proviennent d'environnements variés comme les sédiments miniers (Santini *et al.*, 2000 ; Duquesne *et al.*, 2008), les lacs hyper-salins (Oremland *et al.*, 2005), les sols (Osborne & Ehrlich, 1976), les sources géothermales (Salmassi *et al.*, 2002). Elles sont phylogénétiquement diverses, mais appartiennent principalement aux alpha-, beta- et gamma-protéobactéries, au genre *Chloroflexus*, au phylum des Firmicutes et des Actinobactéries et au genre *Thermus*. Il existe également des archées arsénite-oxydantes comme *Sulfolobus acidocaldarius*, *Sulfolobus tokodaii* ou *Aeropyrum pernix* (Santini *et al.*, 2002).

L'oxydation de l'arsénite suit la réaction suivante :



a) Métabolisme hétérotrophe vs chimiolithoautotrophe

L'As(V) obtenu après oxydation de l'As(III) est moins mobile. L'oxydation peut donc conduire à une détoxification du milieu environnant. Il s'agit ici de l'image que l'on avait de l'arsénite oxydase avant la découverte d'organismes chimiolithoautotrophes (Santini *et al.*, 2000).

Aujourd'hui on distingue donc deux types de bactéries arsénite-oxydantes :

- les hétérotrophes, c'est à dire les bactéries qui ont besoin d'un apport en matière organique. Les souches arsénite-oxydantes hétérotrophes sont nombreuses, notamment *Alcaligenes faecalis* (Phillips & Taylor, 1976), *Hydrogenophaga* sp. NT-14 (Hoven *et al.*, 2004), *Agrobacterium tumefaciens* (Kashyap *et al.*, 2006), *Ochromobacterium tritici* SCII24 (Branco *et al.*, 2009).

- les chimiolithoautotrophes, c'est à dire des organismes qui peuvent utiliser l'As(III) comme donneur d'électrons, avec le CO₂/HCO₃⁻ comme source de carbone et l'O₂ ou le NO₃⁻ comme accepteur terminal d'électrons. On peut citer *Rhizobium* sp. NT-26 (Santini *et al.*,

2000), *Thiomonas arsenivorans* (Battaglia-Brunet *et al.*, 2006), *Alkalilimnicola ehrlichei* MLHE-1 qui utilise le NO_3^- comme accepteur terminal d'électrons (Oremland *et al.*, 2002), les souches *Azoarcus* sp. DAO1 et *Sinorhizobium* sp. DAO10 qui couplent l'oxydation de l'arsénite à la dénitrification (Rhine *et al.*, 2005 ; Rhine *et al.*, 2006).

Chez les micro-organismes hétérotrophes et autotrophes pour lesquels l'oxydation de l'arsénite a été étudiée, la réaction est catalysée par l'arsénite oxydase (Fig. 16-4).

b) L'arsénite oxydase

Cette enzyme est codée par les gènes *aoxA* et *aoxB* qui sont responsables de la synthèse de la petite et la grande sous-unité catalytique de l'enzyme, respectivement. Différentes terminologies ont été attribuées aux gènes codant l'arsénite oxydase. Ainsi, on trouve *aoxAB* (arsénite-**oxydase**) chez *A. tumefaciens* par exemple, *asoBA* (**As oxydante**) chez *A. faecalis* et *aroBA* (arsénite-**oxydase**) chez *Rhizobium* sp. NT-26.

(i) Structure de l'enzyme

A. faecalis est la seule bactérie pour laquelle l'arsénite oxydase a été purifiée et caractérisée (Anderson *et al.*, 1992 ; Ellis *et al.*, 2001). Il s'agit d'un hétérodimère $\alpha\beta$ constitué de deux sous-unités : une petite sous-unité Rieske à centre [2Fe-2S], d'environ 134 aa : AoxA (appelée AroB chez d'autres organismes) et une grande sous-unité catalytique d'environ 825 aa, qui contient un centre molybdoptérine et un centre fer soufre de type [3Fe-4S] : AoxB (appelée AroA chez d'autres organismes). Cette architecture la classe dans la famille des diméthylsulfoxydes (DMSO) réductases (Anderson *et al.*, 1992 ; Ellis *et al.*, 2001).

(ii) Localisation de l'enzyme

Chez *A. faecalis*, l'arsénite oxydase est une protéine périplasmique associée à la membrane interne (Ellis *et al.*, 2001). Chez *Hydrogenophaga* NT-14, l'enzyme serait soluble et localisée dans le périplasme (Hoven *et al.*, 2004). La sous-unité Rieske de l'enzyme contient une séquence signal TAT (Twin Arginine Transporter) en N-term, avec un motif RRGLK très fortement conservé. Dans le cytoplasme, les cofacteurs molybdoptérines et le cluster [Fe-S] sont incorporés au polypeptide. La séquence TAT permet ensuite d'adresser la protéine repliée au périplasme ou à la membrane (Silver & Phung, 2005). D'après

Duquesne *et al.*, contrairement aux systèmes TAT classiques, le peptide signal de AoxA ne serait pas clivé mais servirait à ancrer la protéine dans la membrane (Duquesne *et al.*, 2007).

(iii) Fonctionnement de l'enzyme

D'après la structure cristalline de l'arsénite oxydase de *A. faecalis*, les auteurs ont proposé un mécanisme de réaction d'oxydation (Ellis *et al.*, 2001). Ainsi, la grande sous-unité AoxB présente une structure en entonnoir qui permet à l'As(III) d'atteindre le centre catalytique, puis le relargage de l'As(V). Le site de liaison du substrat se situe au niveau des résidus His195, Glu203, Arg419 et His423 de AoxB (Ellis *et al.*, 2001). Au niveau du site actif on trouve quatre interactions Mo-S et une double liaison à l'oxygène, quand l'enzyme est réduite. L'enzyme oxydée, possède une liaison additionnelle Mo-O. L'arsénite en se liant au cofacteur à molybdène Mo(VI), permet donc une attaque nucléophile de la paire d'électrons de la liaison Mo(VI)=O puis le transfert de deux électrons à l'atome de molybdène. Il en résulte une réduction de Mo(VI) en Mo(IV) et la production d'arséniate. Les électrons arrivent ensuite au centre [3Fe-4S] de la même sous-unité, puis au centre Rieske de la petite sous-unité (Silver & Phung, 2005). Enfin, les électrons sont transférés à une protéine de la chaîne respiratoire : un cytochrome ou une azurine par exemple.

Chez certains micro-organismes, un cytochrome de la famille C_{551} a été mis en évidence comme intervenant dans le transfert d'électrons suite à l'oxydation de l'arsenic. Celui-ci est constitué de 20 aa et contient le motif Cys-X2-Cys-His en N-term, qui est caractéristique d'une liaison à un hème (Hoven *et al.*, 2004). Chez *Hydrogenophoga* NT-14, les électrons seraient transmis à un cytochrome de type C_{551} , puis à une cytochrome oxydase aa_3 . Un gradient de protons se formerait alors *via* le complexe aa_3 et pourrait être utilisé pour la synthèse d'ATP (Hoven *et al.*, 2004).

D/ *Herminiimonas arsenicoxydans* : un organisme de référence

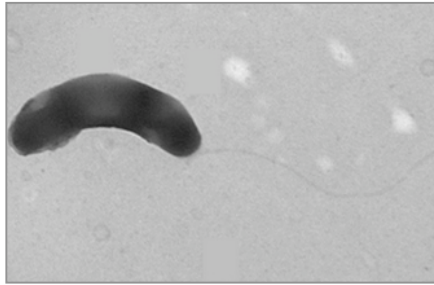


Figure 18 : Photographie en microscopie électronique de *Herminiimonas arsenicoxydans*.

Herminiimonas arsenicoxydans (ULPAs1) (Fig. 18) est une β -protéobactérie hétérotrophe qui a été isolée des boues d'une station d'épuration contaminée par des métaux et l'arsenic. Cette bactérie à Gram négatif est dotée d'un flagelle polaire impliqué dans sa motilité (Muller *et al.*, 2003).

1. Caractéristiques du génome

Son génome a été entièrement séquencé en 2007 par l'équipe en collaboration avec le Genoscope. Ainsi, *H. arsenicoxydans* arbore un chromosome circulaire unique d'environ 3,5 Mb avec 3,294 séquences codantes (CDS) dont 38 % en une fonction inconnue et un contenu en GC (%) modérément riche sur l'ensemble du génome (environ 54%).

L'utilisation d'une banque d'ADN génomique et l'exploration des données du génome ont permis de mettre en évidence la présence d'opérons *ars* multiples chez cet organisme. En effet, quatre loci *ars* ont pu être caractérisés dans la souche ULPAs1, conférant ainsi à cette bactérie une haute résistance à l'arsenic (Muller *et al.*, 2007). Trois de ces loci sont composés de *ArsR*, *ArsB*, une ou deux *ArsC* et un gène additionnel codant une putative flavoprotéine *ArsH* (Muller *et al.*, 2007).

Par ailleurs, l'analyse du génome d'*H. arsenicoxydans* a permis d'identifier un opéron *aox* à proximité d'un opéron *ars* et de gènes dénommés *aoxRS*, formant ainsi un « îlot arsenic ». Un opéron *pst* potentiellement impliqué dans le transport du phosphate, est également retrouvé à proximité de ces derniers (Fig. 19).



Figure 19 : Organisation des gènes impliqués dans l'activité arsénite oxydase chez *Herminiimonas arsenicoxydans*.

L'opéron *aox* est constitué d'un cluster de 4 gènes contigus (*aoxABCD*) : *aoxA* et *aoxB* codent respectivement la petite sous-unité de 174 aa comportant un centre de type Rieske [2Fe-2S] et la grande sous-unité de 826 aa qui contient un centre molybdoptérine et un centre fer soufre de type [3Fe-4S]. Le criblage partiel d'une banque de mutants sur base de l'activité arsénite oxydase, a montré que des mutants d'insertion dans les gènes *aoxA* ou *aoxB* sont dépourvus de toute activité arsénite oxydase (Muller *et al.*, 2007). Une analyse des domaines conservés révèle que la protéine AoxC est composée d'un domaine nitroréductase. Le gène *aoxD* chez *H. arsenicoxydans* possède 35% d'identité de séquence avec le gène codant un cytochrome de type c_{551} chez *Hydrogenophaga* NT-14, suggérant un rôle énergétique de l'enzyme (Muller *et al.*, 2003).

Les gènes codant le système flagellaire sont dispersés dans le génome. Par ailleurs, certains produits de ces gènes sont caractéristiques d'un flagelle polaire (deux clusters avec notamment *flhA/flhB*), alors que d'autres correspondent à une ciliature péritriche (*flhDC* par exemple). Ainsi, les gènes codant le flagelle polaire de cette bactérie sont organisés en une cascade mixte polaire/péritriche (Cf. IIC).

2. Caractéristiques physiologiques et métaboliques

H. arsenicoxydans est capable de résister à l'arsenic par le mécanisme majeur de réduction de l'arséniate mais aussi par oxydation de l'arsénite. Ainsi, elle résiste à 5 mM d'As(III) et à 100 mM d'As(V).

Elle possède non seulement une haute résistance à l'arsenic mais également un chimiotactisme positif vis-à-vis de ce métalloïde la rendant apte à se déplacer vers des milieux contaminés. En effet, une augmentation de la motilité a été mise en évidence après exposition à l'As(III) (Muller *et al.*, 2007).

De plus, par sa capacité à séquestrer cet élément toxique au sein d'une matrice de polymères, la bactérie participe à la détoxification de l'environnement en rendant l'As

beaucoup moins disponible (Muller *et al.*, 2007). En effet, l'observation des cellules en microscopie électronique à transmission, combinée à la spectrométrie à rayons X à dispersion d'énergie (TEM/EDS), montre la présence de granules riches en arsenic piégés dans une gangue de polymères (Fig. 20).

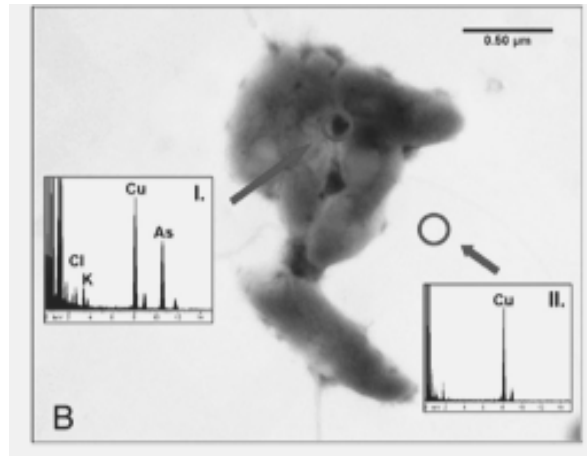


Figure 20 : Synthèse de polymères en réponse à l'arsenic chez *Herminiimonas arsenicoxydans*.

Les ronds représentent les points d'analyse aux rayons X. I. et II. correspondent aux spectres d'analyse obtenus, avec Cl et K, qui représentent les constituants organiques et Cu un constituant de la grille. Le contenu en As est de 16,5% en I. et de 0% en II. (Muller *et al.*, 2007).

L'accès aux données de séquences et à la physiologie de cette bactérie ont coïncidé avec le début de mes travaux de thèse. Cet organisme a donc été défini comme organisme de référence pour l'étude du stress arsénié.

IV.

OBJECTIFS

Le répertoire de micro-organismes colonisant des environnements contaminés par l'arsenic est régulièrement étoffé. L'étendue de la diversité et de la complexité des métabolismes de l'arsenic de ces micro-organismes est aussi grandissante. Toutefois, les mécanismes d'absorption, de métabolisation et la toxicité des différentes formes de l'arsenic ne sont pas totalement élucidés.

Ainsi, il est connu que l'oxydation biologique de l'As(III) en As(V) est catalysée par l'arsénite oxydase. Cependant, *Alcaligenes faecalis* est la seule bactérie pour laquelle cette enzyme a été purifiée et caractérisée. A ce jour, en effet, la physiologie, l'enzymologie, la génétique de l'oxydation de l'arsenic et les processus de régulation restent largement méconnus.

Le sujet d'étude de ma thèse vise à décortiquer la réponse à l'arsenic dans le temps chez les bactéries, en particulier la régulation de l'arsénite oxydase. Une connaissance plus approfondie de ces processus devrait permettre une meilleure compréhension du rôle de l'arsénite oxydase dans des environnements fortement contaminés et aussi d'étendre les connaissances du stress arsénié chez les micro-organismes arsénite-oxydants d'un point de vue physiologique, génétique et moléculaire.

Dans ce but, une approche génétique a été mise en œuvre pour déterminer les régulateurs potentiels de l'*aox*, par criblage de mutants sur base de leur activité arsénite oxydase. Les mutants ont ensuite été analysés par des techniques de biologie moléculaire (q-RT-PCR, immunodétection) afin de déterminer plus précisément le rôle de ces régulateurs (**Chapitre 1**).

En parallèle, une approche protéomique couplée à une approche transcriptomique a été menée chez *Herminiimonas arsenicoxydans*, afin d'identifier les transcrits accumulés et les protéines synthétisées en présence d'arsenic et ainsi de caractériser la réponse adaptative de la cellule bactérienne au stress arsénié (**Chapitre 2**). Ces approches ont ensuite été réitérées pour différents temps d'induction à l'arsenic afin de caractériser la cinétique de la réponse à l'arsenic (**Chapitre 3**).

Les travaux effectués lors de ma thèse ont porté majoritairement sur *H. arsenicoxydans*. Toutefois, les processus d'oxydation de l'arsénite concernent d'autres micro-organismes. Nous nous sommes dès lors interrogés sur les différentes capacités d'adaptation des bactéries à un environnement contaminé par l'arsenic.

Le **chapitre 4** est ainsi consacré à l'analyse comparative des génomes de différentes souches de *Thiomonas*. Cette approche a été menée afin d'identifier le pool de gènes permettant à ces souches de survivre et de coloniser un environnement contaminé.

Enfin, l'analyse du stress arsénié a été étendue à un autre micro-organisme : une bactérie chimiolithoautotrophe *Rhizobium* sp. NT-26. Une approche de génomique descriptive, protéomique et génétique a été menée chez ce micro-organisme afin de caractériser la réponse à l'arsenic chez une bactérie chimiolithoautotrophe (**Chapitre 5**).

Ces travaux personnels sont présentés dans les prochains chapitres et font l'objet de publications parues (Chapitre 1, Chapitre 2 et Chapitre 4), soumise (Chapitre 3) ou en cours de rédaction (Chapitre 5). Ce manuscrit est articulé autour de ces publications scientifiques. Il est complété par l'article consacré aux bactéries du genre *Thiomonas* paru dans le journal du CNRS ('Des bactéries modifient leur génome pour survivre à l'arsenic' **annexe 1**) et le métagénome de Carnoulès (**annexe 2**, en cours de rédaction) qui sont présentés en annexes.

Régulation de l'activité arsénite oxydase
chez *Herminiimonas arsenicoxydans*

A/ Introduction

Les bactéries jouent un rôle important dans le cycle biogéochimique de l'arsenic par leur impact sur la spéciation et la mobilisation de cet élément. Elles résistent majoritairement par les mécanismes de réduction de l'arséniate et d'oxydation de l'arsénite. La réduction de l'arséniate est un processus qui a été largement étudié. L'oxydation de l'arsénite, catalysée par l'arsénite oxydase a quant à elle fait l'objet de peu d'études. Aussi, l'objectif de ce premier chapitre a été de décortiquer les mécanismes moléculaires de régulation de l'activité arsénite oxydase.

Herminiimonas arsenicoxydans est une bactérie oxydant l'arsenic qui a été isolée à partir d'une station d'épuration fortement contaminée par les métaux lourds et les métalloïdes, notamment l'arsenic. Cette bactérie a été choisie comme organisme de référence pour ces études, car la disponibilité des données de séquences coïncide avec mon arrivée au laboratoire. De plus, des études antérieures sur cet organisme ont pu montrer qu'il possède non seulement une haute résistance à l'arsenic mais également un chimiotactisme positif vis-à-vis de ce métalloïde, le rendant apte à se déplacer vers des milieux contaminés. Cette bactérie est capable de séquestrer cet élément toxique, au sein d'une matrice de polymères, elle participe donc activement à la détoxification de l'environnement en rendant l'As moins disponible (Muller *et al.*, 2007). Enfin, la disponibilité d'outils génétiques fait de cette bactérie un outil particulièrement intéressant pour l'étude du métabolisme de l'arsenic et de sa régulation.

L'approche méthodologique qui a été retenue est celle de la construction d'une banque de mutants par insertion aléatoire d'un transposon. En effet, les techniques classiques de microbiologie restent indispensables pour décortiquer des réseaux de régulation. La mutagenèse par transposition est une technique génétique couramment utilisée pour la création *in vitro* ou *in vivo* de phénotypes mutants. Ce type de mutagenèse permet une transposition aléatoire dans le génome d'un organisme. En s'insérant dans un locus, le transposon inactive généralement le gène et le mutant obtenu est déficient dans la fonction codée par ce gène. Le criblage des mutants a été effectué sur base de la perte d'activité arsénite oxydase par un test phénotypique au nitrate d'argent. Cette approche a permis d'obtenir des phénotypes mutants clairement observables. Les mutants ont ensuite été analysés par des techniques de biologie moléculaire (q-RT-PCR, immunodétection) afin

de déterminer plus précisément le rôle de ces différents régulateurs. Ces données ont été complétées par une analyse globale des transcrits après une induction de 8h à l'As(III).

Ma participation à cette étude a porté sur l'identification de certains mutants, la spéciation de l'arsenic et les tests de motilité des différents mutants, la détermination de leur rôle possible dans la régulation post-transcriptionnelle ou traductionnelle par Western-blot, l'analyse des données de transcriptome, l'interprétation des résultats, la rédaction du manuscrit et la réalisation des figures.

L'ensemble de ces travaux a conduit à une publication (Koechler *et al.*, 2010), présentée ci-dessous.

B/ Manuscrit

RESEARCH ARTICLE

Open Access

Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*

Sandrine Koechler^{1†}, Jessica Cleiss-Arnold^{1†}, Caroline Proux², Odile Sismeiro², Marie-Agnès Dillies², Florence Goulhen-Chollet¹, Florence Hommais³, Didier Lièvre¹, Florence Arsène-Ploetze¹, Jean-Yves Coppée², Philippe N Bertin^{1*}

Abstract

Background: Both the speciation and toxicity of arsenic are affected by bacterial transformations, i.e. oxidation, reduction or methylation. These transformations have a major impact on environmental contamination and more particularly on arsenic contamination of drinking water. *Herminiimonas arsenicoxydans* has been isolated from an arsenic-contaminated environment and has developed various mechanisms for coping with arsenic, including the oxidation of As(III) to As(V) as a detoxification mechanism.

Results: In the present study, a differential transcriptome analysis was used to identify genes, including arsenite oxidase encoding genes, involved in the response of *H. arsenicoxydans* to As(III). To get insight into the molecular mechanisms of this enzyme activity, a Tn5 transposon mutagenesis was performed. Transposon insertions resulting in a lack of arsenite oxidase activity disrupted *aoxR* and *aoxS* genes, showing that the *aox* operon transcription is regulated by the AoxRS two-component system. Remarkably, transposon insertions were also identified in *rpoN* coding for the alternative N sigma factor (σ^{54}) of RNA polymerase and in *dnaJ* coding for the Hsp70 co-chaperone. Western blotting with anti-AoxB antibodies and quantitative RT-PCR experiments allowed us to demonstrate that the *rpoN* and *dnaJ* gene products are involved in the control of arsenite oxidase gene expression. Finally, the transcriptional start site of the *aoxAB* operon was determined using rapid amplification of cDNA ends (RACE) and a putative -12/-24 σ^{54} -dependent promoter motif was identified upstream of *aoxAB* coding sequences.

Conclusion: These results reveal the existence of novel molecular regulatory processes governing arsenite oxidase expression in *H. arsenicoxydans*. These data are summarized in a model that functionally integrates arsenite oxidation in the adaptive response to As(III) in this microorganism.

Background

Arsenic is present in various environments, released from either anthropogenic or natural sources. This element is toxic for living organisms and known to be a human carcinogen [1]. Its toxicological effects depend, at least in part, on its oxidation state and its chemical forms, inorganic species being considered as more toxic [2]. The contamination of drinking water by the two inorganic forms, arsenite As(III) and arsenate As(V), has been reported in different parts of the world [3] and constitutes a major threat of public health. Microorganisms are known to take part in the transformation, i.e.

oxidation, reduction or methylation of the metalloid, having a deep impact on arsenic contamination in environment.

Several bacteria and prokaryotes have developed adaptation, resistance and colonization mechanisms, which allow them to live in hostile arsenic contaminated environments. *H. arsenicoxydans* is a Gram-negative β -proteobacterium isolated from an industrial activated sludge plant and exhibiting a remarkable set of arsenic resistance determinants [4]. The *H. arsenicoxydans* adaptive response to arsenic is organized in a complex and sophisticated network. In particular, differential proteome studies have recently demonstrated the synthesis of several proteins encoded by the three *ars* resistance operons, e.g. arsenate reductase ArsC, flavoprotein ArsH and regulator ArsR [5,6] and the induction of oxidative

* Correspondence: philippe.bertin@unistra.fr

† Contributed equally

¹UMR7156 Génétique Moléculaire, Génomique et Microbiologie, CNRS Université de Strasbourg, 28 rue Goethe, 67000 Strasbourg, France

stress protein encoding genes, e.g. catalase (*katA*), superoxide dismutase (*sodB*) and alkyl hydroperoxide reductase (*ahpC*) [7].

One of the most noticeable response to arsenic in *H. arsenicoxydans* is the ability of this bacterium to oxidize As(III) to As(V), a less toxic and less mobile form, via an arsenite oxidase activity. The two genes coding for this heterodimeric enzyme are organized in an operonic structure, and have been named *aoxA* and *aoxB* for the small and the large subunit, respectively [6,8,9]. Homologous genes have been since identified in various microorganisms [6,10-13]. In *Agrobacterium tumefaciens*, a complex transcriptional regulation has been recently suggested, involving As(III) sensing, two-component signal transduction by an AoxS sensor kinase and an AoxR regulator, and quorum sensing [14]. Nevertheless, the molecular mechanisms involved in the control of arsenite oxidase expression remain largely unknown.

To extend our view of arsenic response [7], in particular with respect to the arsenic oxidation process, we conducted a comprehensive transcriptomic, genetic and molecular analysis of *H. arsenicoxydans* following exposure to As(III). These approaches allowed us to identify major determinants involved in the control of arsenite oxidation.

Results

Gene expression profiling in response to arsenic

The response to As(III) was analyzed in exponentially growing cells using microarrays. The data from three biological replicates were combined after normalization and statistical analysis carried out using the R software and packages <http://www.r-project.org>. The set of genes was further refined to include only those genes that showed a valid *p*-value and whose expression was altered by a factor of 2 or more when compared to the level measured in the absence of arsenic.

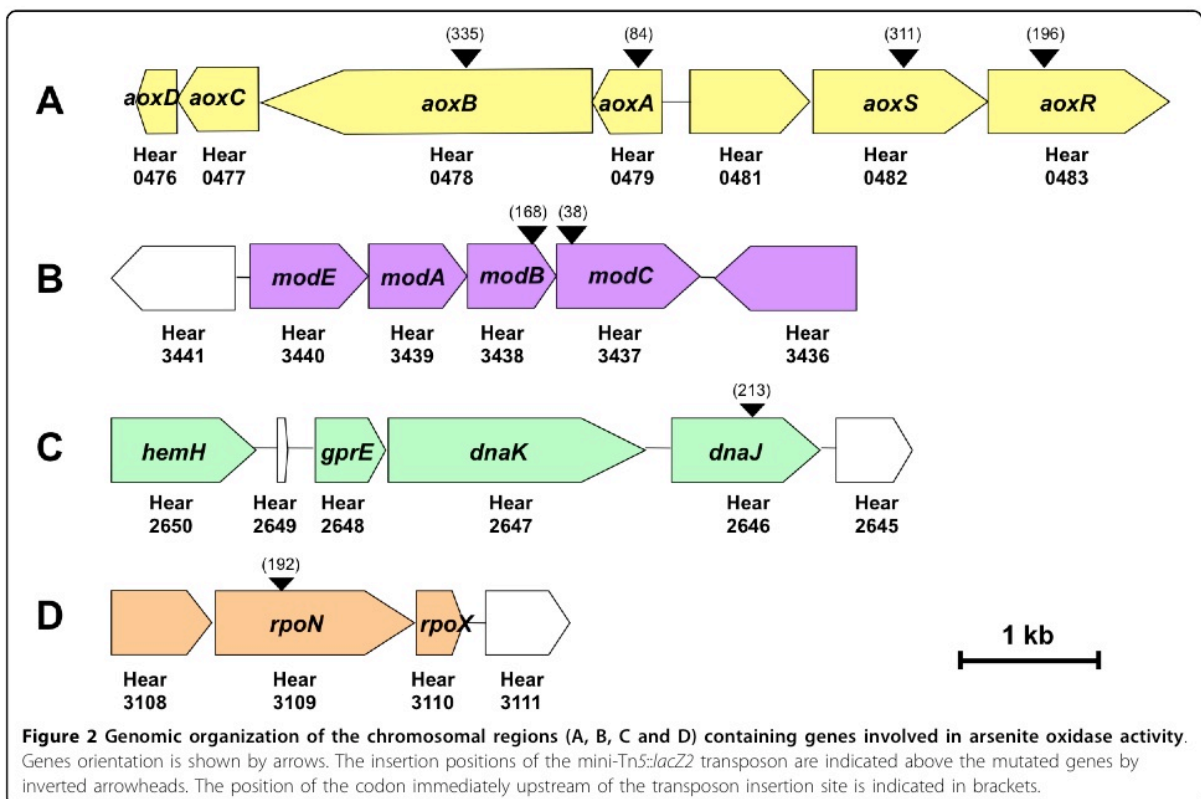
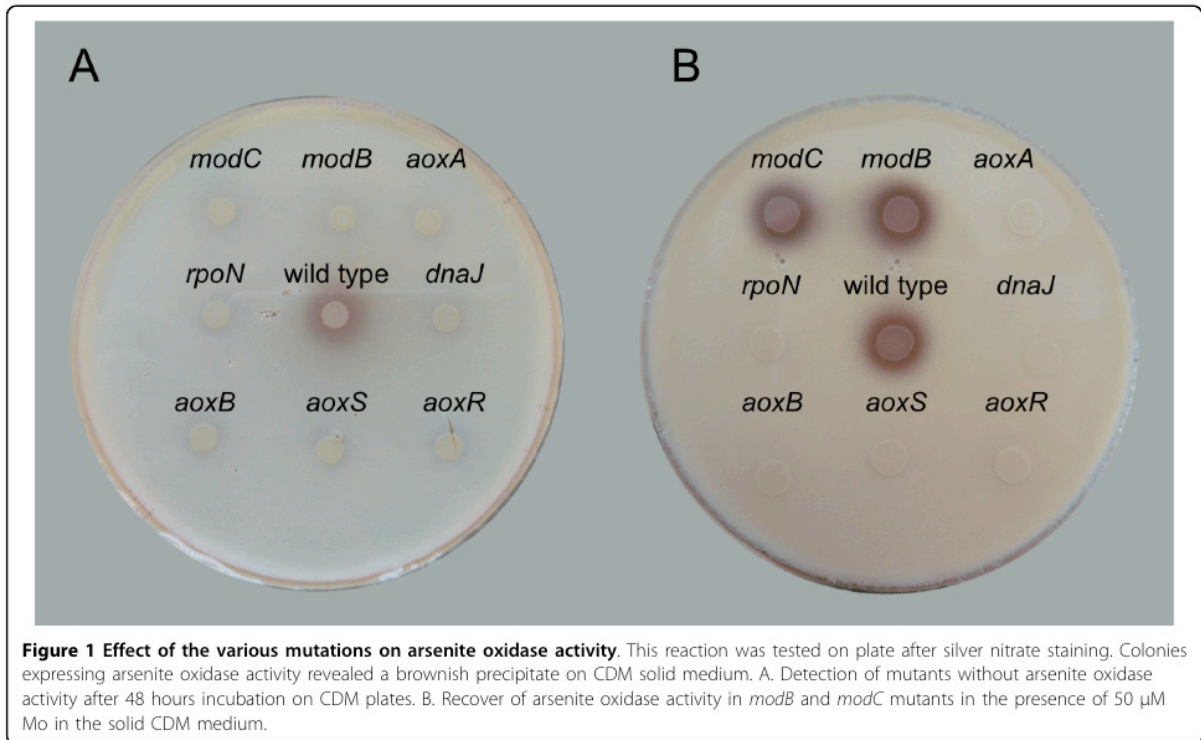
This experiment led to the identification of 293 genes showing an arsenic-induced expression change (> 2 fold ($\log_2 = 1$)). Among these genes, 133 (45%) were up-regulated and the remaining part, i.e. 160 genes, was down-regulated. The relative changes in gene expression ranged from a 11-fold down-regulation (*rpsN* gene encoding a ribosomal protein) to a 126-fold up-regulation (putative gene involved in phosphate transport). A list of those genes is shown in Additional file 1, Table S1. The corresponding HEAR gene numbers are available in the Arsenoscope relational database <http://www.genoscope.cns.fr/agc/mage/arsenoscopevia> the MaGe web interface [15].

The 293 genes differentially expressed were clustered according to the function of the corresponding encoded proteins. Among the 133 genes that were induced by at

least a 2-fold factor, about 11% are involved in metabolism, 26% in transport and binding protein, 26% in cellular processes and 31% have no assigned function. The high percentage of genes with unknown function is in accordance with the proportion of unknown function proteins identified in the genome of *H. arsenicoxydans* [6,7]. In agreement with our previous results, genes involved in arsenic resistance, phosphate transport and flagellar biosynthesis were induced in the presence of As(III) (see Additional file 1, Table S1), further supporting the relationship between these physiological processes [6,7]. Interestingly, only one methyl-accepting chemotaxis protein (MCP) gene was induced in our microarray data, suggesting a role for this protein in the sensing of arsenic. This mechanism is currently under investigation. Genes encoding the putative nitroreductase AoxC and the cytochrome *c*₅₅₂ precursor AoxD as well as the response regulator AoxRS were found to be induced by As(III) (see Additional file 1, Table S1). AoxR has been proposed to act as a positive regulator of the *aox* operon upon phosphorylation by AoxS in *A. tumefaciens* [14]. Our transcriptomic data suggest that the regulation machinery is, at least in part, similar in *H. arsenicoxydans*. Furthermore, genes coding for the arsenite oxidase AoxAB subunits were found to be among the most induced genes in the presence of As(III). However, even though these results extended our knowledge of the arsenic response in *H. arsenicoxydans*, they did not lead to a better understanding of the molecular mechanisms involved in the control of arsenite oxidation. This prompted us to perform a transposon mutagenesis experiment.

Identification of arsenite oxidase accessory genes by screening an Aox activity deficient mutant library

To identify genes possibly involved in the control of arsenite oxidation in *H. arsenicoxydans*, a library of 10,000 kanamycin resistant mutants was constructed by transposon mutagenesis, as previously described [9]. These clones were tested by silver nitrate staining [16] for arsenate production on As(III)-supplemented CDM agar plates. As compared to the wild-type strain, whose arsenite oxidase activity was revealed by a brownish precipitate, 10 mutants with a lack of As(III) oxidase activity were obtained. These strains showed no precipitate (Figure 1A), as observed for the M1 and M2 strains used as negative controls. Indeed, these strains carry a mutation in *aoxA* or *aoxB* genes coding for the small and the large subunit of arsenite oxidase, respectively [9]. Genes disrupted by transposon insertions were identified in these 10 new mutants. As expected, four of the 10 mutants showed insertions in the *aoxAB* operon (Figure 2A). More interestingly, six mutants carried a transposon insertion outside the *aoxAB* operon. Two



mutants were found to be affected in the *aoxRS* two-component signal transduction system (mutants Ha482 and Ha483, respectively) located upstream of the *aoxAB* operon in *H. arsenicoxydans* [6] (Figure 2A). These results further support our transcriptomic data suggesting that these two genes play a role in arsenic response. Two transposon insertions were shown to disrupt genes of the *modEABC* operon coding for a molybdenum high-affinity transport system [17], i.e. *modC* encoding an ATP-binding cassette transport protein (mutant Ha3437) and *modB* encoding a molybdenum transport system permease (mutant Ha3438) (Figure 2B). Remarkably, transposon insertions were also located in *dnaJ* encoding a heat shock protein (Hsp40), (mutant Ha2646) (Figure 2C) and in *rpoN* encoding the alternative nitrogen sigma factor (sigma 54) of RNA polymerase (mutant Ha3109) (Figure 2D).

Two additional experiments were performed to complete the physiological characterization of these mutants with respect to arsenite oxidation. First, arsenic species were quantified by HPLC-ICP-AES on filtered culture supernatants. *H. arsenicoxydans* was grown in liquid medium supplemented with 1.33 mM arsenite and showed 100% transformation of As(III) into As(V) after 48 h, whereas M1 (*aoxA*) and M2 (*aoxB*) mutants used as controls were not able to transform As(III) into As(V). The same loss of arsenite oxidase activity was measured in Ha482 (*aoxS*), Ha483 (*aoxR*), Ha2646 (*dnaJ*) and Ha3109 (*rpoN*) mutants. In contrast to the results obtained on agar plates, Ha3437 (*modC*) and Ha3438 (*modB*) strains showed 100% transformation of arsenite (Table 1, Figure 1A). Previous studies have demonstrated that the bioavailability of metals or trace elements considerably varies according to the type of matrix used for microbial growth [18]. We therefore assumed that Mo might be partly sequestered on CDM agar medium, resulting in a lack of arsenite oxidase

activity on plate. To test this hypothesis, As(III) oxidase tests were performed on CDM agar plates supplemented with 50 µM Mo. The addition of Mo to the solid medium restored As(III) oxidase activity in both Ha3437 (*modC*) and Ha3438 (*modB*) mutants while it had no effect on other mutant strains (Figure 1B).

Second, we have previously demonstrated that the polar flagellum-dependent motility of *H. arsenicoxydans* is increased in the presence of As(III), suggesting that arsenite oxidation may result in a gain of energy [6]. The motility of mutant strains was therefore tested on plates containing different concentrations of As(III), i. e. 0.66 mM, 1.33 mM and 2 mM. The diameter of the swarming rings was measured after 72 h. As shown in Figure 3, the disruption of *aoxA*, *aoxB*, *aoxR*, *aoxS* or *rpoN* genes abolished the improvement of swarming performances in the presence of As(III). Unlike those mutants, a disruption in *dnaJ* completely abolished the motility of *H. arsenicoxydans* in the presence or the absence of As(III). DnaJ is known to be essential for the expression of the *flhDC* flagellar master operon in *Escherichia coli* [19]. The lack of motility observed in the *dnaJ* mutant suggests the existence of a similar *flhDC*-dependent regulation of flagellar genes in *H. arsenicoxydans*. More importantly, the *dnaJ* mutant was affected in both the motility and arsenite oxidation, suggesting that these two mechanisms are co-regulated.

Effect of AoxR, AoxS, RpoN and DnaJ on arsenite oxidase synthesis

To get further insight into the involvement of AoxR, AoxS, RpoN and DnaJ in arsenite oxidase activity, Western immunoblotting experiments were performed using antibodies raised against AoxB. The abundance of this protein was evaluated from total protein extracts of *H. arsenicoxydans* wild-type and mutant strains grown in the presence or not of As(III). AoxB was detected as a single band corresponding to a molecular mass of 92 kDa in As(III)-challenged *H. arsenicoxydans* strain (Figure 4). This single band was not observed in the various mutant strains. Furthermore, arsenite oxidase activity on native gel was only detected in As(III)-challenged wild type total extract (data not shown). Taken together these results suggest that the lack of activity in the mutant strains is due to the absence of AoxB protein, which may result from an effect of AoxR, AoxS, RpoN and DnaJ on *aoxAB* expression.

Effect of AoxR, AoxS, RpoN and DnaJ on the control of arsenite oxidase operon expression

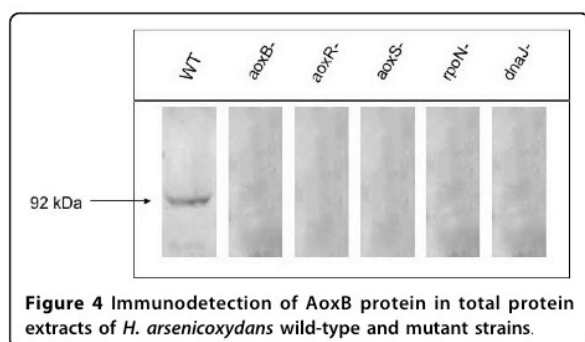
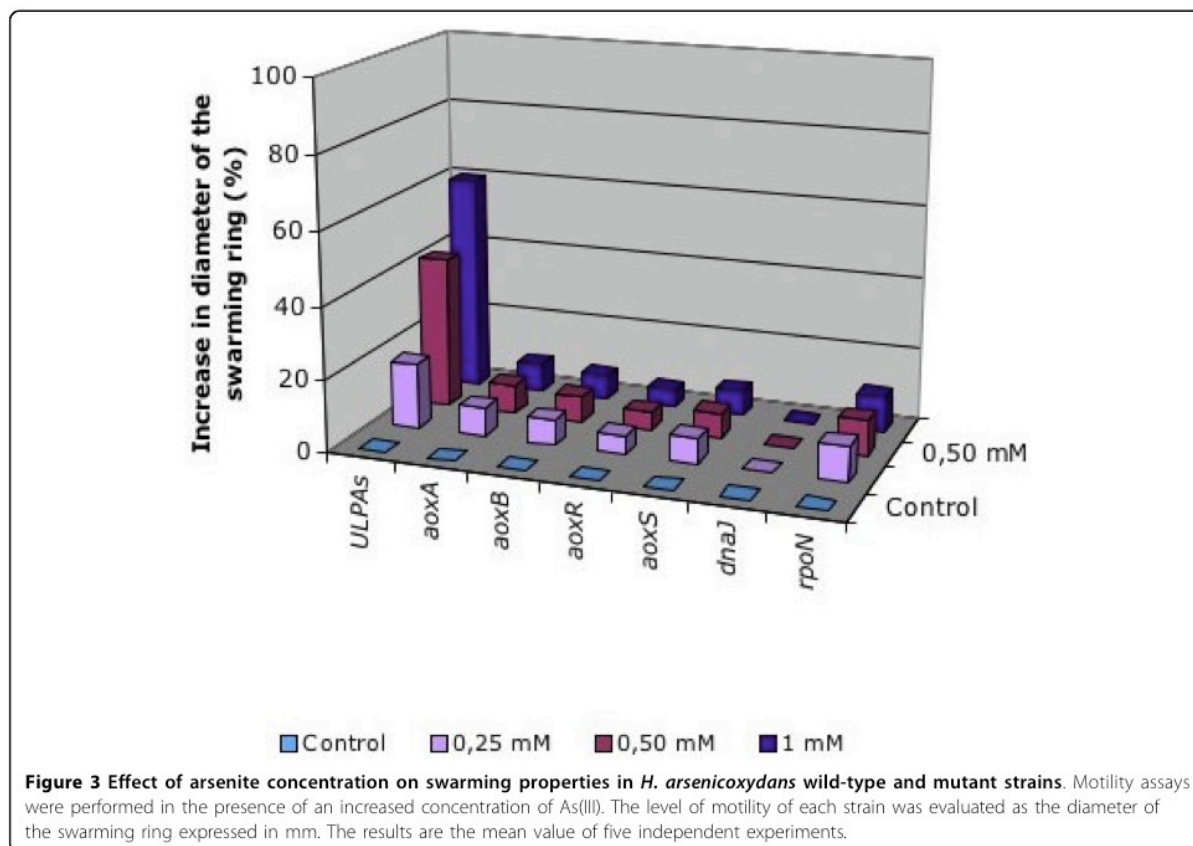
To determine the involvement of *aoxR*, *aoxS*, *dnaJ* and *rpoN* on *aoxAB* transcription, we performed quantitative RT-PCR experiments. For each strain, changes in *aoxB* transcript abundance were compared to two internal

Table 1 Determination of arsenic speciation in *H. arsenicoxydans* wild-type and mutant strains.

Strain	Mutated gene	Arsenic species identified ^a	
		As(III)	As(V)
ULPAs1	/	-	+
M1 ^b	<i>aoxA</i>	+	-
M2 ^b	<i>aoxB</i>	+	-
Ha482	<i>aoxS</i>	+	-
Ha483	<i>aoxR</i>	+	-
Ha2646	<i>dnaJ</i>	+	-
Ha3109	<i>rpoN</i>	+	-
Ha3437	<i>modC</i>	-	+
Ha3438	<i>modB</i>	-	+

^a Determined by HPLC-ICP-AES after 48 h growth in CDM medium containing 100 mg/liter of arsenite.

^b [9]



controls, i.e. the putative RNA methyltransferase gene and the peptide deformylase gene, in cultures challenged or not by As(III). The expression of *aoxB* mRNA was increased by a 9.4 fold factor after As(III) exposure in the *H. arsenicoxydans* wild-type strain. In contrast, *aoxB* expression was not increased in Ha482 (*aoxS*), Ha483 (*aoxR*), Ha3109 (*rpoN*) and Ha2646 (*dnaJ*) mutant strains, suggesting that the corresponding proteins play a crucial role in *aoxAB* operon expression (Table 2).

Table 2 *aoxB* relative expression in *H. arsenicoxydans* wild-type and mutant strains.

Strain	<i>aoxB</i> expression ratio +As(III)/-As(III)	Standard error
Wild type	9.406	0.630
Ha3109 (<i>rpoN</i>)	0.250	0.060
Ha483 (<i>aoxR</i>)	0.111	0.024
Ha482 (<i>aoxS</i>)	0.200	0.029
Ha2646 (<i>dnaJ</i>)	1.156	0.289

Expression ratios of *aoxB* in *H. arsenicoxydans* wild-type and mutant strains without As(III) versus an As(III) 8 hours induction (1.33 mM), as measured by quantitative RT-PCR. Expression of each gene was normalized to the expression of the two housekeeping genes HEAR0118 and HEAR2922 coding for a peptide deformylase and a putative RNA methyltransferase, respectively. Standard errors were calculated from the data of two quantitative PCR replicates obtained from two independent biological replicates. Bold text indicates statistically significant induction.

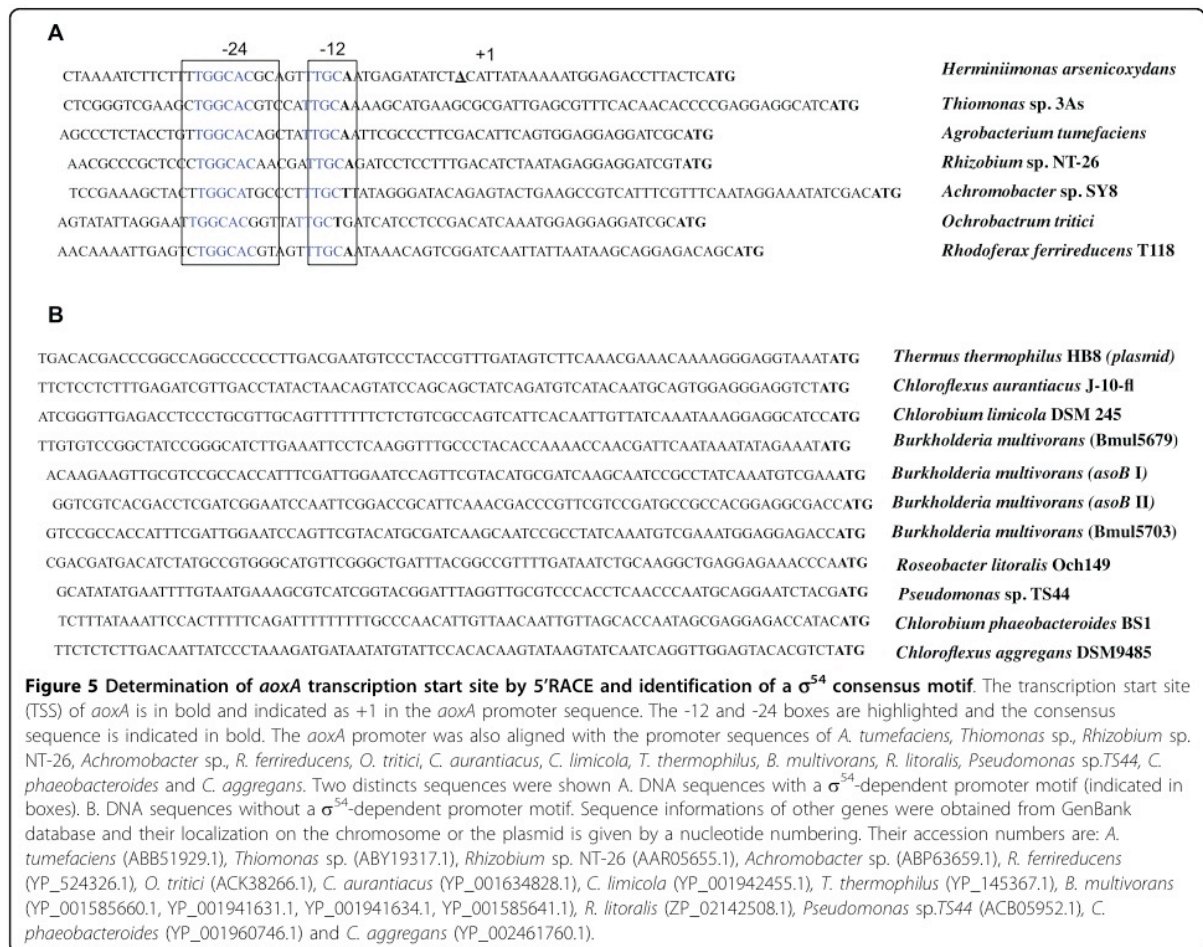
Molecular mechanisms of arsenite oxidase transcription

The *aoxR* and *aoxS* genes encode a two-component system while *rpoN* encodes a sigma factor which recognizes a particular promoter with a specific -12/-24 binding site. These three proteins may therefore play a role in the initiation of *aoxAB* transcription. To get further insight into the molecular interactions between those regulators and the *aoxAB* promoter, we mapped the

transcriptional start site of this operon by the amplification of *aoxAB* cDNA ends and 5'RACE. Messenger RNAs were extracted from induced (1.33 mM As(III)) and non induced *H. arsenicoxydans* wild-type strain cultures. A single transcriptional start site was identified from induced cells at -26 bp relative to the translation start codon, while no transcriptional start site was identified from non induced cells. In agreement with this, a **TGGCACGCAGTTTGC** putative -12/-24 σ^{54} -dependent promoter motif was identified upstream of the *aoxAB* transcriptional start site (Figure 5). In addition, multiple alignment of *aoxAB* promoter sequences present in databases revealed a similarity to promoters recognized by σ^{54} in *A. tumefaciens*, *Thiomonas* sp., *Rhizobium* sp. NT-26, *Achromobacter* sp., *Rhodoferrax ferrireducens*, *Ochrobactrum tritici* (Figure 5A). In contrast, no such σ^{54} -dependent promoter motif was found in several strains containing the *aoxAB* operon but lacking the two-component transduction system *aoxRS* operon, such as *Chloroflexus aurantiacus*, *Chlorobium limicola*,

Thermus thermophilus, *Burkholderia multivorans*, *Roseobacter litoralis*, *Pseudomonas* sp.TS44, *Chlorobium phaeobacteroides* and *Chloroflexus aggregans* (Figure 5B).

Remarkably, a multiple alignment of amino acid sequences revealed that AoxR shares significant homology with a number of σ^{54} RNA polymerase transcriptional activators, i.e. 35.96% identity with ZraR and 35.26% identity with AtoC from *E. coli* K12. AoxR contains three conserved domains shared by most Enhancer Binding Proteins (EBP), namely a N-terminal response regulator receiver domain (amino acids 18-130), a central σ^{54} interaction domain (amino acids 147-368) common to all σ^{54} dependent EBPs (Pfam E-value 10^{-116} ; <http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00158>) and a C-terminal DNA binding helix-turn-helix (HTF) domain (amino acids 421-463) enable to bind to specific upstream activation sequences [20]. AoxR shares similarities with several EBPs of σ^{54} essential for the formation of an open complex formation during σ^{54} -dependent transcriptional initiation, in particular the σ^{54}



activator sequence GAFTGA loop 1 which directly binds to σ^{54} conserved region III (Figure 6) [21]. Taken together, these observations strongly suggest that AoxR interacts directly with RpoN to initiate the transcription of *aoxAB* operon in *H. arsenicoxydans*.

Discussion

Despite many works devoted to arsenic metabolism in microorganisms, little is known about the regulation of arsenite oxidase activity. In the present study, the combination of transcriptomic, genetic and molecular data provided a comprehensive view of the role of various proteins in the control of arsenite oxidation in *H. arsenicoxydans* (Figure 7). We showed that some proteins play an indirect role in this process, as their presence is not essential for AoxAB synthesis. In this respect, the AoxB large subunit contains a Mo site required in arsenite oxidase enzymatic activity [22].

Ha3437 (*modC*) and Ha3438 (*modB*) mutations were located in the molybdenum high-affinity transport system operon, which further support the key role of this element in enzyme activity. In addition, the recovery of As(III) oxidase activity in these two mutants in the presence of an excess molybdenum suggests that Mo may also be transported through an alternative uptake system in *mod* mutants, e.g. a low-affinity uptake system involving non specific permeases such as HEAR0069, HEAR0154, HEAR1749 or HEAR2391 or a sulfate transport system, as described in *E. coli mod* mutants [23].

More importantly, our results suggest that AoxR and RpoN constitute a transcriptional complex that play a major role in the initiation of *aoxAB* operon transcription. Three mutants, i.e. Ha482 (*aoxS*), Ha483 (*aoxR*) and Ha3109 (*rpoN*), were affected in this process. The amino acids sequence analysis of *H. arsenicoxydans* AoxR and AoxS revealed the existence in these proteins

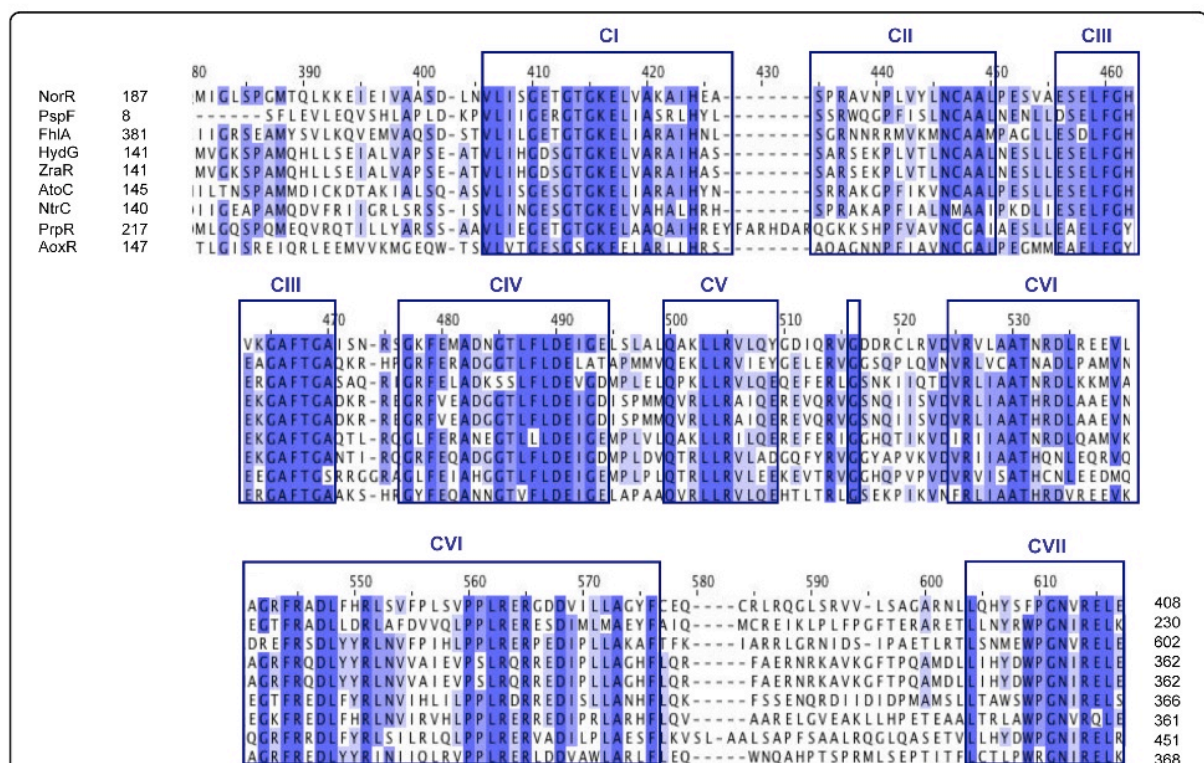
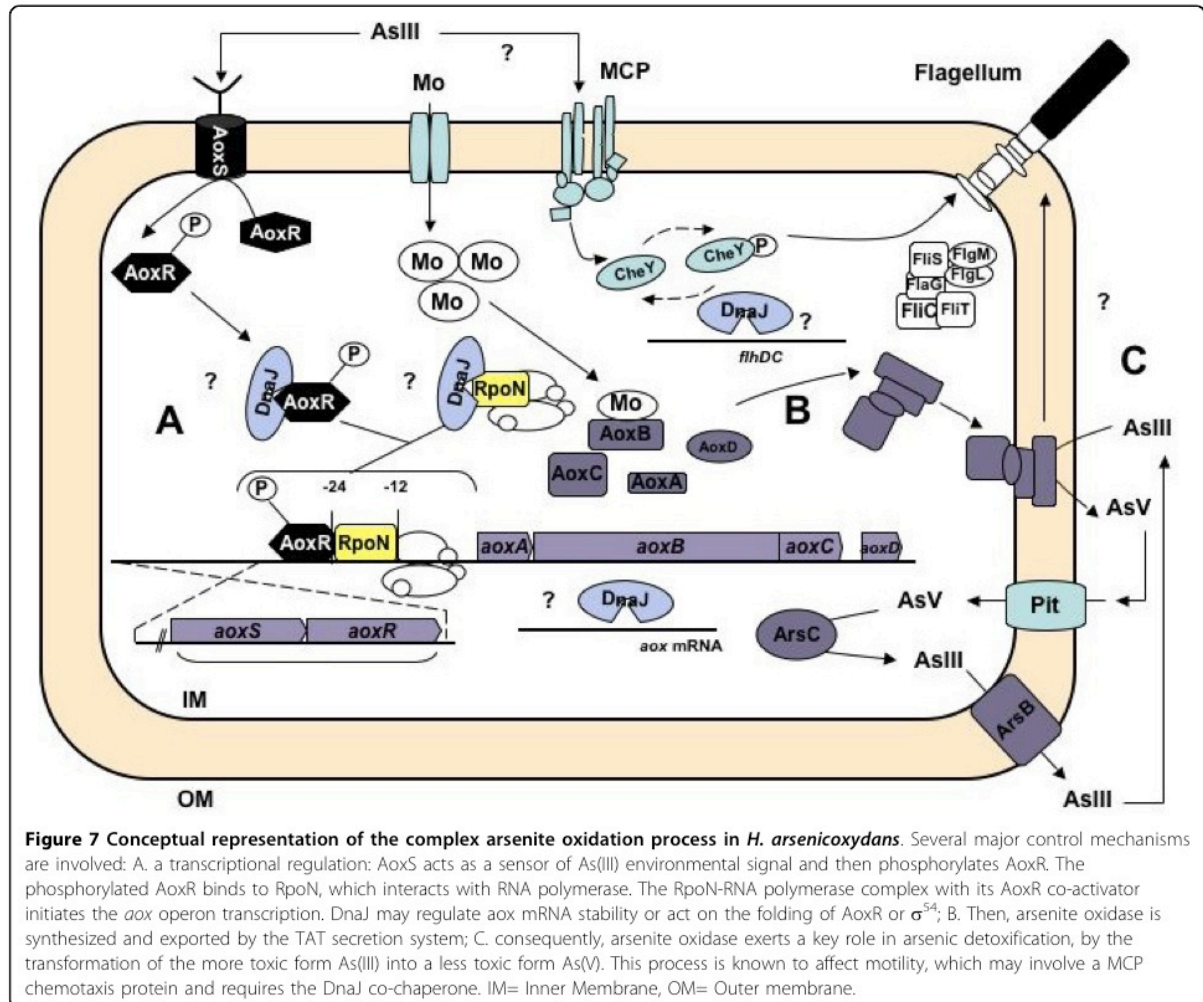


Figure 6 Amino acids conservation between σ^{54} Enhancer Binding Proteins (EBP) and AoxR. Sequence alignment was performed with ClustalW. The conserved amino acids are presented with a blue background and the blue intensity reflects sequence similarities. Only the central binding domain is indicated. Region CI has remarkable similarity to the consensus glycine-rich flexible loop motif (Walker A - consensus motif GxxGxGK), and also contains hydrophobic residues. Region CII is hydrophobic. The region CIII is predicted to fold into two alpha helices separated by a turn. This region is involved in a specific interaction between the EBP and the σ^{54} required for open promoter complex formation via the GAFTGA motif. Region CIV is rich in glycine, negatively charged and contains a consensus sequence of 4 aliphatic residues followed by 2 negatively charged residues (Walker B - consensus motif TVFLDE); in contrast, CVI is positively charged and is rich in aromatic residues and proline. Region CV is found about 80 amino acids away from region CI, and has a consensus sequence QakLLRVLq. Finally, region CVII has a core of eight highly conserved amino acids. Sequence informations of other genes were obtained from Colibri database (Institut Pasteur, Paris).



of structural features common to partners of two-component signal transduction systems, which are composed of a sensor kinase and a response regulator [24]. Moreover, the comparison of AoxS and AoxR protein sequences with those of *A. tumefaciens* revealed similarities. Indeed, the AoxS protein sequence contains short blocks of conserved motifs that are consistent with a role of sensor histidine kinase, e.g. the "H" (amino acids 279 to 287: LAHEVNNPL), the "G2" (amino acids 435 to 441: GRIGLGL) and the "N" (amino acids 380 to 391: VRQIVLNLVLNA) domains. In addition, four highly invariant residues playing a central role in phosphorylation correspond to Asp9, Asp10, Asp57 and Lys107 in the *H. arsenicoxydans* AoxR protein. The three asparagine residues are known to interact together to form an "acid pocket", characteristic of orthodox receivers, into which the side chain of Lys109 protrudes [24]. Finally, a transmembrane region and a 17 amino acids residue cluster possibly exposed to the periplasm are present in

AoxS and could serve as a signal receptor in the presence of As(III) in the medium. The detection of As(III) would then lead to AoxS autophosphorylation at a histidine residue via ATP hydrolysis and phosphotransfer to an aspartate residue in the response regulator AoxR, as recently proposed in *A. tumefaciens* [14].

Remarkably, our results demonstrated for the first time that the alternative N sigma factor (σ^{54}) is essential for the initiation of arsenite oxidase transcription. Indeed, a mutation in the corresponding gene led to a complete loss of As(III) oxidation and *aoxB* transcription in Ha3109 (*rpoN*). σ^{54} is one of the alternative sigma subunits of RNA polymerase responsible for specific binding to DNA. The core RNA polymerase complexed with σ^{54} is usually associated with nitrogen assimilation and fixation, but is also known to play a role in various physiological processes, e.g. flagellar synthesis, carbon source utilization or bacterial virulence [25]. To date, only one report has shown that σ^{54}

participates in the transcription of genes possibly involved in metal tolerance, i.e. the *zraR/S* genes that code for a zinc and lead responsive two-component regulatory system in *E. coli* [26]. RNA polymerase together with σ^{54} binds to a specific promoter site, with the consensus DNA sequence YTGGCACGNNNN TTGCWNNw [27], forming a transcriptionally inactive closed complex. Such a characteristic -12/-24 σ^{54} -dependent promoter motif, i.e. TGGCACG-CAGTTTGC, was identified 26 bp upstream of the transcriptional initiation codon of *aoxAB* with respect to the +1 transcriptional start site (Figure 5), which confirmed the need for RpoN in the initiation of *aoxAB* transcription. Changes in the conformation of σ^{54} -RNA polymerase are nucleotide dependent. Indeed, the DNA melting step absolutely requires the interaction with a transcriptional activator protein. Most of these σ^{54} -dependent activators share three domains found in AoxR, i.e. a C-terminal DNA binding domain that binds to upstream activation sequences, a conserved central domain belonging to the AAA+ (ATPases associated with various cellular activities) protein family to proceed with initiation of transcription and a N-terminal receiver domain that regulates its own AAA+ domain [20,28,29]. A multiple alignment of the central domain revealed a conservation of a common architecture between AoxR and σ^{54} EBPs. Indeed, seven highly conserved sequence motifs corresponding to a σ^{54} interaction domain of AoxR further support the direct interaction of AoxR with RpoN to stimulate the transcription of *aoxAB* operon in *H. arsenicoxydans* (Figure 6). This central σ^{54} interaction domain has been already used to identify new σ^{54} EBPs [30-37]. It contains the highly conserved signature sequence GAFTGA loop (conserved region CIII, Figure 6) proved to be required for direct interaction with σ^{54} [21,38,39], the Walker A (conserved region I, Figure 6) and B (conserved region CIV, Figure 6) ATPase motifs necessary for the open complex formation.

Both multiple sequences alignment and genetic environment analysis of *aox* promoters suggest the existence of a wide diversity in the transcriptional control of the *aox* operon. Indeed, as in *H. arsenicoxydans*, a σ^{54} -dependent promoter signature was identified in bacteria possessing a two-component transduction system AoxRS operon downstream of the *aoxAB* operon, e.g. *A. tumefaciens* and *O. tritici* (Figure 5A). In contrast, no σ^{54} -dependent promoter motif and no *aoxR* homologous gene were found in other bacteria, e.g. *C. aurantiacus* or *C. aggregans* (Figure 5B). These observations suggest that the transcription of the *aox* operon in these bacteria may involve other regulatory proteins and that AoxR may represent a specific co-activator of RpoN in the initiation of the *aox* operon transcription. Finally,

our results provide evidence that the DnaJ co-chaperone is required for As(III) oxidation. DnaJ is part of the DnaK-DnaJ-GrpE Hsp70 machinery. Hsp70 chaperones represent one of the most potent defence cellular mechanism against environmental insults as DnaK-DnaJ-GrpE are known to assist protein folding [40,41] or to be involved in mRNA stability [42]. In the present study we showed that there is no induction of *aoxAB* transcription in the *dnaJ* mutant, resulting in a loss of AoxAB synthesis.

Several possible mechanisms involving DnaJ in the regulation of arsenite oxidase can be hypothesized. DnaJ may be required for the proper folding or activity of the AoxR regulator. Such a function has been demonstrated for the positive regulator CRP in a *dnaJ* deletion mutant in *E. coli* [43]. Similarly, a post-transcriptional regulation of the arsenite oxidase itself can not be excluded. Moreover, a Tat (Twin-Arginine Translocation) signal has been detected in the AoxA sequence of *H. arsenicoxydans* [6]. Proteins secreted to the periplasm via a Tat protein export pathway are known to require a folding by Hsp70 chaperones before their secretion. DnaJ could be one of these chaperones [44,45]. Another possible target of DnaJ may be the RpoN sigma factor, as this chaperone has been demonstrated to play a role in the regulation of σ^5 in various species [46]. Alternatively, several mechanisms are known to be involved in the stability of messenger RNA. For example, in *E. coli*, a long 5' untranslated region (UTR) has been observed upstream of the transcriptional start site of the *flhDC* flagellum master operon. This region plays a crucial role in the stability of the mRNA controlled by CsrA [19]. In the present report, the *aoxAB* transcriptional start site was located 26 bp upstream of the translational start codon, providing evidence that such a long 5'UTR does not exist upstream of the *aox* operon. However, this does not rule out a role of DnaJ on the stability of *aoxAB* mRNA, as HSP proteins are known to function as mRNA stabilizers and to protect them against nuclease degradation [42]. Further experiments will therefore be required to fully elucidate the molecular mechanisms of arsenite oxidase regulation in *H. arsenicoxydans*.

Conclusion

Taken together, our observations provide evidence that multiple proteins play a role in the control of arsenite oxidation in *H. arsenicoxydans*. The following regulatory model is proposed: AoxS responds to the presence of As(III) in the environment and autophosphorylates. The phosphate is then transferred to AoxR, which acts as a positive regulator of the *aox* operon and activates the initiation of the transcription in association with RpoN. In addition, DnaJ acts on the expression or the stability of both arsenite oxidation and motility genes,

demonstrating that these two functions are strongly linked. Our results include the role of RpoN and DnaJ in arsenite oxidase synthesis, which provide further insight into the molecular mechanisms used by *H. arsenicoxydans* to cope with the most toxic form of arsenic in its environment.

Methods

Bacterial strains and growth media

Bacterial strains used in this study are listed in Table 3. *H. arsenicoxydans* ULPAs1 was grown in a chemically defined medium (CDM), supplemented by 2% agar when required [4]. *Escherichia coli* S17-1 strain [47] was cultivated in LB medium (MP Biochemicals). Matings were performed on CDM to which 10% (wt/vol) LB medium was added, as previously described [9]. Tryptone swarm plates containing CDM supplemented with 1% Bacto-Tryptone and 0.25% agar were used to assess bacterial motility.

Transposon mutagenesis

The mini-Tn5::lacZ2 transposon [47] was delivered by mobilization of the suicide vector pUT/mini-Tn5::lacZ2 from *E. coli* S17-1 (λ -pyr) to *H. arsenicoxydans*. Conjugation was performed and transformants were selected as previously described [9].

Selection of arsenite oxidase mutants

Mutants were screened for arsenite oxidase activity as previously described [9]. Agar plates were flooded with a 0.1 M AgNO₃ solution to visualize arsenite oxidation [16]. Mutants affected in molybdenum metabolism were also tested on CDM agar plates supplemented with 50 μ M Na₂MoO₄, 2H₂O and 1.33 mM As(III).

DNA manipulation and insertion mapping

DNA manipulations were carried out according to standard protocols, as described by Sambrook et al. [48].

Total DNA was isolated from mutant strains with the Wizard Genomic DNA purification kit (Promega). Transposon insertion sites were mapped as previously described [9]. Briefly, total DNA was digested and self-ligated. This ligation mixture was used as a template for PCR amplification using mini-transposon specific primers. The PCR products obtained were purified with a PCR purification kit (Qiagen) and sequenced on an Applied Biosystems ABI prism 3130xl capillary sequencer. The resulting sequences were compared to the *H. arsenicoxydans* genome sequence [6] to identify disrupted CDS. Finally, insertion sites and transposon orientations were precisely mapped by sequencing PCR products obtained with two primers hybridizing upstream and downstream, respectively, of the insertion site of each disrupted gene (see Additional file 2, Table S2).

Arsenic speciation determination

H. arsenicoxydans wild type and mutants were grown for 48 hours in CDM medium supplemented with 1.33 mM As(III). Culture supernatants were filtered through sterile 0,22 μ m pore size filters (VWR). Arsenic species were separated by high-performance liquid chromatography (HPLC) and quantified by inductively coupled plasma-atomic emission spectrometry (ICP-AES), as previously described [9].

RNA extraction

Strains were grown at 25°C for 24 h (OD = 0,15) and cultures were induced by addition of 0.66 mM or 1.33 mM As(III) for 8 hours before extraction. Samples were harvested and stored at -80°C. RNA was extracted as previously described [7]. After extraction procedure, RNA integrity was checked by electrophoregram analysis on a BioAnalyser (Agilent) and total RNA concentration was determined spectrophotometrically with a Nanodrop.

Microarrays and data analysis

Microarrays containing 60-mer oligonucleotides for all predicted *H. arsenicoxydans* genes <http://www.genoscope.cns.fr/agc/mage/arsenoscope> were used, as previously described [7]. Briefly, total RNA (5 μ g) was reverse transcribed and indirectly labelled according to manufacturer's instructions with some modifications [7]. The quality and concentration determination as well as hybridization and scanning were performed as previously described [7] Three distinct biological RNA samples were prepared from in each growth condition (with and without As(III) induction) and labelled either by Cy3 or Cy5 in a dye-swap design. Microarray data were deposited in ArrayExpress (accession E-MEXP-2199 and A-MEXP-1594). Data normalization and statistical

Table 3 Bacterial strains used in this study.

Name	Characteristics	Reference
<i>Escherichia coli</i>		
S17-1 (-pyr)	pUT/miniTn5::lacZ2	De Lorenzo et al., 1990
<i>Herminiimonas arsenicoxydans</i>		
ULPAs1	Wild type	Weeger et al., 1999
M1	<i>aoxA</i> ::Tn5lacZ2	Muller et al., 2003
M2	<i>aoxB</i> ::Tn5lacZ2	Muller et al., 2003
Ha482	<i>aoxS</i> ::Tn5lacZ2	This work
Ha483	<i>aoxR</i> ::Tn5lacZ2	This work
Ha3437	<i>modC</i> ::Tn5lacZ2	This work
Ha3438	<i>modB</i> ::Tn5lacZ2	This work
Ha2646	<i>dnaJ</i> ::Tn5lacZ2	This work
Ha3109	<i>rpoN</i> ::Tn5lacZ2	This work

analysis were performed as previously described [7]. Briefly, data were acquired and analyzed by Genepix Pro 6.0 (Axon Instrument). The experiment design included three biological replicates. For each of them, induced and non-induced cells were compared in dye swap experiments. The resulting arrays were analyzed using the R software <http://www.r-project.org>. A slide by slide Loess normalization was performed using the limma package [49]. Valid log₂ expression ratios from replicated spots were averaged on each array so as to get statistically independent ratios for each oligonucleotide included in the array design. For the same reason, dye swap arrays were also averaged. Oligos that had no valid expression ratios on the ten arrays were excluded from the data set for further analysis, which was carried out using the varmixt package and the VM option [50]. The resulting raw *p*-values were adjusted according to a Benjamini and Yekutieli procedure [51]. Genes showing a valid *p*-value and a more than two-fold decreased or increased expression were considered as differentially expressed between the two conditions and were retained for further study.

Quantitative real time PCR

Quantitative PCR experiments were performed with RNA prepared as described for microarrays. RNA aliquots were purified with the RNeasy Plus mini kit (Qiagen) to ensure the elimination of genomic DNA. Total RNA concentration was determined spectrophotometrically using a Nanodrop and RNA integrity was electrophoretically verified. Total RNA (1,9 µg) was reverse transcribed with SuperScript III first-strand synthesis system for RT-PCR (Invitrogen) using random hexamers. Real time quantitative PCR was carried out with a MyiQ single-color Real-time PCR detection system. The reaction mixture contained 12,5 µl of MESA Blue qPCR MasterMix Plus for SYBR Assay with fluorescein (Eurogentec), 5 µl of cDNA and 300 nM of each primer in a total volume of 25 µl. Thermocycling conditions were as follow: 5 min at 95°C and 40 cycles of 15 s at 95°C, 15 s at 61°C and 1 min at 72°C. The PCR efficiency of the genes of interest and internal control genes were optimized to be similar enough by adjusting the primer concentrations to 300 nM each (data not shown). For each quantitative PCR run, non-template controls were performed to identify false positives and negative controls without reverse transcriptase were performed for each cDNA synthesis reaction and verified in real time PCR to determine the presence of contaminating genomic DNA. Two biological replicates (independent cultures) and two quantitative PCR replicates were performed for each experience. Amplification products were designed to be less than 175 bp in size. The pairs of primers used are listed in Additional file 2, Table S2.

Two housekeeping genes, i.e. HEAR2922 coding for a putative RNA methyltransferase and HEAR0118 coding for a peptide deformylase, were used as standards to obtain normalized *aoxB* (HEAR0478) gene ratio [52] in the As(III) induced sample compared to the non-induced sample. These two housekeeping genes showed a stable expression between the two analyzed conditions (without As(III) and after an 8 hours As(III) exposure) when observing the microarrays data. The data were analyzed with the Relative Expression Software Tool [53]. Statistical significance was defined as a *p*-value of ≤ 0.05.

5'RACE experiment

The transcriptional start site of *aoxAB* operon was determined using the 5'RACE system for rapid amplification of cDNA ends (Invitrogen). Total RNA was obtained as described before. Reverse transcriptase reactions were performed using 5 µg total RNA and a gene specific reverse primer (5'-CATGGGCACTT-GAATGTCTTG-3'). Reactions were heated at 70°C for 10 min and immediately prewarmed at 50°C before addition of Super-Script II reverse transcriptase. Reverse transcription was conducted at 50°C for 50 min and stopped at 70°C for 15 min. Purification and tailing of cDNA were performed according to manufacturer's instructions. The resulting cDNA was amplified by PCR using the provided Abridged Anchor Primer and a gene specific primer (5'-ATGCTGTGCGGACGATATCG-3') located upstream of the original cDNA primer.

Preparation of protein extracts, SDS-PAGE and PAGE separation

Western immunoblotting were performed from late exponential phase wild-type and mutant strains grown in 1 liter CDM (with and without the presence of 100 mg/liter As(III)). The cultures were harvested by centrifugation for 10 min at 9,000 × *g*. Cell pellets were resuspended in distilled water and sonicated at 100 A (15 times 1 min with 1 min interval on ice, 80% duty cycle). Cell debris were removed by centrifugation (15 min at 13,000 × *g*). The supernatant was collected (total extract) and stored at -20°C. The protein concentration of each sample was measured with a Bio-Rad protein assay kit. First, fifty micrograms of each protein extract was loaded onto an 11% polyacrylamide-SDS gel. Second, fifty micrograms of each protein extract were loaded onto a polyacrylamide gel (native gel). The assay of arsenite oxidase activity followed the transfer of reducing equivalents from arsenite to 2,4-dichlorophenolindophenol (DCIP) as described by Anderson *et al.* [54]. Briefly, the reduction of DCIP (60 µM) was monitored in the presence of 200 µM sodium arsenite in 50 µM MES, pH 6.0, at 25°C.

Preparation of antibodies and Western blot analysis

Monoclonal antibodies raised against an AoxB peptide were obtained from Proteogenix. Briefly, a hexadecapeptide with the SKNRDRVALPPVNAQK sequence was synthesized. This peptide corresponds to the N-terminal 16 amino acids of the arsenite oxidase large subunit of *H. arsenicoxydans*. The peptide was then coupled to keyhole limpet haemocyanin (KLH). Two rabbits were injected at multiple subcutaneous sites with peptide-KLH at 14 days intervals. Animals were prebled at day 0, bled at day 49 (from an ear vein) and totally bled at day 90. Antibodies were partially purified on an affinity column substituted with the peptide.

After SDS-PAGE electrophoresis, the proteins were electrotransferred to a nitrocellulose membrane (Schleicher and Schuell, BA-85) using a Trans-Blot system (Bio-Rad) at 100 V, 4°C for 1 h. The membranes were washed twice in Tris buffered saline (TBS: 10 mM Tris-HCl pH7.5, 150 mM NaCl) and blocked in TBS with 0,3% bovine serum albumin (BSA). The membrane was then washed three times in Tris-buffered saline with TritonX100 and Tween 20 (TBS-T: 20 mM TrisHCl pH7.5, 500 mM NaCl, 0,2% Triton X-100, 0,05% Tween20), and incubated for 1 h with the AoxB antisera (1:800 dilution) in TBS-T with 0,3% BSA. Excess antibodies were removed by repeated washing with TBS-T. After 30 min incubation in TBS-T containing the secondary antibody (1:800 dilution of goat IgG against rabbit IgG, Sigma) conjugated with alkaline phosphatase, the membrane was washed twice with TBS-T and revealed by NBT/BCIP color reagent using standard procedures.

Additional file 1: Supplemental table S1. Selected genes differentially expressed after 8 hours arsenite stress.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2180-10-53-S1.PDF>]

Additional file 2: Supplemental table S2. Oligonucleotides used in the study. A. Identification of transposon insertion sites in *H. arsenicoxydans* mutants. B. Quantitative RT-PCR.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2180-10-53-S2.PDF>]

Acknowledgements

JCA was supported by a grant from the French Ministry of Education and Research. Financial support came from the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (ANR 07-BLAN-0118 project) and the Université de Strasbourg. This work was done in the frame of the Groupement de Recherche (GDR2909-CNRS): « Métabolisme de l'Arsenic chez les Micro-organismes ».

Author details

¹UMR7156 Génétique Moléculaire, Génomique et Microbiologie, CNRS Université de Strasbourg, 28 rue Goethe, 67000 Strasbourg, France. ²Plateforme technologique Puces à ADN, Institut Pasteur, 28 rue du Dr. Roux,

75724 Paris cedex 15, France. ³UMR5240 Microbiologie, Adaptation et Pathogénie, CNRS Université Lyon 1, Bâtiment André Lwoff, 10 rue Dubois, 69622 Villeurbanne cedex France.

Authors' contributions

SK and JCA wrote the manuscript and performed the genetic experiments. SK carried out the quantitative PCR and 5' RACE experiments. JCA performed the Western immunoblotting analysis. CP, OS, MAD and JYC conceived and performed the transcriptomic experiments and the data analyses. DL performed the chemical experiments. FGC, FH, FAP and PB helped to analyze the data and critically revised the manuscript. PB coordinated and conceived the study. All authors read and approved the final manuscript.

Received: 26 October 2009

Accepted: 18 February 2010 Published: 18 February 2010

References

1. Mead MN: **Arsenic: In search of an antidote to a global poison.** *Environ Health Perspect* 2005, **113**:A378-A386.
2. Rosen BP: **Biochemistry of arsenic detoxification.** *FEBS Lett* 2002, **529**:86-92.
3. Smith AH, Lingas EO, Rahman M: **Contamination of drinking-water by arsenic in Bangladesh: A public health emergency.** *Bull World Health Organ* 2000, **78**:1093-1103.
4. Muller D, Simeonova DD, Riegel P, Mangenot S, Koechler S, Lièvrement D, Bertin PN, Lett MC: ***Herminiimonas arsenicoxydans* sp. nov., a metalloresistant bacterium.** *Int J Syst Evol Microbiol* 2006, **56**:1765-1769.
5. Carapito C, Muller D, Turlin E, Koechler S, Danchin A, Van Dorsselaer A, Leize-Wagner E, Bertin PN, Lett MC: **Identification of genes and proteins involved in the pleiotropic response to arsenic stress in *Caenibacter arsenoxydans*, a metalloresistant beta-proteobacterium with an unsequenced genome.** *Biochimie* 2006, **88**:595-606.
6. Muller D, Medigue C, Koechler S, Barbe V, Barakat M, Talla E, Bonnefoy V, Krin E, Arsene-Ploetze F, Carapito C, et al: **A tale of two oxidation states: bacterial colonization of arsenic-rich environments.** *PLoS genetics* 2007, **3**(4):e53.
7. Weiss S, Carapito C, Cleiss J, Koechler S, Turlin E, Coppee JY, Heymann M, Kugler V, Stauffert M, Cruveiller S, et al: **Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data.** *Biochimie* 2009, **91**:192-203.
8. Alvarez-Martinez CE, Lourenço RF, Baldini RL, Laub MT, Gomes SL: **The ECF sigma factor σ^T is involved in osmotic and oxidative stress responses in *Caulobacter crescentus*.** *Mol Microbiol* 2007, **66**:1240-1255.
9. Muller D, Lièvrement D, Simeonova DD, Hubert JC, Lett MC: **Arsenite oxidase *aox* genes from a metal-resistant beta-proteobacterium.** *J Bacteriol* 2003, **185**:135-141.
10. Santini JM, Hoven Vanden RN: **Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26.** *J Bacteriol* 2004, **186**:1614-1619.
11. Quéménéur M, Heinrich-Salmeron A, Muller D, Lièvrement D, Jauzein M, Bertin PN, Garrido F, Joulain C: **Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-oxidizing bacteria.** *App Environ Microbiol* 2008, **74**:4567-4573.
12. Cai L, Rensing C, Li X, Wang G: **Novel gene clusters involved in arsenite oxidation and resistance in two arsenite oxidizers: *Achromobacter* sp. SY8 and *Pseudomonas* sp. TS44.** *App Microbiol Biotechnol* 2009, **83**(4):715-25.
13. Clingenpeel SR, D'Imperio S, Oduro H, Druschel GK, McDermott TR: **Cloning and in situ expression studies of the *Hydrogenobaculum* arsenite oxidase genes.** *App Environ Microbiol* 2009, **75**:3365-3365.
14. Kashyap DR, Botero LM, Franck WL, Hassett DJ, McDermott TR: **Complex regulation of arsenite oxidation in *Agrobacterium tumefaciens*.** *J Bacteriol* 2006, **188**:1081-1088.
15. Valletet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Médigue C: **MaGe: A microbial genome annotation system supported by synteny results.** *Nucleic Acids Res* 2006, **34**:53-65.
16. Lett M-C, Paknikar K, Lièvrement D: **A simple and rapid method for arsenite and arsenate speciation.** *Biohydrometallurgy - Fundamentals, Technology and Sustainable Development, Part B* Amsterdam: Elsevier ScienceJr VSTCaOG 2001, 541-546, (1348 pp).

17. Mouncey NJ, Mitchenall LA, Pau RN: **Mutational analysis of genes of the mod locus involved in molybdenum transport, homeostasis, and processing in *Azotobacter vinelandii*.** *J Bacteriol* 1995, **177**:5294-5302.
18. Peijnenburg WJGM, Jager T: **Monitoring approaches to assess bioaccessibility and bioavailability of metals: Matrix issues.** *Ecotoxicol Environ Saf* 2003, **56**:63-77.
19. Soutourina OA, Bertin PN: **Regulation cascade of flagellar expression in Gram-negative bacteria.** *FEBS Microbiol Rev* 2003, **27**:505-523.
20. Studholme DJ, Dixon R: **Domain architectures of σ^{54} -dependent transcriptional activators.** *J Bacteriol* 2003, **185**:1757-1767.
21. Rappas M, Schumacher J, Beuron F, Niwa H, Bordes P, Wigneshweraraj S, Keetch CA, Robinson CV, Buck M, Zhang X: **Structural insights into the activity of enhancer-binding proteins.** *Science* 2005, **307**:1972-1975.
22. Ellis PJ, Conrads T, Hille R, Kuhn P: **Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 Å and 2.03 Å.** *Structure* 2001, **9**:125-132.
23. Grunden AM, Shanmugam KT: **Molybdate transport and regulation in bacteria.** *Arch Microbiol* 1997, **168**:345-354.
24. Parkinson JS, Kofoed EC: **Communication modules in bacterial signaling proteins.** *Annu Rev Genet* 1992, **26**:71-112.
25. Kustu S, Santero E, Keener J, Popham D, Weiss D: **Expression of sigma 54 (*ntrA*)-dependent genes is probably united by a common mechanism.** *Microbiol Rev* 1989, **53**:367-376.
26. Leonhartsberger S, Huber A, Lottspeich F, Böck A: **The *hydH/G* genes from *Escherichia coli* code for a zinc and lead responsive two-component regulatory system.** *J Mol Biol* 2001, **307**:93-105.
27. Barrios H, Valderrama B, Morett E: **Compilation and analysis of σ^{54} -dependent promoter sequences.** *Nucleic Acids Res* 1999, **27**:4305-4313.
28. Schumacher J, Joly N, Rappas M, Zhang X, Buck M: **Structures and organisation of AAA+ enhancer binding proteins in transcriptional activation.** *J Struct Biol* 2006, **156**:190-199.
29. Zhang X, Chaney M, Wigneshweraraj SR, Schumacher J, Bordes P, Cannon W, Buck M: **Mechanochemical ATPases and transcriptional activation.** *Mol Microbiol* 2002, **45**:895-903.
30. Yang XF, Alani SM, Norgard MV: **The response regulator Rrp2 is essential for the expression of major membrane lipoproteins in *Borrelia burgdorferi*.** *Proc Natl Acad Sci Unit States Am* 2003, **100**:11001-11006.
31. Stafford GP, Scanlan J, McDonald IR, Murell JC: ***rpoN*, *mmoR* and *mmoG*, genes involved in regulating the expression of soluble methane monooxygenase in *Methylosinus trichosporium* OB3b.** *Microbiology* 2003, **149**:1771-1784.
32. Zhu L, Peng Q, Song F, Jiang Y, Sun C, Zhang J, Huang D: **Structure and regulation of the *gab* gene cluster, involved in the γ -aminobutyric acid shunt, are controlled by a σ^{54} factor in *Bacillus thuringiensis*.** *J Bacteriol* 2010, **192**:346-355.
33. Debarbouille M, Gardan R, Arnaud M, Rapoport G: **Role of *bkdR*, a transcriptional activator of the SigL-dependent isoleucine and valine degradation pathway in *Bacillus subtilis*.** *J Bacteriol* 1999, **181**:2059-2066.
34. Dombrecht B, Marchal K, Vanderleyden J, Michiels J: **Prediction and overview of the RpoN-regulon in closely related species of the Rhizobiales.** *Genome Biol* 2002, **3**(12):RESEARCH0076.
35. Cases I, Ussery DW, De Lorenzo V: **The σ^{54} regulon (sigmulon) of *Pseudomonas putida*.** *Environ Microbiol* 2003, **5**:1281-1293.
36. Endoh T, Habe H, Yoshida T, Nojiri H, Omori T: **A CysB-regulated and σ^{54} -dependent regulator, SfnR, is essential for dimethyl sulfone metabolism of *Pseudomonas putida* strain DS1.** *Microbiology* 2003, **149**:991-1000.
37. Grigoroudis AI, Panagiotidis CA, Lioliou EE, Vlassi M, Kyriakidis DA: **Molecular modeling and functional analysis of the AtoS-AtoC two-component signal transduction system of *Escherichia coli*.** *Biochim Biophys Acta Gen Subj* 2007, **1770**:1248-1258.
38. Bordes P, Wigneshweraraj SR, Schumacher J, Zhang X, Chaney M, Buck M: **The ATP hydrolyzing transcription activator phage shock protein F of *Escherichia coli*: Identifying a surface that binds σ^{54} .** *Proc Natl Acad Sci Unit States Am* 2003, **100**:2278-2283.
39. Dago AE, Wigneshweraraj SR, Buck M, Morett E: **A role for the conserved GAFTGA motif of AAA+ transcription activators in sensing promoter DNA conformation.** *J Biol Chem* 2007, **282**:1087-1097.
40. Hartl FU: **Molecular chaperones in cellular protein folding.** *Nature* 1996, **381**:571-580.
41. Mayer MP, Bukau B: **Hsp70 chaperone systems: Diversity of cellular functions and mechanism of action.** *Biol Chem* 1998, **379**:261-268.
42. Yoon H, Hong J, Ryu S: **Effects of chaperones on mRNA stability and gene expression in *Escherichia coli*.** *J Microbiol Biotechnol* 2008, **18**:228-233.
43. Ohki R, Kawamata T, Katoh Y, Hosoda F, Ohki M: ***Escherichia coli dnaJ* deletion mutation results in loss of stability of a positive regulator, CRP.** *J Biol Chem* 1992, **267**:13180-13184.
44. Berks BC, Sargent F, Palmer T: **The Tat protein export pathway.** *Mol Microbiol* 2000, **35**:260-274.
45. Pérez-Rodríguez R, Fisher AC, Perlmutter JD, Hicks MG, Chanal A, Santini CL, Wu LF, Palmer T, DeLisa MP: **An Essential Role for the DnaK Molecular Chaperone in Stabilizing Over-expressed Substrate Proteins of the Bacterial Twin-arginine Translocation Pathway.** *J Mol Biol* 2007, **367**:715-730.
46. Rodríguez F, Arsène-Pløetze F, Rist W, Rüdiger S, Schneider-Mergener J, Mayer MP, Bukau B: **Molecular Basis for Regulation of the Heat Shock Transcription Factor sigma32 by the DnaK and DnaJ Chaperones.** *Mol Cell* 2008, **32**:347-358.
47. De Lorenzo V, Herrero M, Jakubzik U, Timmis KN: **Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria.** *J Bacteriol* 1990, **172**:6568-6572.
48. Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: a laboratory manual.** Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2. 1989.
49. Smyth GK, Speed TP: **Normalization of cDNA microarray data.** *Methods* 2003, **31**:265-273.
50. Delmar P, Robin S, Daudin JJ: **VarMix: Efficient variance modeling for the differential analysis of replicated gene expression data.** *Bioinformatics* 2004, **21**(4):502-8.
51. Benjamini Y, Yekutieli D: **The control of the false discovery rate in multiple hypothesis testing under dependency.** *Ann Stat* 2001, **29**:1165-1188.
52. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29**(9):e45.
53. Pfaffl MW, Horgan GW, Dempfle L: **Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR.** *Nucleic Acids Res* 2002, **30**(9):e36.
54. Anderson GL, Williams J, Hille R: **The purification and characterization of arsenite oxydase from *Alcaligenes faecalis*, a molybdenum-containing hydroxylase.** *J Biol Chem* 1992, **267**:23674-23682.

doi:10.1186/1471-2180-10-53

Cite this article as: Koehler et al.: Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. *BMC Microbiology* 2010 **10**:53.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



C/ Discussion

Deux études portant sur la régulation de l'arsénite oxydase étaient recensées jusqu'ici. Ainsi, chez *A. tumefaciens*, un système de régulation à deux composants, codés par les gènes *aoxRS* a notamment été démontré comme étant impliqué dans la régulation des gènes *aox* (Kashyap *et al.*, 2006). AoxR présente des domaines de réception de signaux environnementaux transmis par des senseurs kinases et AoxS possède un motif de 14 aa caractéristique d'un senseur kinase qui jouerait le rôle de récepteur du signal périplasmique. Après fixation de l'As(III), AoxS s'autophosphorylerait puis transmettrait son groupement phosphate au régulateur AoxR. AoxR phosphorylé induirait l'expression de l'opéron *aox*.

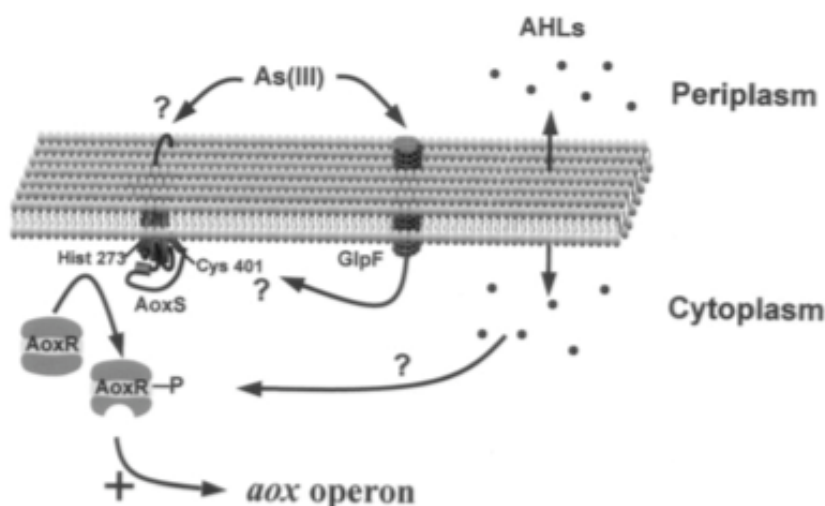


Figure 21 : Régulation de l'opéron *aox* chez *Agrobacterium tumefaciens*.

Le domaine receveur de AoxS est visualisé par une chaîne de 14 aa dans le périplasm. Celui-ci pourrait servir de récepteur du signal environnemental tel que l'As(III) ou un autoinducteur (AHLs). Cys401 est proposé comme domaine receveur potentiel et Hist273 correspond au site de phosphorylation canonique. L'entrée de l'As(III) dans la cellule est visualisée ici *via* une aquaglycéroporine, GlpF (extrait de Kashyap *et al.*, 2006).

Chez *Achromobacter* sp. SY8 un gène additionnel *aoxX* a été identifié et serait impliqué dans la régulation de l'activité arsénite oxydase. Ce gène code une protéine périplasmique impliquée dans la fixation du phosphate. Chez cet organisme l'opéron

aoxABCD serait aussi régulé par le système à deux composants AoxSR. Les auteurs proposent que le signal est perçu directement soit par AoxS soit par AoxX qui fixerait l'As(III) et interagirait avec la boucle périplasmique de AoxS (Cai *et al.*, 2009).

Chez *H. arsenicoxydans* une approche par mutagenèse aléatoire couplée à des analyses des transcrits *aoxB* par q-RT-PCR et de la protéine AoxB par Western-Blot, nous a permis de proposer un modèle conceptuel. Ainsi, l'arsénite serait perçu directement ou indirectement par le senseur AoxS induisant son autophosphorylation. Puis, AoxS phosphorylé transférerait son groupement phosphate à son régulateur AoxR. AoxR phosphorylé se lierait à RpoN pour initier la transcription de l'opéron *aox*.

La protéine DnaJ a également été caractérisée comme régulateur de l'opéron *aox* dans cette étude. Cette protéine est composée (i) d'un domaine N-terminal conservé, appelé domaine J, d'environ 70 aa, (ii) d'une région riche en glycine, le domaine G, (iii) d'un domaine central contenant quatre répétitions d'un motif CXXCXGXG, le domaine CRR et (iv) d'une région C-terminale de 120 à 170 résidus. Chez *Escherichia coli* cette protéine est un co-chaperon impliqué dans la liaison du substrat et sa présentation à DnaK. DnaK est une ATPase qui peut lier à la fois des molécules d'ATP et des polypeptides dénaturés. L'hydrolyse de l'ATP est favorisée par l'interaction du domaine J de DnaJ avec DnaK. Celle-ci est nécessaire pour la fixation du substrat sous une forme étirée, ce qui évite son agrégation. L'interaction avec la protéine GrpE permet ensuite de régénérer le complexe DnaK-ATP, ce qui diminue l'affinité de DnaK pour le substrat et le libère (Genevaux *et al.*, 2007).

Ainsi, dans notre cas, le rôle de DnaJ peut être multiple. Ce co-chaperon associé à DnaK peut aussi bien être impliqué dans la stabilité du transcrit *aox*, le repliement de AoxR ou de tout autre protéine impliquée dans la régulation de l'activité arsénite oxydase. Enfin, une régulation post-traductionnelle des protéines structurales de l'arsénite oxydase AoxAB par DnaJ, ne peut être exclue.

Les résultats obtenus lors de cette étude soulignent l'intérêt d'utiliser des techniques de génétique classique pour décortiquer des mécanismes complexes. Par ailleurs, les outils de biologie moléculaire ont été déterminants dans la validation des hypothèses. Il s'agit de la première étude qui propose un mécanisme de régulation de l'activité arsénite oxydase incluant RpoN, DnaJ et AoxR comme un co-activateur de RpoN.

Mécanismes globaux du stress arsénié
chez *Herminiimonas arsenicoxydans*

A/ Introduction

De nombreux micro-organismes sont capables de coloniser des environnements contaminés par l'arsenic et ont, de ce fait, développé une multitude de mécanismes d'adaptation au stress engendré par cet élément. Les mécanismes de réponse au stress arsénié commencent à être étudiés au niveau de la cellule eucaryote, mais aucune information n'est disponible sur la réponse générale de la cellule bactérienne exposée à un environnement fortement contaminé par l'arsenic. Aussi, cette partie de notre travail visait à étudier la réponse adaptative chez *H. arsenicoxydans* par l'analyse du protéome en présence du métalloïde sous sa forme trivalente (As(III)), considérée comme la plus toxique et l'analyse du transcriptome après induction à l'As(III) pendant 15 min.

Le protéome désigne l'ensemble des protéines synthétisées par une cellule ou un organisme à un instant donné. Celui-ci présente un certain nombre d'avantages évidents : 1) seule la protéine présente un effet biologique, 2) un même génome peut naturellement donner naissance à une variété quasi infinie de protéomes différents et 3) le niveau d'activité d'une protéine ne peut être prédit à partir du seul niveau d'expression des messagers. En effet, certains messagers sont peu, voire pas du tout traduits (Rabilloud, 2002). Cette approche protéomique permet ainsi de recenser l'ensemble des protéines (carte protéomique) impliquées dans l'adaptation de la bactérie au stress arsénié. Toutefois, l'analyse protéomique se heurte à des difficultés, en particulier liées à la quantité considérable de protéines possibles à analyser. Aussi, la stratégie de cette étude a été de coupler cette approche protéomique à une approche transcriptomique.

Le transcriptome consiste à établir les profils d'expression des ARNm à grande échelle. Dans la technique des microarrays, les oligonucléotides correspondants à tous les gènes dont on désire étudier l'expression, sont déposés sur une puce. Les ARNm des cellules à étudier ont été extraits, amplifiés et marqués indirectement à la cyanine, puis les ADNc ont été hybridés sur la puce contenant les oligonucléotides. Dans le cas présent, cette technique a été utilisée pour des analyses différentielles (+ et - As(III)), qui ont pu être visualisées, grâce à la fluorescence, par une intensité de couleur du rouge (sur-exprimés) au vert (sous-exprimés).

Ma participation à ce projet a porté sur l'extraction des ARNs de la souche après induction ou non à l'As(III) pendant 15 min. J'ai ensuite analysé et interprété les données de transcriptome. Ces résultats associés aux données du protéome m'ont permis de proposer un modèle de réponse globale au stress arsénié chez *H. arsenicoxydans*.

L'ensemble de ces travaux a fait l'objet d'une publication (Weiss *et al.*, 2009), présentée ci dessous.

B/ Manuscrit



Available online at www.sciencedirect.com

ScienceDirect

Biochimie 91 (2009) 192–203

BIOCHIMIE

www.elsevier.com/locate/biochi

Research paper

Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data

Stéphanie Weiss^{a,1}, Christine Carapito^{b,1}, Jessica Cleiss^a, Sandrine Koechler^a, Evelyne Turlin^c,
Jean-Yves Coppee^f, Michaël Heymann^{a,2}, Valérie Kugler^a, Magalie Stauffert^a,
Stéphane Cruveiller^{d,e}, Claudine Médigue^{d,e}, Alain Van Dorsselaer^b,
Philippe N. Bertin^a, Florence Arsène-Ploetze^{a,*}

^a Génétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS and Université Louis Pasteur, 28 rue Goethe, 67000 Strasbourg, France

^b Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DSA, ULP, CNRS, 25 rue Becquerel, 67087 Strasbourg, France

^c Génétique des Génomes Bactériens, URA2171, Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France

^d CNRS UMR8030, Génomique Métabolique, 2 rue Gaston Crémieux, 91057 Evry Cedex, France

^e Commissariat à l'Energie Atomique (CEA), Direction des Sciences du Vivant, Institut de Génomique, Genoscope,

Laboratoire de Génomique Comparative, 2 rue Gaston Crémieux, 91057 Evry Cedex, France

^f Plate-forme puces à ADN, Genopole/Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France

Received 16 May 2008; accepted 31 July 2008

Available online 25 September 2008

Abstract

The arsenite-oxidizing strain *Herminiimonas arsenicoxydans* proteome was investigated with gel electrophoresis and tandem mass spectrometry analyses. The comparison of experimental and theoretical M_r and pI , as well as that of peptide sequences identified by MS and predicted protein sequences, allowed the correction of five protein annotations. More importantly, the functional analysis of SDS- and 2D-PAGE proteome maps obtained in the presence of arsenic, combined with partial transcriptomic results indicate that *H. arsenicoxydans* expressed genes and proteins required not only for arsenic detoxification or stress response but also involved in motility, exopolysaccharide synthesis, phosphate import or energetic metabolism. This study provides therefore new insights into the adaptation processes of *H. arsenicoxydans* in response to arsenic. © 2008 Elsevier Masson SAS. All rights reserved.

Keywords: Arsenic; Genome annotation; *Herminiimonas arsenicoxydans*; Proteomic

1. Introduction

Arsenic is widely distributed in the environment, released from both natural and anthropogenic sources. The biogeochemical cycle of this element strongly depends on microbial transformation that affects the mobilization and the distribution

of arsenic species in the environment [1,2]. In particular, the implication of microorganisms in oxidation reactions between the bioavailable As(III) and the less mobile As(V) oxidation states has been demonstrated. Moreover, inorganic forms of arsenic are usually considered as the most toxic. Contamination of drinking water by these soluble forms, arsenite and arsenate, has often been reported, and arsenic has been identified as a major risk for human health in different parts of the world [3].

The beta-proteobacterium *Herminiimonas arsenicoxydans* is a member of the new genus *Herminiimonas*, which is encountered in water [4]. This bacterium has been isolated from the activated sludge of an industrial treatment plant contaminated with heavy metals. Recently, the genome of *H.*

Abbreviation: COG, cluster of orthologous groups of proteins; M_r , relative molecular mass; CAI, Codon Adaptation Index; CDS, CoDing sequence; PAGE, polyacrylamide gel electrophoresis.

* Corresponding author. Tel.: +33 (0)3 9024 1812; fax: +33 (0)3 9024 2028.

E-mail address: florence.ploetze@gem.u-strasbg.fr (F. Arsène-Ploetze).

¹ These two authors contributed equally to this work.

² Present address: Institut de Biologie et Chimie des Protéines (UMR5086), CNRS, University of Lyon 1, IFR 128 BioSciences Lyon-Gerland, Lyon, France.

arsenicoxydans has been fully sequenced and annotated [5]. *H. arsenicoxydans* is able to resist to multiple toxic elements, especially arsenic. Remarkably, the genome sequence revealed the presence of four *ars* operons, which is unusual as compared to other metal/metalloid resistant strains. *H. arsenicoxydans* is also able to oxidize the more toxic As(III) into the less toxic inorganic form, As(V). This oxidation process requires the arsenite oxidase protein, which is encoded by the *aoxAB* operon [6]. Besides genes involved in arsenic resistance and metabolism, *H. arsenicoxydans* has acquired two properties regarding arsenic. First, the strain produces a thick capsule of exopolysaccharides, which has been shown to scavenge arsenic as granules. Second, a significant increase in swimming speed was demonstrated in the presence of As(III). This process requires arsenite oxidase activity, suggesting a gain of energy resulting from arsenite oxidation [5].

Genome sequencing gives an inventory of genes possibly playing a role in adaptive response. The complete proteome map characterization demonstrates which of the predicted genes are actually expressed under particular conditions. A few examples illustrate the relevance of such approaches to study environmental isolates. In *Bradyrhizobium japonicum*, the complete proteome analysis has contributed to the understanding of the interaction mechanisms between plant and bacteroid [7]. Proteome map analysis may also help to discover novel metabolic pathways. For example, the unconventional reduction of CO₂ to methane in *Methanosarcina acetivorans* was deciphered using such an approach [8]. More recently, proteomic studies allowed a complex physiological description of an uncultivable endosymbiont of *Riftia pachyptila* [9]. Applying such a strategy may be useful to deduce the relevance of particular proteins in the physiology of *H. arsenicoxydans*, and more precisely in the adaptation to arsenic contaminated environments.

To evaluate the metabolic pathways that are expressed by *H. arsenicoxydans* in the presence of arsenite, we present here a complete proteome map that extend our previous differential proteomic studies [5,10]. The protein soluble fraction was separated by 2D-PAGE, whereas the membrane fraction was separated by SDS-PAGE. Combining this approach with transcriptomic data, gave rise to an integrated functional view of the cellular processes involved in *H. arsenicoxydans* arsenic-dependent adaptation.

2. Materials and methods

2.1. Bacterial strains, media and growth conditions

The *H. arsenicoxydans* strain ULPAs1 was cultivated in a chemically defined medium (CDM) containing 2.66 mM of As(III) (NaAsO₂) in Falcon tubes placed on a rotary shaker at 25 °C, as previously described [11].

2.2. Preparation of soluble fraction and 2D-PAGE separation

Late exponential phase cells (100 mL) were harvested by centrifugation (7000 × g, 10 min, 4 °C). Cell pellets were

washed with CDM and resuspended in 1 mL of distilled water containing 2.5% (w/v) Protease Inhibitor Cocktail (Complete, Roche). After treatment with DNase (4 µL at 1 mg/ml) and RNase (2 µL at 1 mg/mL), cells were disrupted with a “FP120 FastPrep Cell disruptor” (Bio101) (two-times 30 s at maximum speed with 1-min intervals on ice). Cell debris were removed by ultracentrifugation for 60 min at 90,000 × g. Supernatant is a mixture of cytoplasmic and periplasmic fraction (soluble extract). This fraction was divided into aliquots and stored at –20 °C. A total of 60 µg (for analytical gels) or 150 µg (for preparative gels) of this extract were solubilized in 400 µL of rehydration solution (0.5% (v/v) pharmylate 3–10, 8 M urea, 65 mM DTT, 2% (v/v) NP40), and loaded onto a 18 cm pH 4–7 IPG strip using the in-rehydration technique as previously described [12].

IEF was conducted using the horizontal Multiphor II system (Amersham Biosciences) at a temperature of 20 °C. For both analytical and preparative gels, focusing was performed for 3 h at 300 V, 1 h at 750 V, 30 min at 1500 V, 16 h at 2500 V, and 2 h at 3500 V (total = 50 kVh). IPGs were equilibrated using the in-rehydration technique. The second dimension was performed with 11.5% (w/v) SDS-polyacrylamide gels using the Protean II xi 2D Multicell system (BioRad). Proteins were stained with silver nitrate, and gels were digitized using a GS-800 Calibrated Densitometer (BioRad). Digitized 2D-gel patterns were edited and matched using the PDQUEST software package (PDI, Humington Station). Each 2D-gel was performed in triplicate.

2.3. Preparation of membrane fraction and SDS-PAGE separation

Cell pellets were resuspended in 5 mL lysis buffer, containing 200 µL of Protease Inhibitor Cocktail, 4 µL of a 1 mg/mL DNase solution, and 2 µL of a 1 mg/mL RNase solution, and disrupted two-times 30 s. The supernatant was centrifuged at 13,000 × g for 60 min at 4 °C. The resulting pellet was resuspended in 1 mL lysis buffer and the enzymatic treatment was repeated. After centrifugation, the supernatant was discarded and the pellet was resuspended with 1 mL of 1% (v/v) ABS14. After incubation at 4 °C overnight, the sample was centrifuged (13,000 × g, 4 °C, 90 min) and the supernatant was dialysed with 400 mL of an ABS14 0,1% (v/v) solution at 4 °C, leading to the membrane protein extract. This extract was divided into aliquots and stored at –20 °C. The separation of these proteins by 2D-PAGE failed, therefore 60 µg were separated on 12% SDS-PAGE. Each gel was performed in triplicate, and stained with silver nitrate.

2.4. NanoLC-MS/MS analysis

In situ digestion of gel spots was performed with an automated protein digestion system, MassPREP Station (Waters, Milford, MA). Gel plugs were washed three-times with a mixture of 50%/50% NH₄HCO₃ (25 mM)/ACN. Cysteine residues were reduced with dithiothreitol at 57 °C for 30 min and alkylated with iodoacetamide at room temperature for

20 min. After dehydration with acetonitrile, proteins were digested *in gel* with 10 µL of 12.5 ng/µL modified porcine trypsin (Promega, Madison, WI) in 25 mM NH₄HCO₃ overnight at 37 °C. Then a double extraction was performed, first with 60% acetonitrile in 5% formic acid and, second with 100% acetonitrile. The resulting tryptic peptides were analysed by nanoLC-MS/MS. These analyses were performed on two systems: a CapLC capillary LC system (Waters, USA) coupled to a hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer Q-TOF II (Waters, USA) and on an Agilent 1100 Series nanoflow LC system (Agilent Technologies, Palo Alto, USA) coupled to a HCT™ Plus ion trap (Bruker Daltonics, Bremen, Germany). For nanoLC-Q-TOF II coupling, chromatographic separations were conducted on a reverse-phase (RP) capillary column (Pepmap C18, 75 µm i.d., 15 cm length, LC Packings) with a flow rate of 200 nL/min accomplished by a pre-column split. An external calibration was performed using a solution containing 2 pmol/µL GFP ([Glu1]-Fibrinopeptide B, Sigma). Mass data acquisitions were piloted by MassLynx software (Waters, USA) using automatic switching between MS and MS/MS modes.

For nanoLC-HCT Plus coupling, chromatographic separations were conducted on a reverse-phase (RP) capillary column (Zorbax300SB-C18, 75 µm i.d., 15 cm length, Agilent Technologies) with a 200 nL/min flow rate. In MS mode, scanning was performed using standard enhanced scan mode at 8100 m/z/s whereas in MS/MS mode, the Ultra Scan mode at 26,000 m/z/s was used. The system was fully controlled by Agilent ChemStation and EsquireControl (Bruker Daltonics, Germany) software.

2.5. Protein identification and cellular localization

In a first step, the complete genome sequence of *H. arsenicoxydans* (3.4 Mbps) was used to search MS/MS data (<http://www.genoscope.cns.fr/agc/mage/>). The genome sequence was fragmented into regular segments of 7500 bps with an overlap length of 2500 bps to generate a nucleic acid database. This in-house constructed nucleic acid database was imported into a local Mascot (Matrix Science, London, UK) server and translated on the fly in all six reading frames. The nanoLC-MS/MS data were searched against this constructed genome database. Trypsin was defined as the cleavage enzyme, one missed cleavage was allowed, 0.25 Da error both in MS and MS/MS mode was tolerated, and carbamidomethylation for the cysteines and oxidation for the methionines were set as variable modifications. The predicted function of the identified proteins was determined using Blast (<http://www.ncbi.nlm.nih.gov/BLAST/>) with standard parameters. In parallel, the predicted CDS were extracted from genome annotation data and the MS/MS data were also searched against this predicted proteins database.

The criteria established by the proteomics journals' guidelines were applied for validating the identifications [13,14]. Indeed, protein identifications were validated when at least two peptides with high quality MS/MS spectra (ion scores higher than 35) were detected.

The identification results obtained with complete genome sequence database searches on one side, and with the predicted proteins database (3294 proteins) searches on the other side were then compared with experimental proteome data.

Predictions of protein localization were performed using the PSORTb v.2.0 program [15], keeping the default parameters defined on the website (<http://www.psort.org/psortb/>).

2.6. Differential protein expression analysis

Cells were cultivated in CDM, in the absence or in the presence of As(III) as previously described [10]. Protein extractions were performed as described above. For gel electrophoresis using pH 3–10 IPG strips, proteins separation was performed as described above, and stained with silver nitrate. Gels were analyzed using same parameters that previously allowed a differential analysis [5,10]. When pH 4–7 IPG strip were used, focusing was performed using the IPGPhor system (rehydration 6 h, 6 h at 30 V, 2 h at 150 V, 2 h at 500 V, 2 h at 1000 V, 2 h at 3000 V, and 4 h at 8000 V, total = 50 kVh) (GE Healthcare Biosciences, Australia). The second dimension was performed on 11.5% SDS-PAGE, using the EttanDAlt system (GE Healthcare Biosciences, Australia). Gels were stained with CBB, and digitized using an Image Scanner (Amersham Pharmacia) and the LabScan software (v 3.0, Amersham Pharmacia Biotech). Analysis was performed using the ImageMaster 2D platinum software (v. 6.01, GE Healthcare Biosciences, Australia). 12 gels, i.e. six gels obtained in each condition, were compared. The Student's *t*-value and the "ratio" (average of the relative volumes obtained in the presence of arsenic divided by the average of the relative volumes obtained in the absence of arsenic) were calculated for each spot, as advocated by the manufacturer. Only spots sharing a Student's *t*-value greater than 1.4 (*p*-value less than 0.1) and ratio greater than 2 were analyzed. The selected spots were cut from 2D-gel and stored at –20 °C. Picked spots were washed with 100 µL of 25 mM NH₄HCO₃ buffer and dehydrated with 100 µL of ACN, and the whole process was repeated once more. The samples were vacuum-dried for 10 min, reduced (10 mM DTT/25 mM NH₄HCO₃ buffer (100 µL) at 56 °C for 1 h) and alkylated (25 mM iodoacetamide/25 mM NH₄HCO₃ buffer (100 µL) at room temperature in the dark for 1 h). After three washes for 5 min in 25 mM NH₄HCO₃ and ACN alternately, samples were vacuum-dried, rehydrated overnight at room temperature in the presence of trypsin (three volumes of 12.5 ng trypsin (Promega, V5111)/µL in 25 mM NH₄HCO₃ buffer, freshly diluted). Tryptic peptides were extracted from the gels by sonication for 30 min in 5 µL of 35% H₂O/60% ACN/5% HCOOH. Mass measurements were performed on a BIFLEX III™ MALDI-TOF (Bruker Daltonics, Bremen, GE) equipped with the SCOUT™ High Resolution Optics with X-Y multisample probe and griddle reflectors. The instrument was run in positive ion reflector mode at a maximum accelerating potential of 19 kV. A saturated solution of α -cyano-4-hydroxycinnamic acid (Sigma, Saint Louis, MO) in acetone was used as a matrix. Spreading and fast evaporation of 0.5 µL of matrix solution enabled the formation of a fine layer of crystals, on which

a droplet of 0.5 μ L of aqueous HCOOH (5%) solution was mixed first with 0.5 μ L of peptide-containing digest, and then with 0.3 μ L saturated matrix solution (in 50% H₂O/50% ACN). The preparation was vacuum-dried and washed once with 0.7 μ L of aqueous HCOOH (5%). Mass spectra were internally calibrated with trypsin autolysis peaks ($m/z = 842.510$ and $m/z = 2211.105$). Monoisotopic peptide masses were assigned by the XToF v5.1.5 software (Bruker) and the peak list transferred through MS BioTools™ program (Bruker Daltonics, Bremen, GE) as input to search against the *H. arsenicoxydans* protein database using MASCOT v1.9 software (Matrix Science, London, UK). Tryptic mass searches retained only data with up to one missed tryptic cleavage and optional methionine oxidation, with mass accuracy limited to 50 ppm.

2.7. RNA manipulation

Total RNA was extracted from *H. arsenicoxydans* grown in 50 mL CDM medium for 24 h and induced by 1.33 mM As(III) 15 min before extraction. Bacteria were collected by centrifugation for 15 min at $5000 \times g$ at 4 °C and stored at –80 °C to prevent RNA degradation. To improve RNA extraction, an additional lysis step was performed as follows: bacterial pellets were resuspended in 400 μ L of suspension solution (25 mM Tris–HCl pH 7.6, 10 mM EDTA, 20% glucose (vol/vol)) and transferred into microtubes containing 500 μ L of acidic phenol (pH 4.5) and 0.4 g of 0.1 mm-diameter glass beads (VWR). Cells were mechanically broken with a “FP120 FastPrep Cell Disruptor” (Bio 101) as homogenizer (two 30 s cycles of homogenization at maximum speed with 1 min interval on ice). RNA was then extracted as previously described [10].

A custom oligonucleotide array containing 5516 probes was manufactured by Agilent Technologies. Probes were replicated on the array to reach a final density of 15K probes per array. RNA samples (5 μ g) were indirectly labeled using Atlas PowerScript Fluorescent Labeling kit (Clontech) with a mixture of random hexamers (pdN6), according to the conditions recommended by the manufacturer with the following modifications: after reverse-transcription, RNA was digested with RNase H for 45 min at 37 °C. cDNAs were coupled with cyanines using Cy3 Mono-Reactive Dye or Cy5 Mono-Reactive Dye (Amersham Bioscience) and the resulting fluorescent cDNA were then purified with QIAquick PCR Purification Kit (Qiagen). Their quality and concentration were determined by spectroscopy at 260 nm, 280 nm and 550 nm (Cy3) or 650 nm (Cy5). Cy3 and Cy5 target quantities were normalized at 250 pmol, mixed and thereafter concentrated with Microcon YM-30 (Millipore). Sample volumes were adjusted to 50 μ L in Agilent hybridization buffer. Hybridizations were performed for 17 h at 65 °C in a rotating hybridization oven according to Agilent’s recommendations, i.e. two colors microarray-based gene expression analysis protocol, and slides were washed as recommended. Arrays were scanned with an Axon 4000B scanner with fixed PMT (PMT = 550 for Cy3 and 650 for Cy5). Data were acquired and analyzed by Genepix Pro 5.0 (Axon Instrument). The

experiment design included five biological replicates. For each of them, induced and non-induced cells were compared in dye swap experiments. The resulting ten arrays were analysed using the R software (<http://www.R-project.org>). A slide by slide loess normalization was performed using the limma package [16]. Valid log₂ expression ratios from replicated spots were averaged on each array, so as to get statistically independent ratios for each oligonucleotide included in the array design. For the same reason, dye swap arrays were also averaged. Oligos that had no valid expression ratios on the ten arrays were excluded from the data set for further analysis, which was carried out using the varmixt package and the VM option [17]. The resulting raw *p*-values were adjusted according to a Benjamini and Yekutieli procedure [18].

3. Results

3.1. Proteome map of *H. arsenicoxydans* cultivated in the presence of As(III)

The 3.4 Mbps single chromosome contains a total of 3294 protein-encoding genes (3267 CDSs and 27 fCDSs, as modified from published data, [5]), and 58 RNA-encoding genes. The theoretical proteome map of *H. arsenicoxydans* was constructed using the M_r/pI dataset calculated from annotated CDS available on the MaGe website (<http://www.genoscope.cns.fr/agc/mage/>). The molecular mass of the 3294 proteins ranged from 2.5 to 329 kDa, and *pI* values from 3.33 to 12.7 (Supplementary Fig. S1), with 32% cytoplasmic proteins and 22% of proteins localized in the inner-membrane, periplasm or outer-membrane (Supplementary Fig. S2A). The localization of remaining proteins (46%) could not be predicted with the pSORTb program.

A closer look at the M_r repartition revealed that a large majority (89%) of the predicted proteins was included in the 10–100 kDa range, with only 67 proteins showing a $M_r > 100$ kDa and 307 showing a $M_r < 10$ kDa. Among the proteins in the 10–100 kDa M_r range, 1593 had *pI* values ranging from 4 to 7, corresponding to 48% of the total proteome and a total of 1325 proteins had *pI* values higher than 7. However, 76% of the cytosolic proteins, according to the pSORTb prediction, had a *pI* between 4 and 7, whereas the majority of the membrane and periplasmic proteins (70%) had a *pI* higher than 7. These observations supported the use of a 4–7 IPG strip to separate soluble fraction proteins. Moreover, to complete this soluble protein map and have access to membrane proteins, a SDS-PAGE approach was also used.

The protein content was investigated in *H. arsenicoxydans* grown in minimal medium in the presence of 2.66 mM AsNaO₂ (Fig. 1). The proteomic map was similar to the map obtained in a previous differential analysis that compared proteins accumulated in the presence and in the absence of arsenic [10]. After 2D-PAGE separation of the soluble fraction, 437 spots were analysed by nanoLC-MS/MS (Fig. 1). A total of 365 proteins were identified (Supplementary Table S1) using the validation criteria described in materials and methods. 5% of the proteins identified on the 2D-PAGE had

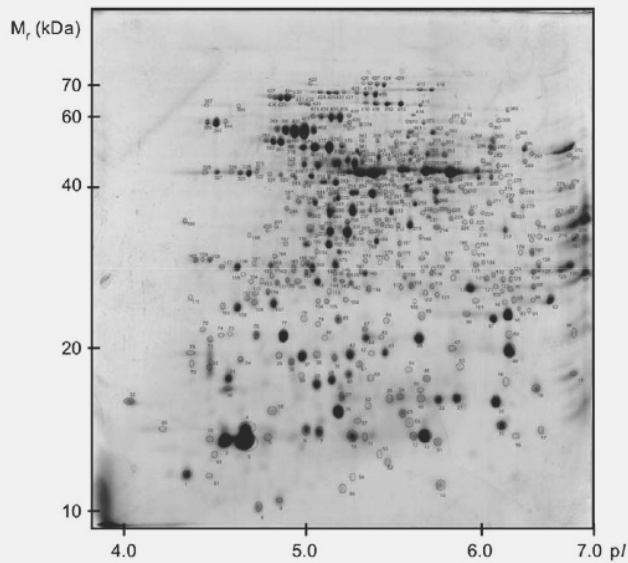


Fig. 1. 2D-PAGE proteome map of the soluble fraction. Spots analysed with nanoLC-MS/MS are surrounded and numbered. Complete identification results of these spots are provided in supplementary Tables S1 and S2. An interactive version of this map can be visualized on <http://inpaact.u-strasbg.fr/~db/index.php> [31].

periplasmic or inner-membrane localization and no protein had outer-membrane localization (Supplementary Fig. S2C).

The analysis of the membrane fraction of *H. arsenicoxydans* separated by SDS-PAGE was performed by a systematic cut-out (every 2 mm) of the gel (Lane 2, Fig. 2). A total of 28 gel slices were then analysed by nanoLC-MS/MS. 139 unique proteins were identified using the same criteria as those used for 2D-PAGE (Supplementary Table S2). The subcellular localization

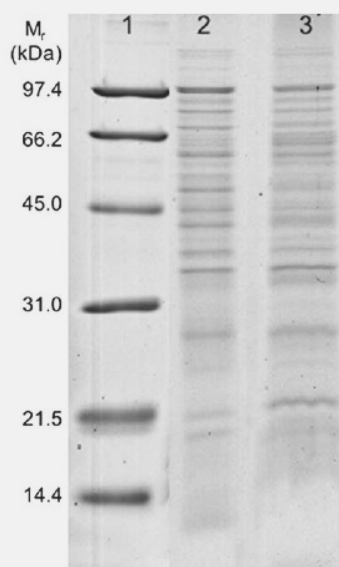


Fig. 2. SDS-PAGE separation of membrane proteins. Lane 1: Protein marker (SDS-PAGE Standards low range, Bio-Rad). Lane 2: Protein extract from cells grown in presence of 2.66 mM As(III). Lane 3: Protein extract from cells grown in absence of arsenic (control).

of the identified proteins was determined with the PSORTb v.2.0 program. Among the 139 unique proteins identified on SDS-PAGE, 39 (28%) were localized in inner- or outer-membrane or in the periplasm (Supplementary Fig. S2D). Fifty-six proteins were identified on both SDS- and 2D-PAGE: these proteins count among the very abundant proteins. All together, a total of 447 unique proteins validating 447 predicted ORFs were identified, representing 13.6% of the total proteome predicted from genome sequence (<http://www.genoscope.cns.fr/agg/mage>). The percents of proteins originated from outer-membrane, periplasm or cytoplasm were similar in the experimental and the theoretical proteomes (Fig. S2B), a slightly reduced percent of the inner-membrane proteins was resolved, as compared to the theoretical proteome (7% versus 18%, Supplementary Fig. S2B).

3.2. Theoretical versus experimental proteome and validation of genome annotation

Previous reports have shown that proteome maps provide experimental data useful to validate genome annotations [19,20]. Indeed, discrepancies between the predicted and experimental M_r/pI of proteins are usual and may result not only from cellular proteolytic cleavage and other post-translational modifications [21–23], but also from incorrect genome annotation, in particular from incorrect start predictions. These start predictions are mainly based on algorithms, which may sometimes predict several start codons for the same protein. MS identification of proteins may therefore help to determine which of those predicted start codons is relevant. Both the experimental relative molecular mass and isoelectric point of each protein were determined on the *H. arsenicoxydans* 2D-PAGE map and compared to the calculated theoretical M_r/pI , to search for annotation mistake among the identified proteins. This strategy led to the detection of one significant anomaly: the 320 amino acid predicted HEAR2960 had M_r/pI values of 34.2 kDa/10.1 while its apparent M_r/pI values on gel were 18.7 kDa/5.23. Sequence alignment, analysis of CDS predictive curve representing the codon usage, and MS/MS data of the sequenced peptides suggest that the length of the protein had been initially overestimated. The start codon was therefore shifted 169 amino acid residues downstream of the previously proposed translational codon, and encodes a valine residue. The M_r and pI values of this 151 amino acid corrected protein (17.1 kDa and 5.2) were close to the values measured on gel. This re-annotated protein shows now 88% identity with the sequence of DksA in *Herbaspirillum seropedicae*. In addition, MS/MS peptidic data were compared to both the complete genome sequence and the predicted protein database (see Section 2). This strategy allowed the correction of four protein start codons, i.e. HEAR0146, HEAR0538, HEAR2634, HEAR3178 (Supplementary Table S3), for which an additional peptide was found upstream of the annotated start codon. For example, the HEAR3178 start codon was corrected on the basis of homologies with NusG protein sequences present in SwissProt and TrEMBL databases. These start codon modifications were

implemented into MaGe annotations. However, it should be emphasized that these modifications did not modify the functional annotation of the five proteins. Finally, experimental proteome data allowed us to validate the existence of 57 hypothetical proteins with no known function (Supplementary Tables S1 and S2).

3.3. Analysis of general metabolic pathways expressed by *H. arsenicoxydans* cultivated in the presence of arsenic

To determine which metabolic pathways are used by *H. arsenicoxydans* when cultivated in the presence of arsenic,

proteins were categorized into functional COG classes using the COGNITOR program [24]. A total of 97% of the proteins identified from gel analysis were classified in at least one COG group whereas 80% of the total predicted proteins were assigned to a COG class. Fig. 3 presents the repartition of the experimental (black) versus predicted (gray) proteins into 18 functional groups. For example, proteins involved in amino acid transport and metabolism, were overrepresented in the experimental proteome map. Indeed, 10.1% of the predicted ORFs are involved in these processes but these proteins accounted for 14.3% of the experimentally identified proteins. Similarly, several enzymes involved in energy production (8.5% in the experimental proteome versus 6.6% in the

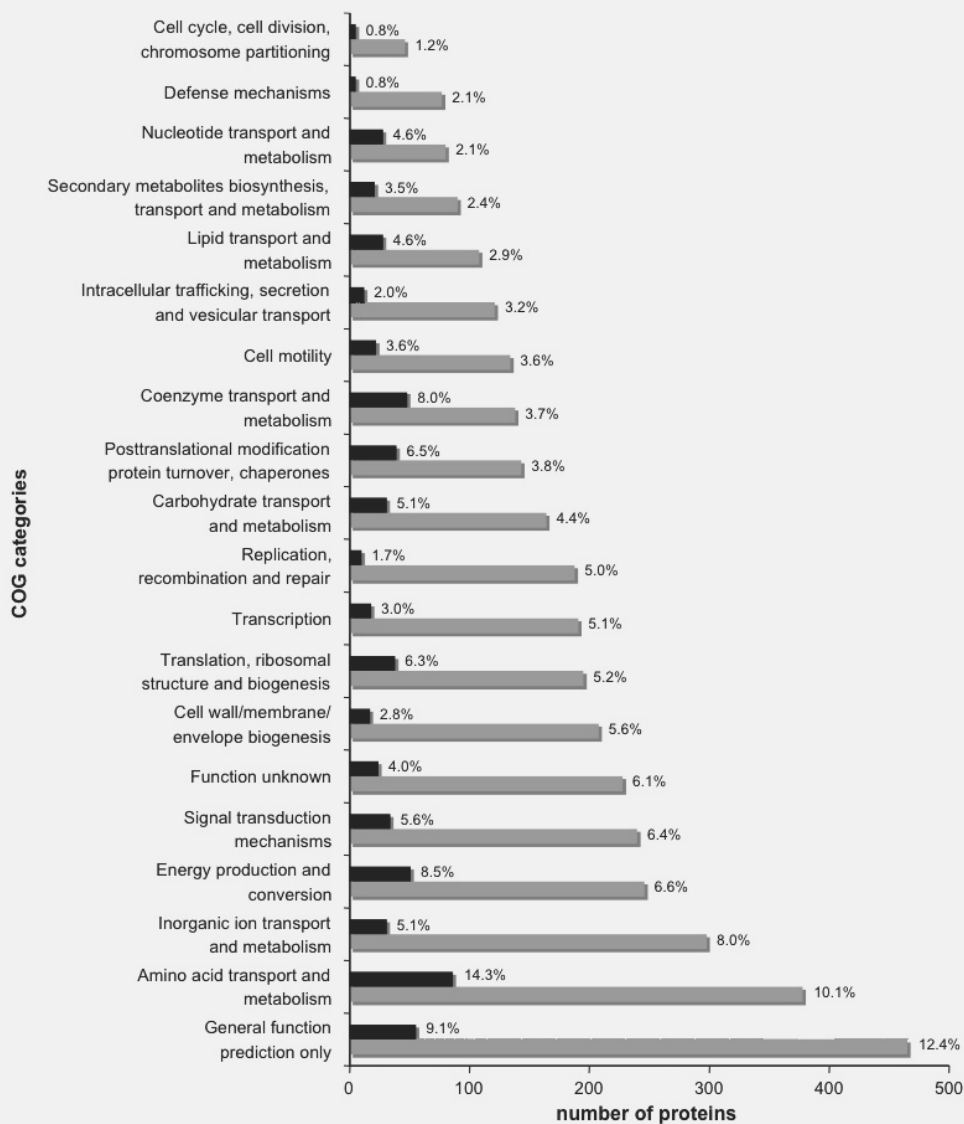


Fig. 3. Classification of proteins in COG categories. Gray bars represent the total number of predicted proteins for each category in the sequenced genome, black bars are the proteins identified with both 2D-PAGE workflow and SDS-PAGE. These groups are represented in an increasing order proportional to their abundance in the predicted proteome (%), which allows an easier abundance comparison in the experimental proteome. It should be noticed that some single proteins were classified into several groups according to COGNITOR.

predicted proteome) or in coenzyme metabolism (8% in experimental data versus 3.7% in the predicted proteome) were well represented. A similar phenomenon was observed for lipid and nucleotide metabolism, which suggests that enzymes involved in general metabolism are overexpressed on the experimental proteomic map. The Codon Adaptation Index [25] measures the potential of expression of any nucleotide sequence. This index has a value comprised between 0 and 1 and is dependent on the codon usage of the strain of interest. The closer this value is to 1, the more the gene may be highly expressed. To test if the overrepresentation of several proteins on the experimental map may be correlated with an unbalance in favor of proteins that have a high CAI value in *H. arsenicoxydans*, such values were calculated from genome sequence data (Fig. 4, in blue). Their repartition showed a maximum peak for values comprised between 0.6 and 0.7. The graphical representation of the calculated CAI values for our experimental proteomic data (Fig. 4, in red) revealed only a small number of identified proteins with a CAI value lower than 0.6. Such a distribution suggests the existence of an experimental unbalance in favor of the most expressed genes.

H. arsenicoxydans is a chemo-organotroph that mainly produces energy from aerobic respiratory process. Lactate has been defined as the optimal carbon source and electron donor promoting *H. arsenicoxydans* growth [11]. The use of lactate requires a lactate permease, allowing entry of the carbon source into the cell. The expert annotation of the genome suggests the presence of two genes possibly encoding a lactate permease (HEAR1916 and HEAR2485). The product of only one of them was experimentally identified (HEAR1916) on our 2D-PAGE

map, suggesting that this protein may be responsible for lactate uptake. Lactate may be then transformed to pyruvate, by the predicted cytoplasmic lactate dehydrogenase (HEAR1710, that shows a *pI* of 8.64 and is not resolved on our 2D-PAGE), and next to phosphoenolpyruvate (PEP) by HEAR1352 identified on our proteome map (Supplementary Tables S1 and S2). Five (HEAR2188, 2610, 0850, 2609, 0572) among the six enzymes required to synthesize fructose 6-P were identified on the proteome map, suggesting that PEP was transformed through gluconeogenesis, further supporting the use of lactate as a carbon source by *H. arsenicoxydans* [11]. On the other hand, the HEAR0746 and HEAR0747 enzymes are both expressed, suggesting that pyruvate has also been used to synthesize acetyl-CoA, a key intermediate of the TCA cycle. Eight enzymes involved in this metabolic pathway were identified on our proteome map (HEAR1771, 1772, 1773, 1775, 1780, 1783, 0343 and 3270), whereas only three proteins predicted to be involved in this pathway were missing on the proteome map (HEAR2526, 3467 and 3327). Finally, at least one protein was found in each step of the respiratory chain, e.g. NADH-ubiquinone oxidoreductase NuoD (HEAR1823), succinate dehydrogenase SdhB (HEAR1775), ubiquinol-cytochrome *c* reductase PetA (HEAR3052), *ccb3*-type cytochrome *c* oxidase CcoP (HEAR1641), ubiquinol oxidase subunit 2 precursor CyoA (HEAR2702) and *aa3*-type cytochrome *c* oxidase CtaC (HEAR2919). These observations further support the role of lactate as an electron donor in *H. arsenicoxydans* aerobic metabolism [11].

3.4. Analysis of the arsenic-specific metabolic pathways

Several proteins identified on our proteome map are involved in arsenic metabolism, including the *ars* encoded proteins. Four *ars* operons have been found in the genome of *H. arsenicoxydans* and proteins experimentally identified were encoded by three of them, i.e. *arsI*, *arsII* and *arsIV*. The *arsII* and *arsIV* loci are localized in genomic regions that were most probably acquired by horizontal gene transfer and carry genes involved in cell response to particular conditions [5]. Indeed, *H. arsenicoxydans* has a GC content of 54.29%. However, eight regions presenting irregular GC contents have been identified [5]. The number of CDS predicted in these various regions, the number of proteins experimentally identified and the percentage of accumulated versus predicted proteins were calculated for each of them (Table 1). Except for the low GC regions 1, 3 and 7, the percentage of expressed proteins was significantly lower in the irregular GC regions, i.e. a mean of 2.5%, as compared to 15.1% in the remaining part of the genome (Table 1). The regions 1 and 7 encode the *arsII* and *arsIV* operon, respectively, and the region 3 encodes enzymes involved in exopolysaccharide biosynthesis (Supplementary Fig. S3).

To know whether proteins encoded by the *arsIII* operon were missing in our previous differential analysis [5,10], we enlarged this analysis using Coomassie staining or pH3–10 gradient (Supplementary Table S4). As expected, we identified the proteins involved in arsenic reduction or extrusion (Table 2),

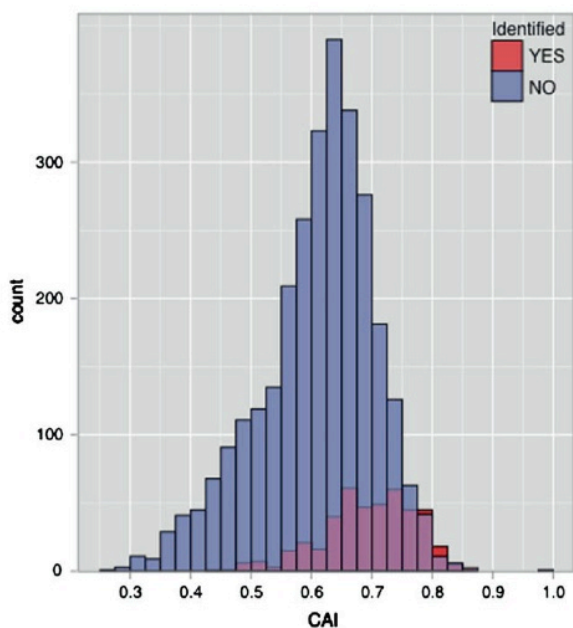


Fig. 4. Number of identified proteins in each range of CAI value, compared to the theoretical proteome. Blue: proteins of the complete theoretical proteome, red: proteins identified in experimental 2D- and SDS-PAGE proteome.

Table 2
Arsenic-induced proteins identified in proteome map analysis, in differential 2D-PAGE analysis, and in differential transcriptome analysis

Gene number (HEAR)	Theoretical M_r/pI	Protein function (gene name)	Fold induction		pI gradient used	Reference
			Protein	RNA		
<i>A – Cell envelope and cellular processes</i>						
0500	18470/6.07	Arsenate reductase (<i>arsC2a</i>)	>10	>10	4–7	This work, [5]
0502	15625/5.20	Arsenate reductase (<i>arsC2b</i>)	>10	>10	4–7	This work, [5]
0503	28290/8.86	Putative flavoprotein (<i>arsH</i> -like)	3	>10	4–7	This work, [5]
1513	15804/9.22	Putative universal stress protein (<i>uspA</i>)	>10		3–10	This work
1883	18004/9.74	Flagellar FliL protein (<i>fliL</i>)	4		4–7	[5]
2338	15162/5.68	Putative universal stress protein (<i>uspA</i>)	4		3–10	This work
2584	25143/6.17	3-demethylubiquinone-9 3-methyltransferase (<i>ubiG</i>)	3		4–7	[10]
2648	19682/4.81	HSP-70 cofactor (<i>grpE</i>)	2		3–10	This work
2809	35890/5.38	D-alanine-D-alanine ligase B (<i>ddlB</i>)	3		4–7	[10]
3207	18402/5.46	Arsenate reductase (<i>arsC</i>)	>10	>10	4–7	This work, [5]
3208	28144/5.89	Arsenical resistance protein (<i>arsH</i>)	3	>10	4–7	This work, [5]
3303	12209/5.44	Arsenical resistance transcriptional regulator (<i>arsR</i>)	>10	>10	4–7	This work, [5]
<i>B – Transport and binding protein</i>						
0490	31431/5.73	Phosphate import ATP-binding protein (<i>pstB</i>)	3		4–7	This work
1000	41561/9.30	Putative TRAP dicarboxylate transporter	6		4–7	This work
1107	36577/9.14	Phosphate-binding periplasmic protein precursor (<i>pstS2</i>)	4		3–10	This work
1111	26473/5.40	Phosphate transport system protein PhoU (<i>phoU2</i>)	–2		4–7	[10]
3439	26537/8.76	Molybdate-binding periplasmic protein precursor (<i>modA</i>)	6		3–10	This work
<i>C – Information and regulation pathway</i>						
0248	45491/6.08	Tyrosyl-tRNA synthetase (<i>tyrS</i>)	>10		4–7	[10]
0491	24701/5.13	Putative phosphate uptake regulator (<i>phoU</i> -like)	4		4–7	This work
0807	12166/5.14	Peptidyl-prolyl cis-trans isomerase (<i>fbp2</i>)	>10		3–10	This work
1278	26299/6.09	23S rRNA (guanosine-2'-O)-methyltransferase (<i>rlmB</i>)	3		3–10	This work
2708	24065/5.39	L-isoaspartate O-methyltransferase	3		4–7	[10]
3169	77404/5.40	Elongation factor G (EF-G) (<i>fusA</i>)	3		4–7	This work
<i>D – Intermediary metabolism</i>						
0343	41623/5.22	Succinyl-CoA synthetase, beta subunit (<i>sucC</i>)	>10		4–7	[10]
3064	26969/5.05	Imidazole glycerol phosphate synthase, subunit F (<i>hisF</i>)	2		4–7	[10]
<i>E – Energy metabolism</i>						
1822	17569/5.07	NADH-quinone oxidoreductase subunit E (<i>nuoE</i>)	>10		4–7	This work
3408	18824/5.04	Putative ATPase delta chain (<i>atpH</i>)	2		4–7	This work
<i>F – Transcription</i>						
1829	75923/5.26	Polyribonucleotide nucleotidyltransferase (<i>pnp</i>)	>10		4–7	This work

enhance the genome annotation quality, with at least two methods to detect erroneous start codon: the presence of an additional peptide upstream of the predicted start, which leads to a longer CDS prediction, or a high discrepancy between theoretical and experimental M_r/pI , resulting in a shorter CDS. In our study, the start codon of 1% of the predicted CDS was changed, which further supports the high quality of the expert annotation of *H. arsenicoxydans* genome.

13.6% of the total proteome was identified in our proteomic experiments, which is an usual percentage for this type of studies [27,28]. Moreover, all COG categories and all fractions (inner-membrane, outer-membrane, cytosolic or periplasmic proteins) were represented in the experimental proteome. All together, these observations suggest that even if this data are not exhaustive, our established proteome map is representative of the theoretical proteome. Therefore, the proteome map data may be useful to deduce which of the predicted metabolic pathways (deduced from genome sequence) is expressed, in particular when *H. arsenicoxydans* is cultivated in the presence of arsenic. The over-representation of proteins involved in general metabolism, like amino acid metabolism, can be explained by the fact that bacteria were grown in minimal

medium which required macromolecule biosynthesis. Similarly, enzymes involved in energy metabolism are required to ensure optimal growth, which could explain their over-representation in our experimental proteome. On the other hand, an unbalance in favor of proteins sharing a high CAI value was observed, suggesting that the overexpression of the corresponding genes was independent on the presence of arsenic but rather due to codon usage. The presence of arsenic did not dramatically modified *H. arsenicoxydans* cell metabolism, which further supports its high level of adaptation to this toxic metalloid. The proteome map revealed that not only the arsenic-specific proteins, but also proteins involved in other adaptive processes were accumulated. This led us to propose a schematic representation of the *H. arsenicoxydans* arsenic response (Fig. 5).

First, the entry of As(III) into the cell is not yet completely understood. Some studies suggest that this arsenic form enters via a GlpF aquaglyceroporin in *E. coli* [29], but no GlpF homologous protein was identified either on the *H. arsenicoxydans* proteome map, or in the theoretical proteome. As(III) may therefore enters into the cell through another transport component. As(V) is a structural analog of

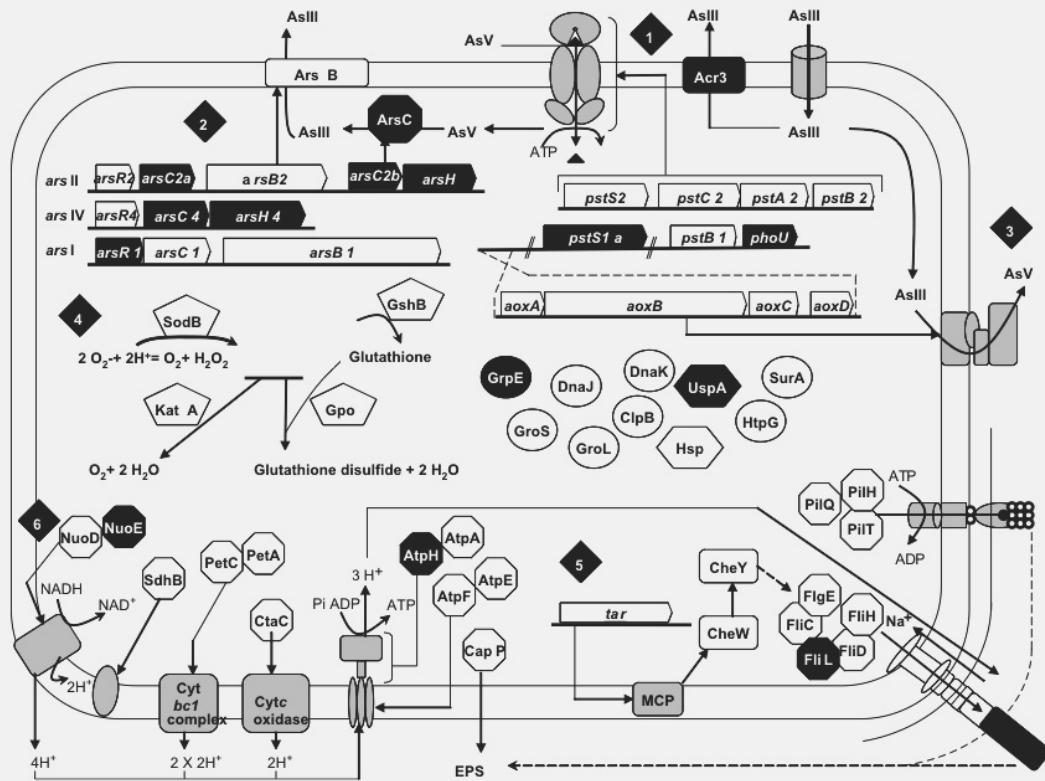


Fig. 5. Cellular model of the *H. arsenicoxydans* arsenic response after exposure to As(III). Cellular events are represented on this model according to proteomic results. Genes and/or proteins induced in the presence of As(III) are in black front.

phosphate, and we previously hypothesized the existence of a link between arsenate and phosphate transport [5]. Nine proteins involved in phosphate import were identified on the proteome map, and the accumulation of some of these proteins in the presence of arsenic was shown (Table 2). These observations as well as the localization of *pst* genes nearby the *arsII* operon further support the possible link between arsenate and phosphate metabolism. Moreover, two putative phosphate uptake regulator (*phoU*-like) proteins were differently regulated in response to arsenic: the PhoU-like protein encoded by the gene localized in the vicinity of *arsII* was induced (Table 2), whereas the HEAR1111 paralog encoding gene was previously shown to be repressed by arsenic [10]. This observation suggests that a complex regulation of the two *pst* operons may allow *H. arsenicoxydans* to adapt rapidly to the presence of arsenic. Indeed, in the presence of arsenite, As(III) is oxidized to As(V) by arsenite oxidase. Consequently, the high affinity phosphate transport system may be overexpressed while, at the same time, the low affinity phosphate transport system may be repressed. This would favor the transport of phosphate and block the entry of As(V) [5].

Second, the main defense mechanism against As(III) exposure results from the induction of the *ars* operon (Fig. 5). Three *ars* operon are induced in the presence of arsenic ([5,10], Table 2), each of them encoding the repressor ArsR, the arsenical pump ArsB, the arsenate reductase ArsC and the

putative flavoprotein ArsH. ArsC reduces As(V) to As(III), which is then extruded by ArsB or Acr3 efflux pump.

Third, the AoxAB arsenite oxidase (AoxA/AoxB/AoxC/AoxD) which catalyses As(III) oxidation to As(V), is involved in the partial detoxification of As(III). Indeed, previous studies revealed that arsenite oxidase allows bacteria to resist to higher level of As(III) [5]. Neither AoxA nor AoxB polypeptides was identified in our proteome map. The localization of these proteins could not be predicted with the pSORTb program. It has been previously proposed that the arsenite oxidase is anchored in the inner-membrane but exposed to the periplasmic side [6]. A careful fractionation of periplasmic, inner-membrane and outer-membrane proteins may help to reveal additional proteins and to precisely determine their localization. This strategy will be used in a future work in order to resolve the arsenite oxidase localization. Nevertheless, the absence of this protein was not due to a specific lost of membrane or periplasmic proteins, since such proteins were identified either by 1D- or 2D-PAGE and were well-represented in our proteome map (Fig. S2). Alternatively, previous studies revealed a weak abundance of the *aoxA* transcript [10]. The arsenite oxidase may be poorly synthesized in such a way that its amount was not sufficient to detect it on our proteome map.

Fourth, arsenic is known to induce oxidative stress in cells. Oxidative damage is a primary toxic effect of As(III)

poisoning in bacteria. O₂^{•-} can react with reduced glutathione and many [4Fe–4S]-containing enzymes, reducing the protection against As(III) exposure. Due to the presence of arsenic in the growth medium, many proteins involved in oxidative stress were identified on the proteome map. The GshB protein is an enzyme essential for glutathione synthesis. In *E. coli*, a mutant lacking the *gshB* gene was found to be more sensitive to As(V) than the wild-type strain [30]. Moreover, several chaperones were identified on the proteome map, including GrpE and UspA that were accumulated in the presence of arsenic. These observations suggest that these proteins may be required to prevent damages resulting from the exposure to arsenic.

Fifth, two proteins, probably involved in exopolysaccharide (EPS) production were accumulated, in agreement with previous data demonstrating that EPS synthesis is induced in response to arsenic, and probably plays a major role in resistance to this toxic element by scavenging nanoparticle inside the EPS matrix [5]. Intriguingly, in this study, no proteins of the putative EPS operon previously described [5] were identified. In contrast, a protein CapE, from an operon, which has been shown in synteny with the capsular biosynthesis operon in *Staphylococcus aureus*, was expressed. In *H. arsenicoxydans*, the operon was located in the low GC region 3 and then probably acquired by horizontal transfer. In addition, *H. arsenicoxydans* motility has been shown to be increased in the presence of As(III) [5]. PilI, PilJ and PilH proteins, which are involved in twitching motility, were identified on the proteome map. These two processes, i.e. motility and exopolysaccharide biosynthesis, should play a key role in the colonization by *H. arsenicoxydans* of its environment.

Finally, all these processes require energy, and the oxidative phosphorylation is an important pathway for bacteria growing aerobically. Proteins involved in each step of this metabolism were identified in this study (Fig. 5, Supplementary Tables S1 and S2).

5. Conclusion

Our data suggest that multiple proteins are expressed when cells are exposed to arsenic. They are involved in oxidation-reduction, efflux, stress response, motility, exopolysaccharide synthesis, and energy metabolism. This suggests that a complex adaptive process is required to protect cells against arsenic stress. With this respect, the combination of proteome and genome data allowed us to highlight major properties of *H. arsenicoxydans* in the context of its adaptation to toxic environments.

Acknowledgements

Stéphanie Weiss and Jessica Cleiss were supported by a grant from the French Ministry of Education and Research and we thank the CNRS and the Bruker Daltonics Company for Christine Carapito's PhD fellowship. All MS/MS spectrometry mass analyses were performed in the Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DSA, UMR7178

CNRS/ULP, and MALDI-TOF spectrometry analysis in the Plateforme Protéomique Esplanade, IBMC-IFR1589. Financial support came from the Institut Pasteur, the Université Louis Pasteur and the Centre National de la Recherche Scientifique. This work was performed within the framework of the 'Groupement de recherche: Métabolisme de l'Arsenic chez les Procaryotes' (GDR2909-CNRS).

Appendix. Supplementary material

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.biochi.2008.07.013

References

- [1] R.S. Oremland, J.F. Stolz, The ecology of arsenic, *Science* 300 (2003) 939–944.
- [2] R. Mukhopadhyay, B.P. Rosen, L.T. Phung, S. Silver, Microbial arsenic: from geocycles to genes and enzymes, *FEMS Microbiol. Rev.* 26 (2002) 311–325.
- [3] A.H. Smith, P.A. Lopipero, M.N. Bates, C.M. Steinmaus, Public health: enhanced: arsenic epidemiology and drinking water standards, *Science* 296 (2002) 2145–2146.
- [4] D. Muller, D.D. Simeonova, P. Riegel, et al., *Herminiimonas arsenicoxydans* sp. nov., a metalloresistant bacterium, *Int. J. Syst. Evol. Microbiol.* 56 (2006) 1765–1769.
- [5] D. Muller, C. Médigue, S. Koechler, et al., A tale of two oxidation states: bacterial colonization of arsenic-rich environments, *PLoS Genet.* 3 (2007) e53.
- [6] D. Muller, D. Lièvreumont, D.D. Simeonova, J.-C. Hubert, M.-C. Lett, Arsenite oxidase *aox* genes from a metal-resistant β -proteobacterium, *J. Bacteriol.* 185 (2003) 135–141.
- [7] A.D. Sarma, D.W. Emerich, Global protein expression pattern of *Bradyrhizobium japonicum* bacteroids: a prelude to functional proteomics, *Proteomics* 5 (2005) 4170–4184.
- [8] D.J. Lessner, L. Li, Q. Li, et al., An unconventional pathway for reduction of CO₂ to methane in CO₂-grown *Methanosarcina acetivorans* revealed by proteomics, *Proc. Natl. Acad. Sci.* 103 (2006) 17921–17926.
- [9] S. Markert, C. Arndt, H. Felbeck, et al., Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*, *Science* 315 (2007) 247–250.
- [10] C. Carapito, D. Muller, E. Turlin, et al., Identification of genes and proteins involved in the pleiotropic response to arsenic stress in *Caeniobacter arsenoxydans*, a metalloresistant β -proteobacterium with an unsequenced genome, *Biochimie* 88 (2006) 595–606.
- [11] W. Weeger, D. Lièvreumont, M. Perret, et al., Oxidation of arsenite to arsenate by a bacterium isolated from an aquatic environment, *BioMetals* 12 (1999) 141–149.
- [12] T. Rabilloud, C. Valette, J. Lawrence, Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension, *Electrophoresis* 15 (1994) 1552–1558.
- [13] S. Carr, R. Aebersold, M. Baldwin, A. Burlingame, K. Clauser, A. Nesvizhskii, The need for guidelines in publication of peptide and protein identification data: working group on publication guidelines for peptide and protein identification data, *Mol. Cell. Proteomics* 3 (2004) 531–533.
- [14] M.R. Wilkins, R.D. Appel, J.E.V. Eyk, et al., Guidelines for the next 10 years of proteomics, *Proteomics* 6 (2006) 4–8.
- [15] J.L. Gardy, M.R. Laird, F. Chen, et al., PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis, *Bioinformatics* 21 (2005) 617–623.

- [16] G.K. Smyth, T.P. Speed, Normalization of cDNA microarray data, *Methods* 31 (2003) 265–273.
- [17] P. Delmar, S. Robin, D. Tronik-LeRoux, J.-J. Daudin, Mixture model on the variance for the analysis of gene expression data, *J.R. Stat. Soc. C* 54 (2005) 31–50.
- [18] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple hypothesis testing under dependency, *Ann. Stat.* 29 (2001) 1165–1188.
- [19] A. Tebbe, C. Klein, B. Bisle, et al., Analysis of the cytosolic proteome of *Halobacterium salinarum* and its implication for genome annotation, *Proteomics* 5 (2005) 168–179.
- [20] I. Mailet, P. Berndt, C. Malo, et al., From the genome sequence to the proteome and back: evaluation of *E. coli* genome annotation with a 2-D gel-based proteomics approach, *Proteomics* 7 (2007) 1097–1106.
- [21] A. Lèvine, F. Vannier, C. Absalon, et al., Analysis of the dynamic *Bacillus subtilis* Ser/Thr/Tyr phosphoproteome implicated in a wide variety of cellular processes, *Proteomics* 6 (2006) 2157–2173.
- [22] S. Morandell, T. Stasyk, K. Grosstessner-Hain, et al., Phosphoproteomics strategies for the functional analysis of signal transduction, *Proteomics* 6 (2006) 4047–4056.
- [23] K. Pethe, P. Bifani, H. Drobecq, et al., Mycobacterial heparin-binding hemagglutinin and laminin-binding protein share antigenic methyllysines that confer resistance to proteolysis, *Proc. Natl. Acad. Sci.* 99 (2002) 10759–10764.
- [24] R.L. Tatutov, D.A. Natale, I.V. Garkavtsev, et al., The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res.* 29 (2001) 22–28.
- [25] P.M. Sharp, W.-H. Li, The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res.* 15 (1987) 1281–1295.
- [26] S. Li, R. Waters, *Escherichia coli* strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination, *J. Bacteriol.* 180 (1998) 3750–3756.
- [27] G. Bernardini, G. Renzone, M. Comanducci, et al., Proteome analysis of *Neisseria meningitidis* serogroup A, *Proteomics* 4 (2004) 2893–2926.
- [28] N. Hansmeier, T.C. Chao, S. Daschkey, et al., A comprehensive proteome map of the lipid-requiring nosocomial pathogen *Corynebacterium jeikeium* K411, *Proteomics* 7 (2007) 1076–1096.
- [29] Y.L. Meng, Z. Liu, B.P. Rosen, As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*, *J. Biol. Chem.* 279 (2004) 18334–18341.
- [30] K.L. Oden, T.B. Gladysheva, B.P. Rosen, Arsenate reduction mediated by the plasmid-encoded ArsC protein is coupled to glutathione, *Mol. Microbiol.* 12 (1994) 301–306.
- [31] P.N. Bertin, C. Medigue, P. Normand, Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems, *Microbiol.* 154 (2008) 347–359.

C/ Discussion

Jusqu'ici, la grande majorité des études concernant le stress arsénié avait été réalisé sur des lignées cellulaires eucaryotes. De nombreuses études mettent ainsi en évidence la production de radicaux libre (ROS : Reactive Oxygen Species) en présence d'As(III). Ces ROS seraient responsables du stress et de la toxicité de l'arsenic. En effet, les effets toxiques de l'arsenic peuvent être réduits par l'ajout de système antioxydant de nature enzymatique (superoxyde dismutase, catalase ou peroxydase) ou non enzymatique (glutathion réduit) (Del Razo *et al.*, 2001).

Dans ce chapitre, nous avons montré pour la première fois que les enzymes impliquées dans les mécanismes de réponse au stress oxydatif, tels que la catalase ou la glutathione réductase, sont également induites en présence d'As(III) chez les procaryotes.

Par ailleurs, l'arsenic est connu pour induire un grand nombre de protéines de stress telles les « heat shock proteins » (Hsp) chez les eucaryotes. Nous avons également démontré cette induction chez les procaryotes, notamment les protéines GrpE et UspA.

Différents mécanismes viennent s'ajouter à la réponse au stress arsénié chez les procaryotes. Il s'agit des systèmes de résistance à l'arsenic (Ars) et d'autres non encore décrits tels que le transport de phosphate (Pst), la phosphorylation oxydative (Nuo, Atp) ou encore le chimitotactisme (Che) et la motilité (Fli). L'analyse protéomique couplée à l'analyse transcriptomique a ainsi permis de caractériser diverses protéines impliqués dans la réponse à l'arsenic et de démontrer que l'adaptation au stress arsénié est un processus complexe chez *H. arsenicoxydans*.

Cinétique du stress arsénié chez
Herminiimonas arsenicoxydans

A/ Introduction

Dans le second chapitre il a été montré que la réponse au stress arsénié est un processus complexe chez *H. arsenicoxydans*. Toutefois, cette approche ne permet d'avoir qu'une image partielle de la réalité. En effet, les bactéries dans leur habitat naturel ne sont pas sujettes à une croissance exponentielle continue. Une réponse au stress chez les micro-organismes implique donc souvent différentes phases de croissance.

Aussi, l'objectif de ce chapitre a été de décortiquer la cinétique de réponse à l'As(III) chez *H. arsenicoxydans*. Pour ce faire, la cinétique des réactions d'oxydoréduction a été établie puis d'après ces données, les transcriptomes après induction à l'As(III) de 15 min, 6 et 8 h, ont été analysés et hiérarchisés grâce au logiciel ArrayMiner®, afin d'identifier des groupes de gènes qui varient de la même manière en réponse à l'As(III). Ces résultats ont été complétés par des tests physiologiques ou biochimiques.

Ma participation à cette étude a portée dans un premier temps, sur l'extraction des ARNs de la souche après les différents temps d'induction ou non à l'As(III). La construction des puces à ADN permettant d'analyser le transcriptome, ainsi que la hiérarchisation des données ont ensuite été effectuées en collaboration avec l'Institut Pasteur. J'ai ensuite pris en charge l'ensemble de cette étude : aussi bien l'analyse des données des transcriptomes et leur interprétation, les tests physiologiques et biochimiques, l'élaboration des figures et la rédaction du manuscrit.

L'ensemble de ces travaux fait l'objet d'une publication soumise (Cleiss-Arnold *et al.*, submitted), présentée ci-dessous.

B/ Manuscrit

Temporal transcriptomic response during arsenic stress in *Herminiimonas arsenicoxydans*

Jessica Cleiss-Arnold¹, Sandrine Koechler¹, Caroline Proux², Marie-Laure Fardeau³, Marie-Agnès Dillies², Jean-Yves Coppee², Florence Arsène-Ploetze¹ and Philippe N. Bertin^{1,*}

Address : ¹UMR7156 Génétique Moléculaire, Génomique et Microbiologie, CNRS
Université de Strasbourg, 28 rue Goethe, 67000 Strasbourg, France

² Plate-forme technologique Puces à ADN, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris
cedex 15, France

³ IRD Laboratoire de Microbiologie, ESIL, 163 Avenue de Luminy 13288 Marseille cedex
9, France

Email : Jessica Cleiss-Arnold - cleiss.j@unistra.fr ; Sandrine Koechler -
sandrine.koechler@unistra.fr ; Caroline Proux - karo@pasteur.fr ; Marie-Agnès Dillies -
marie-agnes.dillies@pasteur.fr ; Marie-Laure Fardeau - marie-laure.fardeau@univmed.fr ;
Florence Arsène-Ploetze - ploetze@unistra.fr ; Jean-Yves Coppee - jycoppee@pasteur.fr ;
Philippe N. Bertin – philippe.bertin@unistra.fr.

* Corresponding author

Abstract

Background : Arsenic is present in numerous ecosystems and microorganisms have developed various mechanisms to live in such hostile environments. *Herminiimonas arsenicoxydans*, a bacterium isolated from arsenic contaminated sludge, has acquired remarkable capabilities to cope with arsenic. In particular our previous studies have suggested the existence of a temporal induction of arsenite oxidase, a key enzyme in arsenic metabolism, in the presence of As(III).

Results : Microarrays were designed to compare gene transcription profiles under a temporal As(III) exposure. Transcriptome kinetic analysis demonstrated the existence of two phases in arsenic response. The expression of approximately 14% of the whole genome was significantly affected by an As(III) early stress and 4% by an As(III) late exposure. The early response was characterized by arsenic resistance, oxidative stress, chaperone synthesis and sulfur metabolism. The late response was characterized by arsenic metabolism and associated mechanisms such as phosphate transport and motility. The major metabolic changes were confirmed by chemical, transcriptional, physiological and biochemical experiments. These early and late responses were defined as general stress response and specific response to As(III), respectively.

Conclusion : Gene expression patterns suggest that the exposure to As(III) induces an acute response to rapidly minimize the immediate effects of As(III). Upon a longer arsenic exposure, a broad metabolic response was induced. These data allowed to propose for the first time a kinetic model of the As(III) response in bacteria.

Bacteria live in changing environments and are subjected to a variety of environmental stresses such as pH, temperature, osmolarity or heavy metals. Arsenic is found in numerous disturbed or natural ecosystems where it can exist in multiple oxidation states, the most common being arsenite As(III) and arsenate As(V) [1]. This metalloid is known to generate oxidative stress in cells by its capability to induce the formation of reactive oxygen species (ROS). The damages caused by ROS to lipids, proteins and DNA are likely to contribute to As toxicity [2, 3]. In addition, one property of arsenite consists in a high reactivity with sulphhydryl groups and this binding affects the activity of many proteins. Microorganisms have developed remarkable capabilities to cope with arsenic. The most common arsenic resistance mechanism depends on the presence on plasmid or chromosome of *ars* genes encoding proteins involved in the reduction and/or the efflux of the toxic element. In addition, arsenic stress is associated in various microbial species with arsenite oxidase activity, biofilm formation, motility, oxidative stress or sulfur assimilation [4, 5]. For example, the biofilm development by *Thiomonas arsenivorans* has been described as a physical barrier decreasing As(III) access to sessile cells [6]. Remarkably, some organisms such as *Rhizobium sp.* NT-26, have evolved specific metabolic pathways allowing them to oxidize As(III) as an energy source [7, 8] or others use As(V) in anaerobic respiration [9]. In the heterotrophic prokaryote *Herminiimonas arsenicoxydans*, genome sequencing revealed the presence of four *ars* operons involved in the reduction of As(V) to As(III) and of one *aoxAB* operon involved in the oxidation of the most toxic form, As(III) to the less toxic form, As(V) [10]. In addition to this detoxification process, this bacterium exhibits positive chemotaxis and motility towards arsenic, and metalloid scavenging by exopolysaccharides. The availability of the *H. arsenicoxydans* complete genome sequence offers an opportunity to study its physiology by functional genomic approaches [10]. Our previous transcriptomic studies have demonstrated that a large number of genes encoding proteins involved in

oxidative stress, import of phosphate or DNA repair are induced after 15 min As(III) exposure. However, no variation was found in the genes coding for arsenite oxidase, a key enzyme in arsenic response [11] recently shown to be controlled by a complex regulation [12]. In addition, little is known regarding the kinetics of arsenic stress response in microorganisms. To address this process, physiological analyses coupled with Western immunoblotting experiments and DNA microarrays were used to examine the temporal changes in transcriptome profiles during the transition from As(III) to As(V) species due to As(III) oxidation. Our work represents, to our knowledge, the first kinetic analysis of transcription pattern in bacteria exposed to arsenic, leading to propose a global model of arsenic response in *H. arsenicoxydans*.

Results and Discussion

Characterization of arsenic oxidoreduction kinetics

To study the oxidoreduction kinetic in *H. arsenicoxydans*, the two arsenic species (As(III) and As(V)) were quantified by HPLC-ICP-AES from filtered culture supernatants at several times after As(III) or As(V) induction (0, 15 min, 1, 2, 4, 6 and 8 hours) (Figure 1). In the presence of As(III) (Figure 1A), three distinct phases were observed : in phase A1 (early exposure : 0 to 4 hours), no oxidation was observed since arsenic was present as 100% As(III). In phase A2, which corresponds to a partial oxidation of As(III) into As(V), both species were present, a proportion of 50% As(III) and 50% As(V) being observed after 6 hours. The transformation of As(III) into As(V) was complete after 8 hours (phase A3).

Overview of transcriptome profiles

To get insight into the mechanisms involved in arsenic response, including oxidoreduction processes, transcriptomic analyses were performed at different steps of the experimental time course. Cells were collected from cultures of *H. arsenicoxydans* induced 15 min (early), 6 hours (intermediary), 8 hours (late) or not with As(III) (Figure 1A) or As(V) (Figure 1B). Under As(V) stress, 128 genes were up-regulated and 107 genes were down-regulated in the early phase (B1) (7% of the genome). In the late phase (B2) the number of genes up-regulated (239 genes) and down-regulated (161 genes) increased (12% of the genome). Under As(III) stress, the proportion was inverted. The major shift in gene expression occurred during the onset of the early phase (A1) with 214 genes up-regulated and 245 genes down-regulated under As(III) stress. This represents approximately 14% of the genes whose expression was significantly affected by As(III). As cell growth progressed further into the late phase (A3), the number of genes up-regulated (60 genes) and down-regulated (73 genes) decreased, representing 4% of the genome. The large number of genes responding to As(III), especially in the early phase, suggests a major remodelling of cellular physiology. In addition, oxidoreduction kinetic was only observed under As(III) stress (Figures 1A and 1B), suggesting that the Aox-dependent oxidation largely predominate the Ars-dependent reduction process. This study therefore focused on the As(III) response in *H. arsenicoxydans*.

Hierarchical clustering kinetics

To classify the transcripts on the basis of their temporal dynamics after exposure to arsenite, we performed a hierarchical clustering with the ArrayMiner® software (<http://www.optimaldesign.com/ArrayMiner/ArrayMiner.htm>). Genes were sorted into groups that exhibited similar transcription kinetics as cells transitioned from A1 to A3

phases. The set of genes regulated by As(III) shown in Figure 2 was classified into 20 clusters. The profiles similarity of those clusters enabled us to identify two major classes: 1) class I: (clusters 0, 2, 3, 6, 8, 10, 13, 14, 15 and 19) containing genes whose expression increased immediately after 15 min induction and decreased after 6 and 8 hours and class II: (clusters 1, 4, 5, 7, 9, 11, 12 and 16) containing genes whose expression levels increased only after 6 and 8 hours induction. The product of each either upregulated or downregulated gene by a fold change of two or more was identified to gain insight into the changes in physiology (Table 1). Next, to examine how the two classes are distributed with regard to their functions, we further classified these genes according to the Gene Classification based on COG functional categories (Figure 3). The overall gene expression pattern suggests the existence of two major physiologic phases (early and late) during the adaptation to As(III) stress.

i) Early phase. Our microarray analysis highlighted a set of stress responsive genes whose transcription was early affected by As(III). This phase contained the early up-regulated genes whose expression levels gradually decreased after the induction (Class I) or down-regulated genes whose expression levels gradually increased after the induction (Class II).

First, σ^E and several genes encoding proteins involved in the synthesis of cell envelope components were down-regulated (Class II, Figure 3B (regulatory function, protein synthesis and fatty acid)) whereas genes encoding protein involved in transport were up-regulated (Class I, Figure 3A(transport)). The bacterial envelope is a major defense against threats from the environment. Due to its crucial importance, the integrity of the cell envelope, including the presence of outer membrane proteins (OMP), is closely monitored to ensure its functionality. The σ^E system is the major pathway [13] required for proper folding of OMPs and their turnover [14], phospholipids and lipopolysaccharide (LPS) biosynthesis,

signal transduction and the expression of inner and outer membrane proteins [15]. In *Escherichia coli*, the σ^E regulon is induced specifically in response to the unbalanced synthesis of outer membrane proteins [14] and to the misfolding of polypeptides that have been translocated across the cytoplasmic membrane [16]. Our results suggest that both the organization (lipoproteins, peptidoglycan and LPS) and the functioning (transport, signal transduction) of the bacterium may be changed in the early phase of the response to As(III).

Second, early As(III) stress strongly stimulated expression of genes encoding components of sulfur assimilation, sulfur oxidation and glutathione (GSH) biosynthesis pathways. In fact, mRNA levels of genes encoding all the enzymatic steps required for sulfate uptake, its conversion into cysteine and GSH were increased. The expression of these genes was induced about 3 to 22 fold after 15 min induction with As(III). The expression of some of these genes was down-regulated at 6 and 8 hours (Class II, Table 1). In eukaryotic cells, As(III) exposure results in increased levels of reduced glutathione [17] and an As(III)-triglutathione complex has been found in human liver excreta [18], leading to the conclusion that GSH is important in reducing As(III) toxicity in humans. Most of the effects of arsenite result from its interaction with thiol groups of proteins, its inophoretic properties and its ability to directly or indirectly generate free radicals and hence induce oxidative stress [19]. Microorganisms have therefore developed defense mechanisms that either keep the concentration of the O₂-derived radicals at acceptable levels or repair oxidative damages. In our previous proteomic experiments, oxidative stress proteins, such as KatA, SodB, were shown to be induced in response to As(III) stress in *H. arsenicoxydans* [11]. In the present study, one important change that could be observed was the positive regulation of three genes (*sodC*, *katA*, *ahpC*) involved in the protection against oxidative stress. The *sodC* gene encodes the superoxide dismutase catalyzing the dismutation of the superoxide radical to H₂O₂, which is less toxic, and O₂. In addition, *katA* and *ahpC* genes encoding a catalase and

a peroxidase, respectively, are responsible for scavenging endogenously produced H₂O₂. Finally, As(III) early exposure induced various thioredoxin encoding transcripts and a glutaredoxin (Table 1). In *H. arsenicoxydans* the induction of genes encoding proteins involved in sulfate and GSH pathways suggests that GSH may play a key role in the cell defense against oxidative stress and metal toxicity by maintaining an intracellular reducing environment. GSH may detoxify arsenite by i) chelation and then removing of arsenite from the cell, ii) protection against oxidation caused by metals since GSH serves as the main redox buffer in the cell; and iii) binding to reactive sulfhydryl groups of proteins thereby protecting them from irreversible metal binding and/or oxidative damages [20]. Thioredoxin is believed to serve as a cellular antioxidant by maintaining the intracellular redox state, which is greatly influenced by the presence or absence of intracellular ROS, by reducing disulfide bonds produced by various oxidants and by interacting with proteins by direct association or thiol reactivity. Glutaredoxin is also known as a thioltransferase and like thioredoxin possesses an active center disulphide bond. All these cellular activities (GSH, catalase, peroxidase and thioredoxin-related) are thought to be part of a general set of reactions involved in responding to As(III)-mediated production of reactive oxygen species in *H. arsenicoxydans*.

Third, the adaptative responses of bacteria to sudden environmental changes, usually involved the induction of many major heat shock proteins MHSPs, including chaperones, proteases, transcriptional regulators and other structural proteins. As(III) induced the expression of several genes encoding MHSPs (Table 1). The MHSPs play a role in repairing and preventing damages caused by an accumulation of unfolded proteins resulting from diverse environmental stresses. In agreement with this, our microarray data showed that As(III) early exposure induced the transcription of two *uspA* genes that might be involved in protecting bacterial cells against As(III) stress. The protein UspA was also induced by

As(III) in previous proteomic data [11]. The expression of the universal stress protein A (UspA) is known to be induced by a large variety of stress conditions [21], our results show that these conditions include As(III).

Finally, arsenic resistance in *H. arsenicoxydans* is partially mediated by proteins encoded by three different *ars* operons and one *aox* operon [10]. The clusters of *ars* genes encode an ArsR regulator, an As(III) extrusion pump ArsB/Acr3, an ArsH putative flavoprotein and one or two arsenate reductases ArsC. One of the loci contains an Acr3-type transporter, the others associate an ArsB-type transporter with ArsC reductases. ArsC is an enzyme that reduces As(V) to As(III) which can be pumped out by the ArsB or Acr3 membrane proteins. Both *acr3* and *arsB* genes were induced in the presence of As(III). However, Acr3 mRNAs were only detected in 15 min As(III) exposed cells, which demonstrates that Acr3 is specific to the early response. Remarkably, the arsenite oxidase genes were not induced in early phase. The lack of induction of the *aox* genes is in agreement with the detection by HPLC-IC-AES of 100 % As(III) (Figure 1).

ii) *Late phase*. This phase contains genes whose expression levels gradually increased (class II) or decreased (class I), after arsenic exposure. Multiple genes of cellular processes, cell envelope and metabolism (energy, intermediary, DNA...) were highly downregulated in class I (Figure 3B). Class II was characterized by the induction of several functional categories, i.e. genes involved in metabolism, regulatory functions and cellular processes (Figure 3A). These genes are suggested to be specific for the late response to As(III).

First, *H. arsenicoxydans* reacted to a late As(III) exposure by activating genes involved in the transport and assimilation of phosphate as well as other phosphorous compounds (Table 1). Indeed, after late exposure, phosphate specific transporter (Pst) and phosphate inorganic transporter (Pit) were induced. As(V) is a structural analog of

phosphate and we demonstrated here that the oxidation of As(III) to As(V) only occurred in late phase (Figure 1). The strong induction of various genes involved in phosphate uptake may be a response to the accumulation of As(V). Indeed, As(V) produced by the oxidation of As(III) may compete with phosphate. These observations suggest that a complex regulation of the *pst* and *pit* operons allow *H. arsenicoxydans* to maintain its intracellular phosphate level despite the accumulation of As(V). This further supports our previous hypothesis on the physiological link between arsenic and phosphate metabolism [10].

Second, genes involved in chemotactic machinery and flagellum assembly were up-regulated after late exposure to As(III). Our results from the microarray experiments during exposure to As(III) combined to previous observations [10] clearly demonstrate that the chemotactic response is induced by As(III). In *H. arsenicoxydans* the flagellar machinery is organized in a mixed peritrichous/polar cascade which is most probably synthesized through sequential hierarchy of gene activation events initiated by the expression of the master transcriptional regulator FlhDC [10]. The microarray results strongly suggest that motility is gradually induced in response to As(III) in *H. arsenicoxydans*.

Finally, the *aox* operon was induced after late exposure. The *aox* operon contains *aoxA* encoding the small subunit of arsenite oxidase, *aoxB* encoding the large subunit, *aoxC* encoding a nitroreductase and *aoxD* encoding a cytochrome c552. The *aoxRS* operon is involved in the regulation of *aoxABCD* operon [12]. All of them were induced by late As(III) stress (Class II, Table 1), which strongly suggest that arsenite oxidase activity is specific to late exposure.

Transcriptional and physiological analysis of the major metabolic changes

Our microarray data suggest that several genes play a major role in the transition from the absence of As(III) to early or late arsenic exposure. To confirm this hypothesis, chemical, transcriptional, physiological and biochemical experiments were performed.

First, to further support the involvement of glutathione synthesis and sulfate metabolism in early As(III) response (early phase) in *H. arsenicoxydans*, we performed quantitative RT-PCR experiments. Transcripts abundance of *gloA* and *soxC* were compared to two internal controls, i.e. the putative RNA methyltransferase gene and the peptide deformylase gene, in cultures challenged 15 min, 8 hours or not by As(III). The expression of *soxC* and *gloA* mRNA was increased by a 312 and 7 fold factor, respectively, after 15 min As(III) exposure. No such induction was observed in late exposure.

To address the possible role of As(III) in the GSH pathway, the total content of GSH was measured. The pool of GSH increased after 15 min (~3 fold) and further rose over time (5 fold after 6 hours). After 8 hours the level remained unchanged, as compared with 6 hours. These observations suggest that the accumulation of GSH in response to As(III) was specific to the early phase. In addition, genes encoding sulfate uptake were up-regulated after 15 min and down regulated after 6 and 8 hours (Table 1). We therefore evaluated the sulfate uptake by measuring the extracellular sulfate level at several time after As(III) exposure. Interestingly, we observed that the sulfate level remained unchanged after early and late exposure to As(III). These observations suggest that a balance exists between sulfate uptake and biosynthesis in the cell. Similarly, genes involved in the sulfur oxidation pathway were also induced, i.e. *soxC*, *soxD* (Table 1). We can hypothesize that the Sox system, in particular the sulfur dehydrogenase SoxCD [22] is activated in the cell to maintain an optimal intracellular concentration of sulfate. The accumulation of GSH is correlated with the induction of genes involved in sulfate uptake and biosynthesis (early

phase). Consequently, the imported sulfate may be redirected toward GSH biosynthesis in As(III) exposed cells.

On the other hand, to confirm the oxidative stress response resulting from arsenic exposure, peroxidase activity was measured from cultures induced or not, 15 min, 6, 8 hours with As(III). The activity was induced 5 fold after 15 min and remained unchanged after late exposure. This observation demonstrated that peroxidase activity is also specific to the early exposure to As(III).

Second, the induction of arsenite oxidase activity in late phase was demonstrated in this study (Figure 1). The oxidation was complete after 8 hours exposed cells. On agar plate, the activity was detectable after 48 hours growth in presence of As(III), by a silver nitrate staining test (data not shown). The accumulation of arsenite oxidase was tested by Western immunoblotting experiments with anti-AoxB antibodies. AoxB was detected in protein extract from cells induced for 6 and 8 hours. In contrast, no protein was detected after 15 min induction (Figure 4) which explains the lack of *aox* gene expression after 15 min in the presence of As(III) observed in a previous study [11]. Arsenite oxidase activity is therefore specific to late exposure to As(III) in *H. arsenicoxydans*.

Similarly, to support the effect of arsenic on flagellar motility and phosphate transport in late phase, transcripts abundance of *fliC* and *pstSb* was measured. *pstSb* mRNAs increased by a 455 fold factor after 8 hours As(III) exposure but no such induction was observed in early exposure. Remarkably, *fliC* was induced by a 14 or 11 fold change after 15 min or 8 hours, respectively.

The motility of *H. arsenicoxydans* was analyzed in the presence and the absence of As(III). As shown in Figure 5 a swarming diameter of 10 mm was measured for cells

incubated 48 hours without arsenic on the semi-solid medium. In the presence of As(III), cells were mobile on semi-solid medium only after 72 hours incubation (12 mm), corresponding to the late phase in liquid medium. These results suggest that the presence of As(III) delays cell motility. However, no significant difference in the expression of the flagellar genes was observed between early and late exposure (Table 1). Moreover, the expression of *motA* and *motB* genes, encoding proteins responsible for the rotary activity of flagellum motor [23], was not affected by early or late exposure to As(III). These observations clearly demonstrated that the flagellum biosynthesis was not altered in response to As(III), unlike its functioning. In *H. arsenicoxydans*, the rotation of this flagellar appendage is driven by a sodium motive force [10]. Since the oxidation of arsenite releases four protons [24], we proposed that arsenite oxidase activity may indirectly increase motility. The exploration of the genome of *H. arsenicoxydans* revealed the possible synthesis of a Na⁺/H⁺ transporter. The protons released from arsenite oxidation may extrude Na⁺ through this antiporter and generate the electron potential for the Na⁺-driven flagellar motor, as described in the differentiation of flagellar system of *Vibrio alginolyticus* [25]. This hypothesis could explain how arsenite oxidase contributes to the motility of *H. arsenicoxydans* after As(III) late exposure.

Conclusion

Many microorganisms have acquired the capacity to survive in the presence of toxic metals or metalloids by inducing the expression of resistance genes. The specific nature of these resistance mechanisms is the result of a cleverly designed genetic circuit that is tightly controlled by regulatory proteins. However, the multiple physiological responses induced by arsenite stress in *H. arsenicoxydans* are not exclusively associated with the expression of classical arsenic resistance operons. Indeed, in this study we demonstrated that the

adaptation to As(III) is characterized by an early and a late response. A highly sophisticated regulatory network may thus modulate the expression of genes in response to arsenic. These effects are mediated in part through the activation or repression of mRNA transcripts by DNA-binding proteins and signal transduction systems, such as ArsR, AoxRS and others. All these transcriptomic observations are summarized in a kinetic and metabolic model (Figure 6).

The early response is characterized by the induction of genes involved in general stress. First, GSH is a protagonist of the primary cellular responses to As(III). The genes encoding enzymes involved in sulfate uptake (A), sulfur oxidation (B), cysteine synthesis (C) leading to cysteine containing proteins and their conversion in GSH (D) were induced. As(III)-exposed cells may thus channel a large part of the assimilated sulfur into GSH for metal chelation, cellular redox buffering and possibly also protein glutathionylation. As(III) may react with GSH to form an As(SG)₃ complex (E). Interactions of these As-thiol group non specific complexes with molecular oxygen lead to the formation of reactive oxygen species such as H₂O₂, resulting in oxidative stress within the cell. The As(III) is extruded by ArsB or Acr3 (F). Acr3 was demonstrated to be specific to early exposure. The mechanism proposed here is in agreement with the observations made with HPLC-ICP-AES, showing that no oxidation occurred in early As(III) stress and only As(III) species was present.

The late response was characterized by the induction of genes specific for arsenic. Indeed, the GSH response is downregulated (A) and the transcripts encoding arsenic resistance (B), especially arsenite oxidase and its regulation (AoxRS) (C, D), and associated mechanisms such as phosphate uptake (E) and flagellum biosynthesis (F) were upregulated. The arsenite oxidase catalyses the oxidation of As(III) in As(V) (D). As(V) may directly enters to the cell via the Pst or Pit transporter (E), where it is reduced to As(III) by the arsenate reductase and extruded into the periplasm *via* ArsB (F). Then As(V) produced by

the oxidation of As(III) is extruded outside the cell (G). The mechanism proposed here for the late response to As(III) is in agreement with the oxidoreduction kinetics and underlies the links between arsenic, phosphate and motility.

In this study we demonstrated that the detoxification and metabolism processes are gradually expressed in *H. arsenicoxydans* to adapt to arsenic-rich environments. In bacteria, the response to As(III) seems therefore to be characterized by a stereotypic just-in-time transcription of specific metabolic pathways, resulting in a two steps response.

Experimental procedures

Bacterial strains and growth media

H. arsenicoxydans ULPAs1 was grown in a chemically defined medium (CDM), supplemented by 2% agar when required [26]. Tryptone swarm plates containing CDM supplemented with 1% Bacto-Tryptone and 0.25% agar were used to assess bacterial motility.

Arsenic speciation and sulfate determination

H. arsenicoxydans was grown in CDM medium and cultures were induced when required by addition of 0.66 mM As(III) for 15 min, 1 hour, 2, 4, 6 or 8 hours. Similarly, cultures were induced or not for 15 min, 6 hours, 8 hours with 13.3 mM As(V). Culture supernatants were filtered through sterile 0,22 µm pore size filters (VWR). Arsenic species were separated by high-performance liquid chromatography (HPLC) and quantified by inductively coupled plasma-atomic emission spectrometry (ICP-AES), as previously described [27]. Sulfate extracellular concentration was determined by ion chromatography (IC) using a Metrohm Compact IC 761 system. The column used was an MetrosepA SUPP 1-250 (4.6+250 mm) with an eluent concentration of 3 mM sodium carbonate.

Peroxidase and total glutathione assay

Peroxidase activity and total glutathione were measured from serially diluted (10^0 - 10^{-3}) 48 hours-old cultures of *H. arsenicoxydans* exposed or not to As(III). The peroxidase activity was determined using the Amplex Red Hydrogen Peroxide Assay kit (Invitrogen) and the total glutathione concentration was measured using the colorimetric glutathione detection kit (PromoKine). This assay was performed according to the protocol provided by the manufacturer.

Preparation of protein extracts, SDS-PAGE separation

Protein extracts and Western immunoblotting experiments were performed as described previously in Koechler *et al.* [12]

RNA extraction

Strains were grown at 25°C for 24 hours (OD=0,15) and cultures were induced or not by addition of 0.66 mM As(III) for 15 min, 6 hours or 8 hours before extraction. Similarly, 13.3 mM As(V) were added or not to a 24 hours cultures during 15 min, 6 hours or 8 hours. Samples were harvested and stored at -80°C. RNA was extracted as previously described [11]. After extraction procedure, RNA integrity was checked by electrophoregram analysis on a BioAnalyser (Agilent) and total RNA concentration was determined spectrophotometrically with a Nanodrop.

Microarrays and data analysis

Microarrays containing 60-mer oligonucleotides for all predicted *H. arsenicoxydans* genes (<http://www.genoscope.cns.fr/agc/mage/arsenoscope>) were used, as previously described

[11]. Briefly, total RNA (5 µg) was reverse transcribed and indirectly labelled and next hybridized to the arrays. Three distinct biological RNA samples were prepared from each growth condition at different time after induction with As(III) or As(V) (0 min and after 15 min, 6 hours or 8 hours) and labelled either by Cy3 or Cy5 in a dye-swap design. Microarray data were deposited in ArrayExpress (accession E-MEXP-2199 and A-MEXP-1594). Scanning, data normalization and statistical analysis was performed as described [11]. Genes showing a valid *p*-value and a more than two-fold decreased or increased expression were considered as differentially expressed between the conditions tested and were retained for further study.

Clustering analysis was performed using ArrayMiner (<http://www.optimaldesign.com/>) V. 5.3.4 with the Gaussian clustering and Pearson correlation options selected. The data table included log₂ ratios at three different time points (15 min, 6 and 8 hours) with t=0 being a reference. 2349 genes were significantly up- or down-regulated in at least one of the three time points. The number of classes has to be tuned prior to the run. It was chosen to 20, as it appeared to provide homogeneous profiles within classes with a reasonable number of unclassified genes.

Quantitative real time PCR

Quantitative PCR experiments were performed as described by Koechler *et al.* [12], with the following modification of thermocycling conditions : 5min at 95°C and 40 cycles of 15s at 95°C, 15s at 60°C and 1 min at 72°C. Two biological replicates (independent cultures) and two quantitative PCR replicates were performed in each experience. Amplification products were designed in order to obtain less than 175 bp fragment. The pairs of primers used are listed in Table S1.

Acknowledgments

J.C-A. was supported by a grant from the French Ministry of Education and Research. Financial support came from the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (ANR 07-BLAN-0118 project) and the Université de Strasbourg. This work was done in the frame of the Groupement de Recherche (GDR2909-CNRS) : « Métabolisme de l'Arsenic chez les Microorganismes ».

1. Cullen WR, Reimer KJ: **Arsenic speciation in the environment**. *Chemical Reviews* 1989, **89**(4):713-764.
2. Abernathy CO, Liu YP, Longfellow D, Aposhian HV, Beck B, Fowler B, Goyer R, Menzer R, Rossman T, Thompson C *et al*: **Meeting on Arsenic: Health Effects, Mechanisms of Actions, and Research Issues, Hunt Valley, Maryland, 22-24 September 1997**. *Environmental Health Perspectives* 1999, **107**(7):593-597.
3. Mandal BK, Suzuki KT: **Arsenic round the world: A review**. *Talanta* 2002, **58**(1):201-235.
4. Thorsen M, Lagniel G, Kristiansson E, Junot C, Nerman O, Labarre J, Tamos MJ: **Quantitative transcriptome, proteome, and sulfur metabolite profiling of the *Saccharomyces cerevisiae* response to arsenite**. *Physiological Genomics* 2007, **30**(1):35-43.
5. Parvatiyar K, Alsabbagh EM, Ochsner UA, Stegemeyer MA, Smulian AG, Hwang SH, Jackson CR, McDermott TR, Hassett DJ: **Global analysis of cellular factors and responses involved in *Pseudomonas aeruginosa* resistance to arsenite**. *Journal of Bacteriology* 2005, **187**(14):4853-4864.
6. Michel C, Jean M, Coulon S, Dictor MC, Delorme F, Morin D, Garrido F: **Biofilms of As(III)-oxidising bacteria: formation and activity studies for bioremediation process development**. *Applied Microbiology and Biotechnology* 2007, **77**(2):457-467.
7. Santini JM, Sly LI, Schnagl RD, Macy JM: **A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: Phylogenetic, physiological, and preliminary biochemical studies**. *Applied and Environmental Microbiology* 2000, **66**(1):92-97.
8. Santini JM, Kappler U, Ward SA, Honeychurch MJ, vanden Hoven RN, Bernhardt PV: **The NT-26 cytochrome c552 and its role in arsenite oxidation**. *Biochimica et Biophysica Acta - Bioenergetics* 2007, **1767**(2):189-196.
9. Stolz JF, Oremland RS: **Bacterial respiration of arsenic and selenium**. *FEMS Microbiology Reviews* 1999, **23**(5):615-627.
10. Muller D, Medigue C, Koechler S, Barbe V, Barakat M, Talla E, Bonnefoy V, Krin E, Arsene-Ploetze F, Carapito C *et al*: **A tale of two oxidation states: bacterial colonization of arsenic-rich environments**. *PLoS genetics* 2007, **3**(4).
11. Weiss S, Carapito C, Cleiss J, Koechler S, Turlin E, Coppee JY, Heymann M, Kugler V, Stauffert M, Cruveiller S *et al*: **Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data**. *Biochimie* 2009, **91**(2):192-203.

12. Koechler S, Cleiss-Arnold J, Proux C, Sismeiro O, Dillies M-A, Goulhen- Chollet F, Hommais F, Lièvremon D, Arsène-Ploetze D, Coppée J-Y *et al*: **Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. *BMC Microbiology* 2010, **10**:53.**
13. Brooks BE, Buchanan SK: **Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochimica et Biophysica Acta - Biomembranes* 2008, **1778**(9):1930-1945.**
14. Mecsas J, Rouviere PE, Erickson JW, Donohue TJ, Gross CA: **The activity of σ (E), an *Escherichia coli* heat-inducible σ -factor, is modulated by expression of outer membrane proteins. *Genes and Development* 1993, **7**(12 B):2618-2628.**
15. Rezuchova B, Miticka H, Homerova D, Roberts M, Kormanec J: **New members of the *Escherichia coli* σ^E regulon identified by a two-plasmid system. *FEMS Microbiology Letters* 2003, **225**(1):1-7.**
16. Missiakas D, Betton JM, Raina S: **New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Molecular Microbiology* 1996, **21**(4):871-884.**
17. Ochi T: **Arsenic compound-induced increases in glutathione levels in cultured Chinese hamster V79 cells and mechanisms associated with changes in γ -glutamylcysteine synthetase activity, cystine uptake and utilization of cysteine. *Archives of Toxicology* 1997, **71**(12):730-740.**
18. Kala SV, Neely MW, Kala G, Prater CI, Atwood DW, Rice JS, Lieberman MW: **The MRP2/cMOAT transporter and arsenic-glutathione complex formation are required for biliary excretion of arsenic. *Journal of Biological Chemistry* 2000, **275**(43):33404-33408.**
19. Bernstam L, Nriagu J: **Molecular aspects of arsenic stress. *Journal of Toxicology and Environmental Health - Part B: Critical Reviews* 2000, **3**(4):293-322.**
20. Pompella A, Visvikis A, Paolicchi A, De Tata V, Casini AF: **The changing faces of glutathione, a cellular protagonist. *Biochemical Pharmacology* 2003, **66**(8):1499-1503.**
21. Kvint K, Nachin L, Diez A, Nyström T: **The bacterial universal stress protein: Function and regulation. *Current Opinion in Microbiology* 2003, **6**(2):140-145.**
22. Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J: **Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? *Applied and Environmental Microbiology* 2001, **67**(7):2873-2882.**
23. Chevance FFV, Hughes KT: **Coordinating assembly of a bacterial macromolecular machine. *Nature Reviews Microbiology* 2008, **6**(6):455-465.**
24. Ellis PJ, Conrads T, Hille R, Kuhn P: **Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 Å and 2.03 Å. *Structure* 2001, **9**(2):125-132.**
25. Yorimitsu T, Homma M: **Na⁺-driven flagellar motor of *Vibrio*. *Biochimica et Biophysica Acta - Bioenergetics* 2001, **1505**(1):82-93.**
26. Muller D, Simeonova DD, Riegel P, Mangenot S, Koechler S, Lièvremon D, Bertin PN, Lett MC: ***Herminiimonas arsenicoxydans* sp. nov., a metalloresistant bacterium. *International Journal of Systematic and Evolutionary Microbiology* 2006, **56**(8):1765-1769.**
27. Muller D, Lièvremon D, Simeonova DD, Hubert JC, Lett MC: **Arsenite oxidase *aox* genes from a metal-resistant beta-proteobacterium. *Journal of Bacteriology* 2003, **185**(1):135-141.**

Figure legends

Figure 1 : Temporal dynamics of arsenic oxidoreduction in *H. arsenicoxydans*. (A) Time course measurement of As(III) and As(V) species after As(III) exposure.

Figure 2 : Patterns of gene expression changes during As(III) exposure. Expression graphs of the 20 clusters obtained by plotting the log₂ fold-change values (y-axis) at each of the three time points (x-axis). The average for all genes in each cluster is highlighted by a blue line. Class I and Class II contained genes whose expression levels gradually decreased or increased after the induction, respectively. Class II contained genes specific of the intermediary phase.

Figure 3 : Distribution of genes differentially expressed under arsenic stress and specific to Class I or II associated in clusters of functional categories. (A) induced and (B) repressed.

Figure 4 : Immunodetection of AoxB protein in total protein extracts of *H. arsenicoxydans* induced or not for 15 min, 6 or 8 hours with As(III). AoxB was detected as a single band at 90 kDa.

Figure 5 : Effect of temporal exposure to arsenite on *H. arsenicoxydans* motility. The level of motility at 24, 48, 72, 96 and 120 hours was evaluated as the diameter of the swarming ring in mm. The results presented are the mean value of five independent experiments.

Figure 6 : **A.** Conceptual representation of the metabolic response after temporal exposure to As(III) in *H. arsenicoxydans*. (A) after 15 min induction (early phase) and (B) after 8

hours induction (late phase). Genes repressed or induced after As(III) exposure are shown with a red or green background, respectively

Figure 1

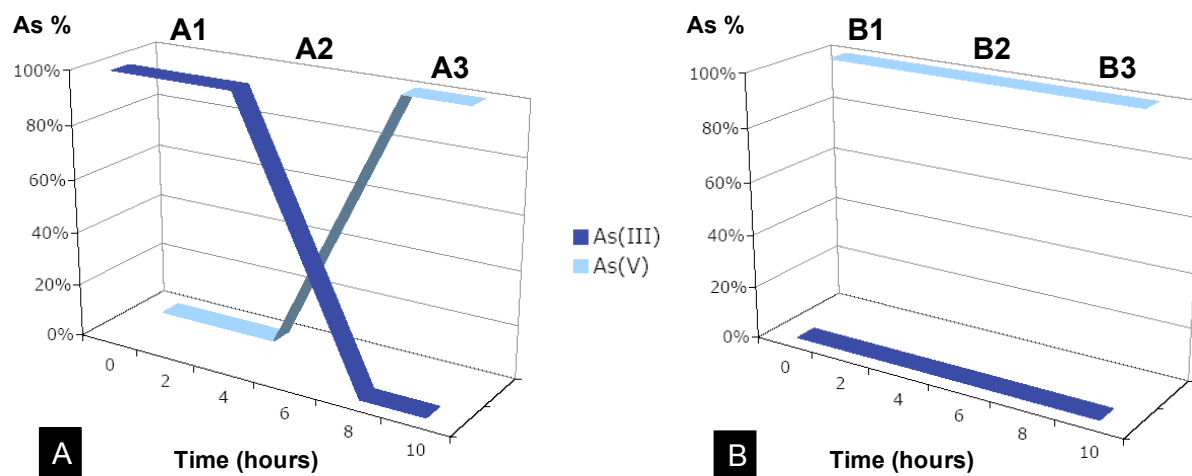


Figure 2

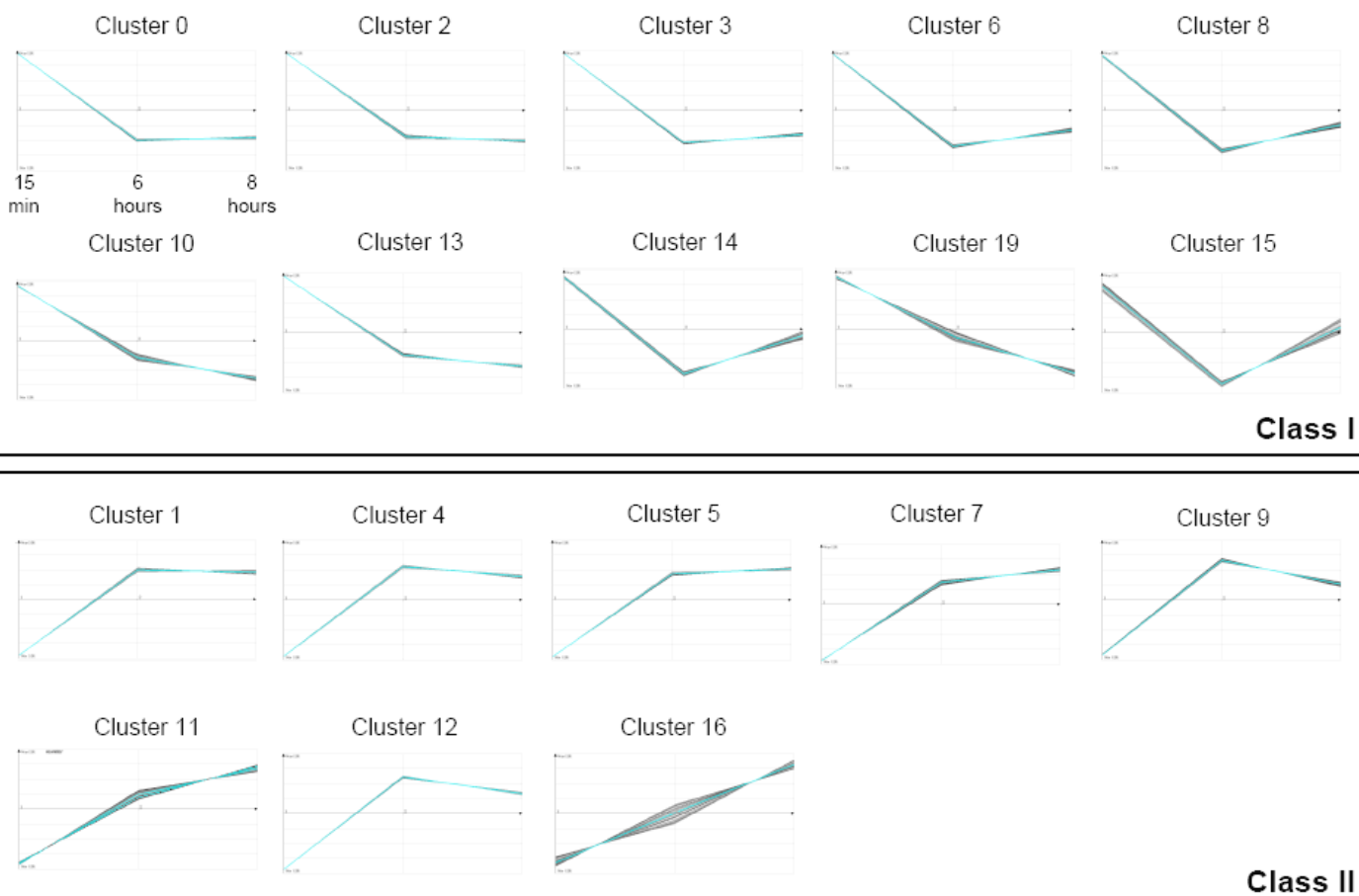


Figure 3

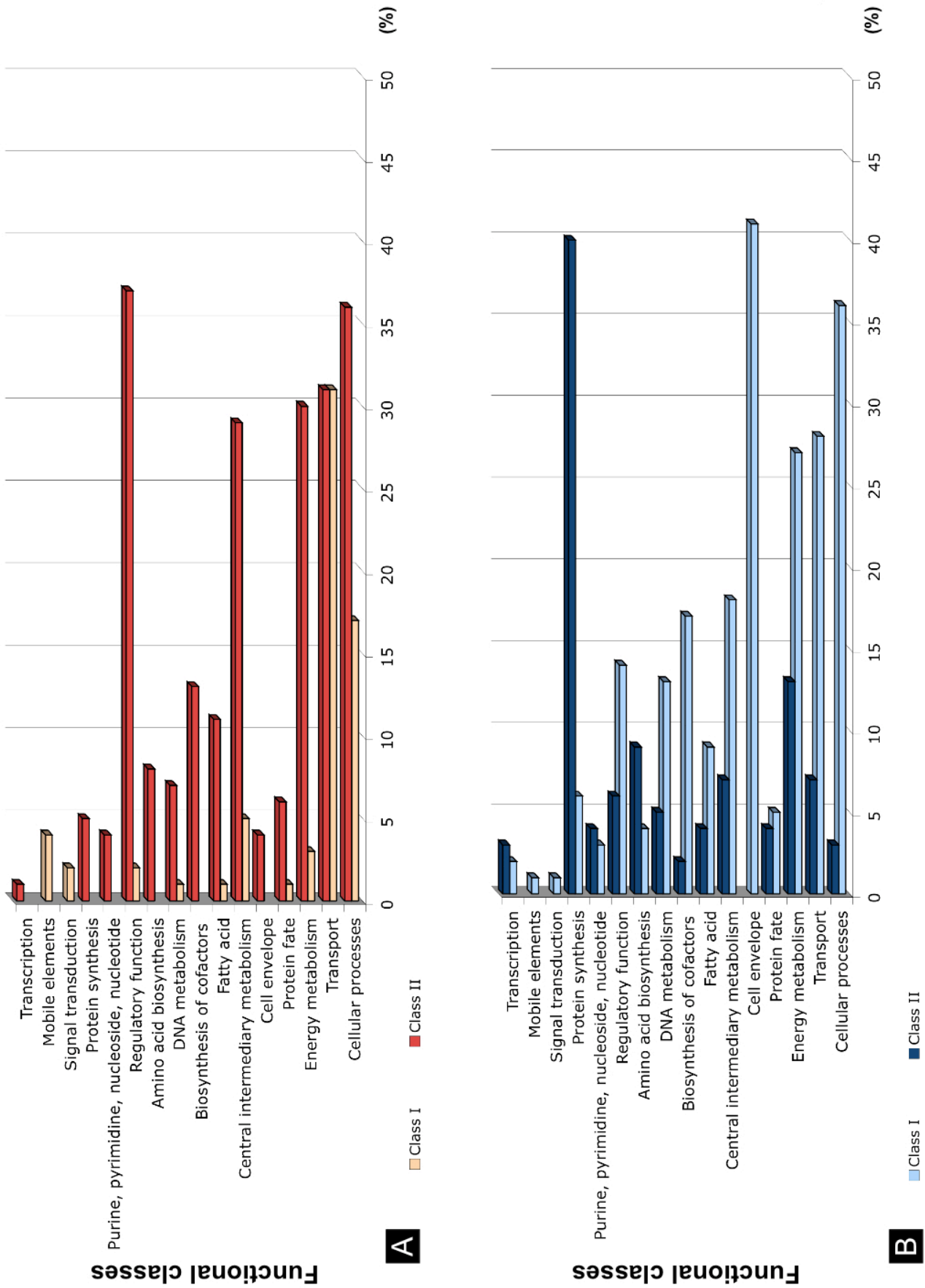


Figure 4

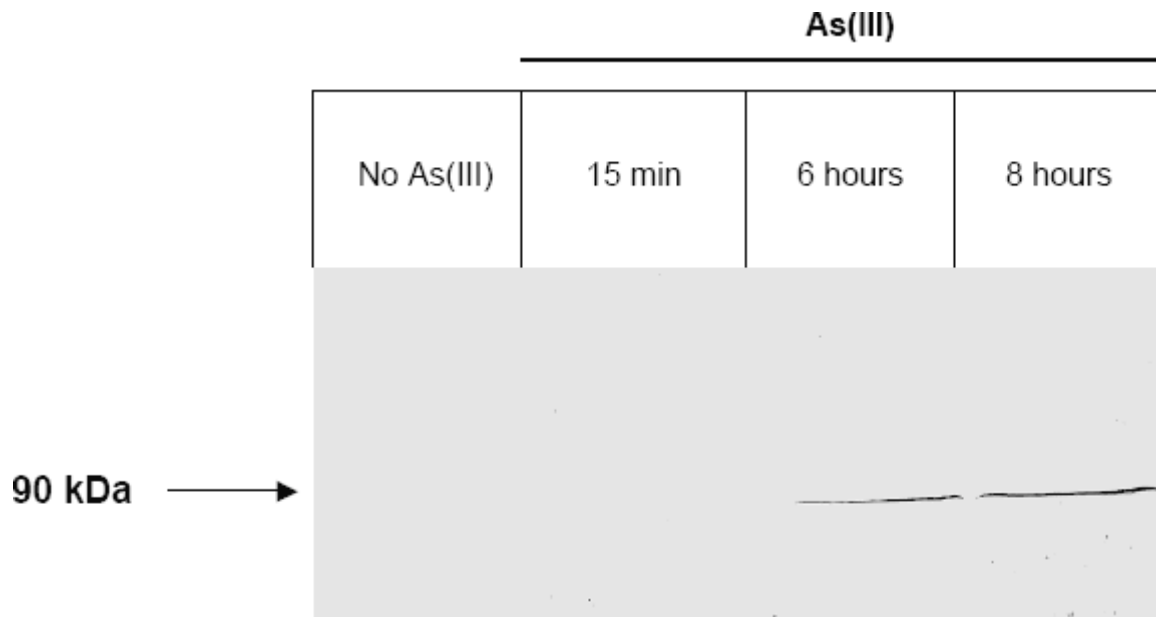


Figure 5

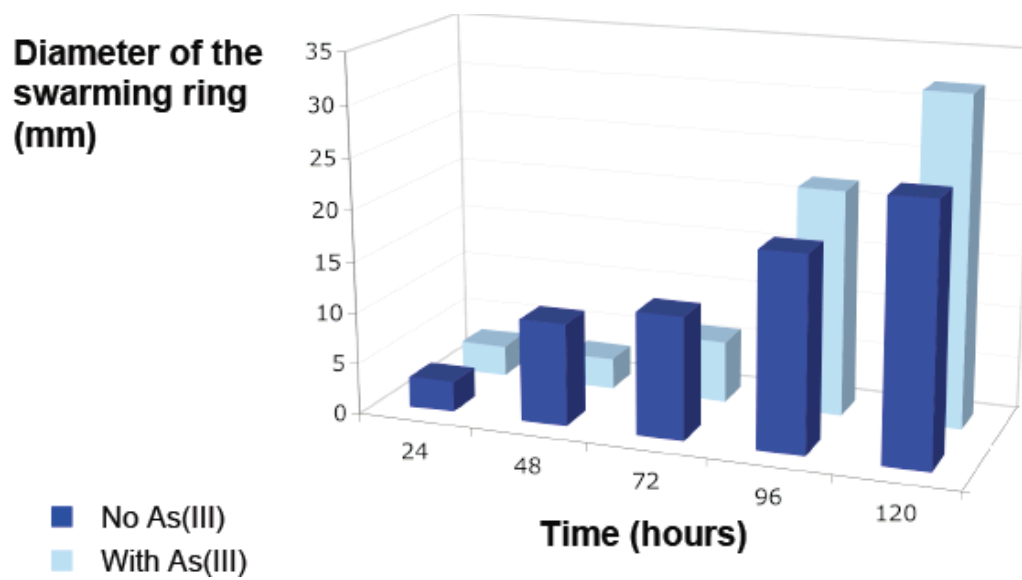
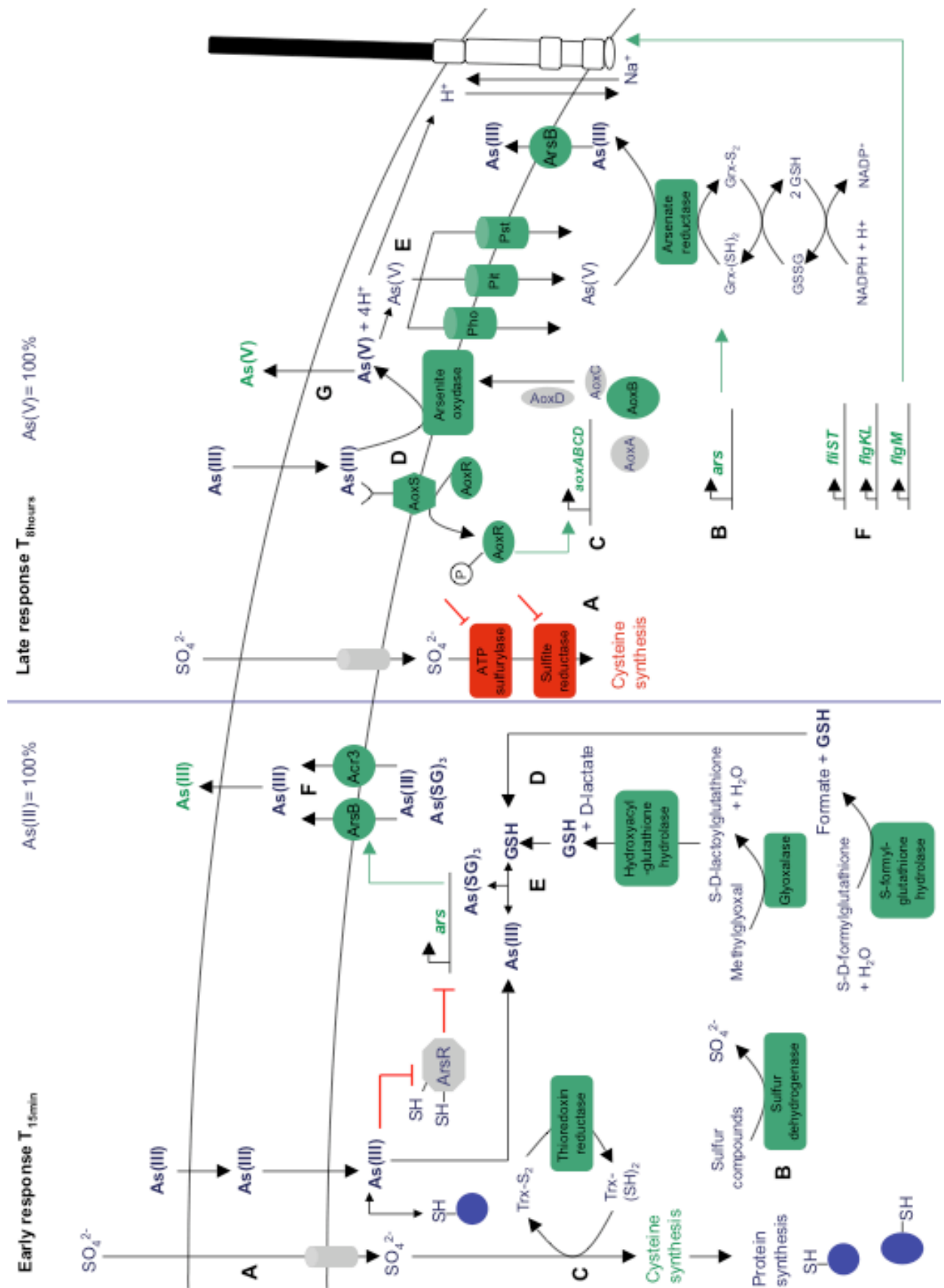


Figure 6



C/ Discussion

Dans ce chapitre, nous avons pu montrer pour la première fois qu'il existe une cinétique de la réponse au stress arsénié chez les micro-organismes. Cette étude a été significativement facilitée par des techniques telles que les puces à ADN ou la RT-PCR quantitative. En effet, en comparant les deux conditions (- et + As(III) à différents temps), ces méthodes nous ont permis d'identifier les gènes qui sont exprimés différemment à un temps donné. Les puces à ADN se sont révélées être un outil puissant pour mesurer et visualiser très rapidement les différences d'expression entre les gènes et ceci à l'échelle d'un génome complet. La RT-PCR quantitative quant à elle a été un excellent compromis pour quantifier l'expression d'un nombre limité de gènes. Ainsi, chez *Herminiimonas arsenicoxydans*, une réponse précoce (15 min d'induction) a été identifiée et est caractérisée par une induction des gènes codant des protéines chaperons ou impliqués dans le métabolisme de l'arsenic (*ars* uniquement), le stress oxydatif, le métabolisme du soufre. Aussi, elle a été qualifiée de réponse générale. Une réponse tardive a également été identifiée. Celle-ci est caractérisée par une induction des gènes impliqués dans le métabolisme de l'arsenic en particulier l'arsénite oxydase, la motilité, le transport du phosphate. Elle a été qualifiée de réponse spécifique.

Outre les réponses biologiques qu'ils apportent, les résultats présentés dans ce chapitre confirment l'utilité d'étudier la réponse à un stress à différents temps. En effet, ce chapitre reprend deux caractéristiques introduites dans la première partie :

- De nombreuses réponses au stress se traduisent par une régulation de l'expression des gènes. Dans notre cas, en moyenne 10% des gènes sont exprimés différemment en réponse au stress arsénié. Toutefois, la distribution des transcrits exprimés différemment, n'est pas la même aux différents temps étudiés.
- Il existe : (i) des réponses générales au stress, c'est-à-dire qu'elles sont indépendantes de la nature du stress et (ii) des réponses spécifiques. A notre connaissance, il s'agit de la première étude qui montre qu'il existe une réponse générale et spécifique en réponse à l'arsenic et que celles-ci sont organisées dans le temps.

Caractérisation des souches de
Thiomonas sp. par génomique descriptive

A/ Introduction

Dans les chapitres précédents, nous avons caractérisé les mécanismes moléculaires de régulation de l'arsénite oxydase et la réponse au stress arsénié chez une bactérie hétérotrophe, *H. arsenicoxydans*, mais les processus d'oxydation de l'arsénite concernent d'autres micro-organismes, en particulier les bactéries chimiolithoautotrophes chez lesquelles ils restent largement inexplorés.

L'étude présentée dans ce chapitre, a été motivée par la découverte de plusieurs souches de *Thiomonas* isolées de différents sites contaminés par l'arsenic. Plusieurs souches ont été isolées par exemple des sites miniers de Carnoulès dans la Gard (Duquesne *et al.*, 2003 ; Bruneel *et al.*, 2003) (Battaglia-Brunet *et al.*, 2006) et de Chéni dans la Limousin (Bryan *et al.*, 2009). La souche *Thiomonas* 3As est une β -protéobactérie chimiolithoautotrophe facultative, qui a été isolée des eaux de drainages miniers acides (DMA) de Carnoulès. Cette bactérie à Gram négatif se présente sous forme d'un bâtonnet doté d'un flagelle polaire. Elle est résistante à de nombreux métaux et métalloïdes, notamment l'arsenic. *Thiomonas* 3As et plusieurs souches isolées de DMA sont capables d'oxyder l'As(III) en As(V). Par contre, l'une des bactéries isolées du même site que *Thiomonas* 3As, ne possède pas cette capacité à oxyder l'As(III). Ces bactéries seraient donc distribuées de façon ubiquitaire dans les DMA, mais elles s'adaptent différemment.

L'objectif de cette étude a été d'identifier les éléments génétiques qui confèrent à ces différentes souches la capacité de croître et de coloniser efficacement un environnement toxique. Afin d'analyser ces génomes, une approche par génomique comparative a été menée sur quatre souches de la même espèce (CB1, CB2, CB3 et CB6) et quatre espèces différentes (3As, *T. arsenivorans*, *T. perometabolis* et *Thiomonas* sp. Ynys1).

Ma participation à ce projet a porté sur la détermination des concentrations minimales inhibitrices de chaque souche vis-à-vis de différents métaux ou métalloïdes, notamment l'As(III) et l'As(V). Par ailleurs, une annotation experte du génome de la souche *Thiomonas* 3As et une exploration des gènes impliqués dans le métabolisme de l'arsenic a été réalisée.

L'ensemble de ces travaux a fait l'objet d'une publication (Arsène-Ploetze *et al.*, 2010), présentée ci-dessous.

B/ Manuscrit

Structure, Function, and Evolution of the *Thiomonas* spp. Genome

Florence Arsène-Ploetze¹, Sandrine Koechler¹, Marie Marchal¹, Jean-Yves Coppée², Michael Chandler³, Violaine Bonnefoy⁴, Céline Brochier-Armanet⁴, Mohamed Barakat⁵, Valérie Barbe⁶, Fabienne Battaglia-Brunet⁷, Odile Bruneel⁸, Christopher G. Bryan^{1,9a}, Jessica Cleiss-Arnold¹, Stéphane Cruveiller^{6,9}, Mathieu Erhardt¹⁰, Audrey Heinrich-Salmeron¹, Florence Hommais¹¹, Catherine Joulian⁷, Evelyne Krin¹², Aurélie Lieutaud⁴, Didier Lièvre¹, Caroline Michel⁷, Daniel Muller^{1,9b}, Philippe Ortet⁵, Caroline Proux², Patricia Siguier³, David Roche^{6,9}, Zoé Rouy⁶, Grégory Salvignol⁹, Djamilia Slyemi⁴, Emmanuel Talla⁴, Stéphanie Weiss¹, Jean Weissenbach^{6,9}, Claudine Médigue^{6,9}, Philippe N. Bertin^{1*}

1 Génétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS and Université de Strasbourg, Strasbourg, France, **2** Genopole, Plate-forme puces à ADN, Institut Pasteur, Paris, France, **3** Laboratoire de Microbiologie et Génétique Moléculaire, UMR5100, Toulouse, France, **4** Laboratoire de Chimie Bactérienne, UPR9043 CNRS, Institut de Microbiologie de la Méditerranée, Marseille, France, **5** Institut de Biologie Environnementale et de Biotechnologie, CEA-CNRS-Université Aix-Marseille II, Saint-Paul-lez-Durance, France, **6** Institut de Génomique, CEA-DSV, Génoscope, Evry, France, **7** Environnement et Procédés, Ecotechnologie, BRGM, Orléans, France, **8** Laboratoire Hydrosociétés Montpellier, UMR 5569 CNRS, IRD and Universités Montpellier I and II, Montpellier, France, **9** Génomique Métabolique, Laboratoire de Génomique Comparative, CNRS UMR8030, Evry, France, **10** Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, Strasbourg, France, **11** Unité Microbiologie, Adaptation, Pathogénie, CNRS-INSA-UCB UMR 5240, Université Lyon 1, Villeurbanne, France, **12** Génétique des Génomes Bactériens, URA2171, Institut Pasteur, Paris, France

Abstract

Bacteria of the *Thiomonas* genus are ubiquitous in extreme environments, such as arsenic-rich acid mine drainage (AMD). The genome of one of these strains, *Thiomonas* sp. 3As, was sequenced, annotated, and examined, revealing specific adaptations allowing this bacterium to survive and grow in its highly toxic environment. In order to explore genomic diversity as well as genetic evolution in *Thiomonas* spp., a comparative genomic hybridization (CGH) approach was used on eight different strains of the *Thiomonas* genus, including five strains of the same species. Our results suggest that the *Thiomonas* genome has evolved through the gain or loss of genomic islands and that this evolution is influenced by the specific environmental conditions in which the strains live.

Citation: Arsène-Ploetze F, Koechler S, Marchal M, Coppée J-Y, Chandler M, et al. (2010) Structure, Function, and Evolution of the *Thiomonas* spp. Genome. PLoS Genet 6(2): e1000859. doi:10.1371/journal.pgen.1000859

Editor: Nancy A. Moran, University of Arizona, United States of America

Received: July 28, 2009; **Accepted:** January 25, 2010; **Published:** February 26, 2010

Copyright: © 2010 Arsène-Ploetze et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by the ANR 07-BLAN-0118 project (Agence Nationale de la Recherche). Annotation and database management were supported by a grant from ANR 07-PFTV MicroScope project. CGB was supported by a grant from the University of Strasbourg; MM by a grant from ANR COBIAS project (PRECODD 2007, Agence Nationale de la Recherche); JC, AHS, SW, and DS by a grant from the Ministère de la Recherche et de la Technologie; and AL by a grant from ADEME-BRGM. This work was performed within the framework of the French research network "Arsenic metabolism in microorganisms" (GDR2909-CNRS)(http://gdr2909.u-strasbg.fr). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: philippe.bertin@unistra.fr

^{9a} Current address: Centre for Bioprocess Engineering Research (CeBER), Dept. Chemical Engineering, University of Cape Town, Rondebosch, South Africa

^{9b} Current address: Université Lyon 1, CNRS, UMR5557, Ecologie Microbienne, Villeurbanne, France

Introduction

In environments such as those impacted by acid mine drainage (AMD), high toxic element concentrations, low levels of organic matter and low pH make growth conditions extreme. AMD is generally characterized by elevated sulfate, iron and other metal concentrations, in particular, inorganic forms of arsenic such as arsenite (As(III)) and arsenate (As(V)) [1,2]. While these waters are toxic to the majority of prokaryotic and eukaryotic organisms, a few Bacteria and Archaea are not only resistant to but also able to metabolize some of the toxic compounds present [1]. Members of the *Thiomonas* genus are frequently found in AMD and AMD-impacted environments, as *Thiomonas* sp. 3As and "*Thiomonas arsenivorans*" [2–6]. These *Betaproteobacteria* have been defined as

facultative chemolithoautotrophs, which grow optimally in mixotrophic media containing reduced inorganic sulfur compounds (RISCs) and organic supplements. Some strains are capable of autotrophic growth and others are capable of organotrophic growth in the absence of any inorganic energy source [5,7,8]. Recently described species and isolates include "*Tm. arsenivorans*" [9], *Tm. delicata* [10], *Thiomonas* sp. 3As [5] and Ynys1 [11]. *Thiomonas* sp. 3As as well as other recently isolated strains from AMD draining the Carnoulès mine site (southeastern France) containing a high arsenic concentration (up to 350 mg L⁻¹) [3,12], present interesting physiological and metabolic capacities, in particular an ability to oxidize As(III).

Over the past few years an increasing number of genomes has been sequenced, revealing that bacterial species harbor a core

Author Summary

Recent advances in the field of arsenic microbial metabolism have revealed that bacteria colonize a large panel of highly contaminated environments. Belonging to the order of Burkholderiales, *Thiomonas* strains are ubiquitous in arsenic-contaminated environments. The genome of one of them, i.e. *Thiomonas* sp. 3As, was deciphered and compared to the genome of several other *Thiomonas* strains. We found that their flexible gene pool evolved to allow both the surviving and growth in their peculiar environment. In particular, the acquisition by strains of the same species of different genomic islands conferred heavy metal resistance and metabolic idiosyncrasies. Our comparative genomic analyses suggest that the natural environment influences the genomic evolution of these bacteria. Importantly, these results highlight the genomic variability that may exist inside a taxonomic group, enlarging the concept of bacterial species.

genome containing essential genes and a dispensable genome carrying accessory genes [13]. Some of these accessory genes are found within genomic islands (GEIs) [14] and have been acquired by horizontal gene transfer (HGT). These GEIs are discrete DNA segments (from 10 to 200 kbp) characterized by nucleotide statistics (G+C content or codon usage) that differ from the rest of the genome, and are often inserted in tRNA or tRNA-like genes. Their boundaries are frequently determined by 16–20 bp (up to 130 bp) perfect or almost perfect direct repeats (DRs). These regions often harbor functional or cryptic genes encoding integrases or factors involved in plasmid conjugation or related to phages. GEIs encompass other categories of elements such as integrative and conjugative elements (ICE), conjugative transposons and cryptic or defective prophages. Such GEIs are self-mobile and play an important role in genome plasticity [14]. In almost all cases, GEIs have been detected *in silico*, by the comparison of closely related strains. Nevertheless, the role of GEIs in genome plasticity has also been experimentally demonstrated in several pathogenic bacteria such as *Staphylococcus aureus* or *Yersinia pseudotuberculosis* [15,16] or in *Pseudomonas* sp. strain B13 isolated from a sewage treatment plant [17].

Deciphering dispensable genomes has revealed that the loss or gain of genomic islands may be important for bacterial evolution [18]. Indeed, these analyses allow the determination of the genome sequence, called pan-genome or supragenome, not just of individual bacteria, but also of entire species, genera or even bacterial kingdom [19,20]. These data result in debates on taxonomic methods used to define the bacterial species [21,22], e.g. pathogens such as *Streptococcus agalactiae* [21,23] or environmental bacteria such as *Prochlorococcus* [24,25] or *Agrobacterium* [26]. However, beyond these well-known and cultivable microorganisms, the diversity of bacteria, in particular those found in extreme environments, has hitherto been comparatively poorly studied. Genome analysis of such extremophiles may yet reveal interesting capacities since these bacteria may express unexpected and unusual enzymes [27]. Since the role of GEIs in extremophiles has not been yet well explored, little is known about their evolution.

In the present study, the genome of *Thiomonas* sp. 3As was sequenced and analyzed. It was next compared to the genome of other *Thiomonas* strains, either of the same species or of other species of the same genus. This genome exploration revealed that *Thiomonas* sp. 3As evolved to survive and grow in its particular extreme environment, probably through the acquisition of GEIs.

Results

General Features of the *Thiomonas* sp. 3As Genome

The genome of *Thiomonas* sp. 3As comprises a 3.7 Mbp circular chromosome and a 46.8 kbp plasmid (Table 1). The single circular chromosome contains 3,632 coding sequences (CDSs) (Table 1, Figure S1A). The mean G+C content of the *Thiomonas* sp. 3As genome is 63.8%. However, the distribution along the genome revealed several regions with a G+C content clearly divergent from this mean value (Figure S1). This suggests that several genomic regions are of exogenous origin. Indeed, 196 genes having mobile and extrachromosomal element functions were identified in the genome, among which a total of 91 ISs (Figure S1A, Table S1) representing 2.5% of total CDS. None of these ISs were found as part of composite transposons, while several were identified as neighbors of phage-like site-specific recombinases.

The plasmid, pTHI, contains 68 predicted CDS. 21 genes were found in synteny with genes carried by the *R. eutropha* JMP134 plasmid pJP4, and among them, *par/tjf/pem* genes necessary for plasmid partitioning, stability and replication (Figure S1B). These observations suggest that pTHI, as JMP134, belongs to the IncP-1 β group [28]. pTHI contains 13 of the 14 genes involved in conjugation (*vir* and *tra* genes) and genes that could fulfill the function of the missing components were found on the chromosome. Therefore, *Thiomonas* sp. 3As may be able to express a complete Type IV secretory system (T4SS) of the Vir/Tra type required for pTHI conjugal transfer. IncP-1 β members are known to carry multi-resistance determinants and degradative cassettes [28], and plasmid pTHI indeed carries a Tn3-related transposon. This transposon contains part of a mercury resistance operon found in many other Gram negative bacterial transposon such as Tn21, Tn501 and Tn5053 [29].

Carbon and Energy Metabolism

Thiomonas sp. 3As is able to use organic compounds as a carbon source or electron donor [5,8]. However, under certain conditions this bacterium may also be able to grow autotrophically [5]. A complete set of *cbb/rbc/cso* genes involved in carbon fixation *via* the Calvin cycle, and genes involved in glycogen, starch and polyhydroxybutyrate (PHB) biosynthesis pathways were identified (Figure 1 and [5]). Fructose, glucose, several amino acid, C4-dicarboxylates, propionate, acetate, lactate, formate, ethanol and glycerol are potential carbon sources or electron donors, since genes involved in their import or degradation *via* the glycolysis, the Entner-Doudoroff, the TriCarboxylic Acid (TCA) or the “rubisco shunt” pathways are present in the genome. The presence of all genes involved in the oxidative phosphorylation pathway (Figure 1) suggests that *Thiomonas* sp. 3As has a respiratory metabolism. Moreover, since several genes coding for terminal oxidases (i.e. *cbh3*, *bd* or *aa3*) were found, this respiratory metabolism may occur over a wide range of oxygen tensions. Finally, the presence of a nitrate reductase and of several formate dehydrogenases suggests that *Thiomonas* sp. 3As is able to use nitrate anaerobically as an electron acceptor and formate as electron donor. In the absence of carbohydrates, *Thiomonas* sp. 3As is a chemolithotroph and may use reduced inorganic sulfur compounds (RISCs) as an electron donor [5]. The presence of *soxRCDYZAXB*, *dsr*, *sorAB*, *sqr* and *fccAB* genes revealed that *Thiomonas* can oxidize thiosulfate, sulfite, S⁰ or H₂S to sulfate (Figure 1) [30,31].

Adaptive Capacities of *Thiomonas* sp. 3As to Its Extreme Environment

Thiomonas sp. is a moderate acidophile. Its optimum pH is 5 but this bacterium can withstand to pH as low as 2.9 (Slyemi, Johnson

Table 1. General genome and plasmid features.

Molecule	Category	Feature	Value	
Plasmid	General characteristics	Size (bp)	46,756	
		GC content (%)	60.49	
		Coding density (%)	88.19	
		Predicted CDSs	68	
	Proteins with predicted function	Percent of total CDSs	48.52	
		Secretion (%)	11.76	
		Partitioning	2.94	
		Replication, recombination	8.82	
		Inorganic ion transport and metabolism (Hg)	2.94	
	Proteins without predicted function	Conserved hypothetical proteins (%)	13.24	
		No homology to any previously reported sequences (%)	38.24	
		Percent of total CDSs	51.48	
	Genome	General characteristics	Size (bp)	3,738,778
			GC content (%)	63.8
Coding density (%)			90.01	
16S-23S-5S rRNA operons			1	
tRNAs			43	
Proteins with predicted function		Percent of total CDSs	74.2	
		Heavy metal resistance (%)	1.6	
		Related to arsenic metabolism/transport (%)	0.6	
Proteins without predicted function		Conserved hypothetical proteins (%)	13.54	
		Hypothetical proteins (%)	11.95	
		Percent of total CDSs	25.8	
Mobile and extrachromosomal element functions		Percent of total CDSs	5.4	
		Transposases (nb CDSs)	101	
		Phage related (nb CDSs)	61	
Repeated regions (%)		7.62		

doi:10.1371/journal.pgen.1000859.t001

and Bonnefoy, personal communication). *Thiomonas* sp. 3As pH homeostasis mechanisms may therefore be strictly controlled as previously described [32,33]. First, genes encoding a potassium-transporting P type ATPase (*kdpABC*) are present in the *Thiomonas* sp. 3As genome. This ATPase could be involved in the generation of a positive internal potential produced by a greater influx of potassium ions than the outward flux of protons. Second, to strengthen the membrane, likely by lowering membrane proton conductance, *Thiomonas* sp. has cyclopropane fatty acids [5]. Accordingly, two putative *cfa* genes encoding cyclopropane fatty acid synthase have been detected. Third, cytoplasmic buffering can be mediated either by amino acid decarboxylation and/or by polyphosphate granules. Genes encoding decarboxylases for lysine (4 CDS), phosphatidyl serine and glycine are present on *Thiomonas* sp. 3As genome. Moreover, urea (formed from arginine by arginase) may be degraded by urease (*ure* genes) or urea carboxylase and allophanate hydrolase. Urease encoding genes are known to be involved in acid tolerance in *Helicobacter pylori* [34]. Protons may be captured during polyphosphate synthesis. Polyphosphate granules have indeed been observed in electron micrographs of thin sections of *Thiomonas* sp. 3As [5]. Genes involved in such mechanisms (*ppk*, *pap*, *ppx*) were found in 3As genome. Fourth, primary and secondary proton efflux transporters

were predicted by genome sequence analysis, including four putative Na⁺/H⁺ exchangers and voltage gated channels for chloride involved in the extreme acid resistance response in *E. coli* (*clcAB*) [35]. Finally, the elimination of organic acids can lead to pH homeostasis. Some organic acid degradation pathways have been detected in *Thiomonas* sp. such as an acetyl-CoA synthetase-like. Moreover, formate oxidation was observed (Slyemi, Johnson and Bonnefoy, personal communication) and two formate dehydrogenases are encoded by the *Thiomonas* sp. 3As genome, these enzymes could convert acetate to acetyl-CoA and formate to CO₂ and hydrogen, respectively.

The Carnoulès AMD contains a high concentration of heavy metals such as zinc or lead. To resist to heavy metals, bacteria usually develop several resistance mechanism including toxic compounds extrusion pumps [36] or biofilm synthesis [37]. Flagella are important for the first steps of biofilm formation and all genes involved in motility, twitching and chemotaxis, were found in its genome. *Thiomonas* sp. 3As is motile but, unlike *H. arsenicoxydans*, this motility was not affected by arsenic concentration (Table S2, [38]). Finally, *Thiomonas* sp. 3As is able to synthesize exopolysaccharides (Table S2), and one *eps* operon involved in their synthesis was identified in the genome, as well as two *mdoDG* clusters involved in glucan synthesis. Several genes

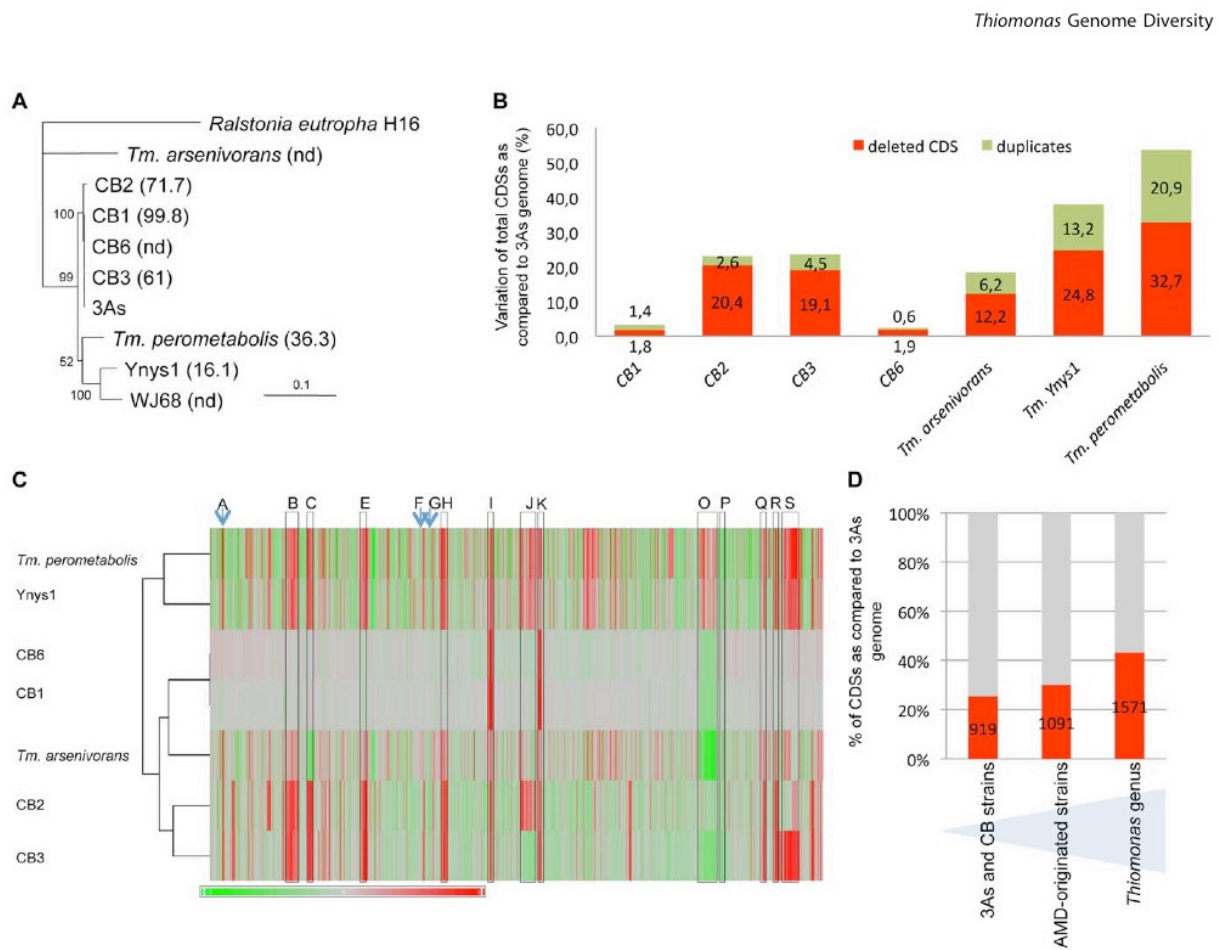


Figure 2. Genomic diversity among *Thiomonas* strains. (A) Phylogenetic dendrogram of the SuperGene construct of both the 16S rRNA and *rpoA* genes of the *Thiomonas* strains used in this study. *Ralstonia eutropha* H16 served as the outgroup. The DNA-DNA hybridization percentage is included in brackets. nd: not determined (CB1 and CB6 were almost identical at the genetic content level. For this reason, DNA-DNA hybridization analyses were performed only with CB1). Numbers at the branches indicate percentage bootstrap support from 500 re-samplings for ML analysis. NJ analyses produced the same branch positions (not shown). The scale bar represents changes per nucleotide; (B) number of variable CDS exhibited by each strain determined by CGH, expressed as a percent of the total number of 3As CDS; (C) hierarchical clustering established based on CGH experiments, represented as a composite view of genome diversity in *Thiomonas* strains compared to 3As. The column represents genes as they are found along 3As genome, starting from the origin of replication (THI0001). Results for each strain are shown in each row. Red color indicates absence or strong divergence leading to reduced hybridization efficiency as compared to the corresponding *Thiomonas* sp. 3As gene ($\text{Log}_2(A_{635 \text{ nm}}/A_{532 \text{ nm}}) \leq -1$); Green color indicates a $\text{Log}_2(A_{635 \text{ nm}}/A_{532 \text{ nm}}) \geq 1$, suggesting duplication of this gene. The regions corresponding to ThGEI-A, B, C, E, F, G, H, I, J, K, O, P, Q, R and S are indicated by a blue arrow or a grey rectangle; (D) percent of CDS found in core (in grey) or in dispensable (in red) genome, in 3As and CB strains (1st column), in AMD-originated strains (3As, CB strains and “*Tm. arsenivorans*”, 2nd column), or in all *Thiomonas* strains (3rd column).

doi:10.1371/journal.pgen.1000859.g002

highly divergent in at least one strain, accounting for 25.3% of strain 3As genes (Figure 2D). The remaining conserved CDS (2713 CDS, 74.7% of the genome of strain 3As) represent a common backbone of the “core” genes of this species.

In order to enlarge our comparative analysis, genomic similarities were similarly searched for in other *Thiomonas* species: an arsenite-oxidizing strain, “*Tm. arsenivorans*”, and two closely related strains that are unable to oxidize arsenite, *Tm. perometabolis* and *Thiomonas* sp. Ynys1 (Table S2, Figure 2, ArrayExpress database, accession number E-MEXP-2260). No significant hybridization was observed with oligomers corresponding to the plasmid, suggesting that pTHI is absent in all these strains. 18.4, 37.9 and 53.6% of the 3As CDS were flexible in “*Tm. arsenivorans*”, Ynys1 and *Tm. perometabolis*, respectively. Altogether, 1571 CDS accounting for at least 43.3% of the *Thiomonas* sp. 3As genome were found in the *Thiomonas* genus dispensable genome (Figure 3D).

Finally, these CGH experiments revealed that the *Thiomonas* core genome contains 2,061 CDS (56.7% of the *Thiomonas* genome). Interestingly, almost all genes involved in acid resistance described above, were found in this core genome, as for example genes involved in polyphosphate granule synthesis, *cfa* and *kdp* genes, genes encoding ion transporter amino acid decarboxylase, formate dehydrogenase and other hydrogenases. One *ars* operon involved in arsenic resistance, i.e. *arsI*, and almost all genes involved in DNA repair were also conserved in all strains.

Among the flexible pool, 19 regions (ThGEI-A - ThGEI-S) had similarities with GEIs found in other bacterial genomes, suggesting that they were possibly acquired by horizontal gene transfer: (i) an abnormal deviation of the codon adaptation index (CAI) and the GC content at the 3rd nucleotide position of each codon (GC₃) was observed in these regions as compared to the rest of the genome (Figure 3), (ii) many of their genes formed syntenic blocks that

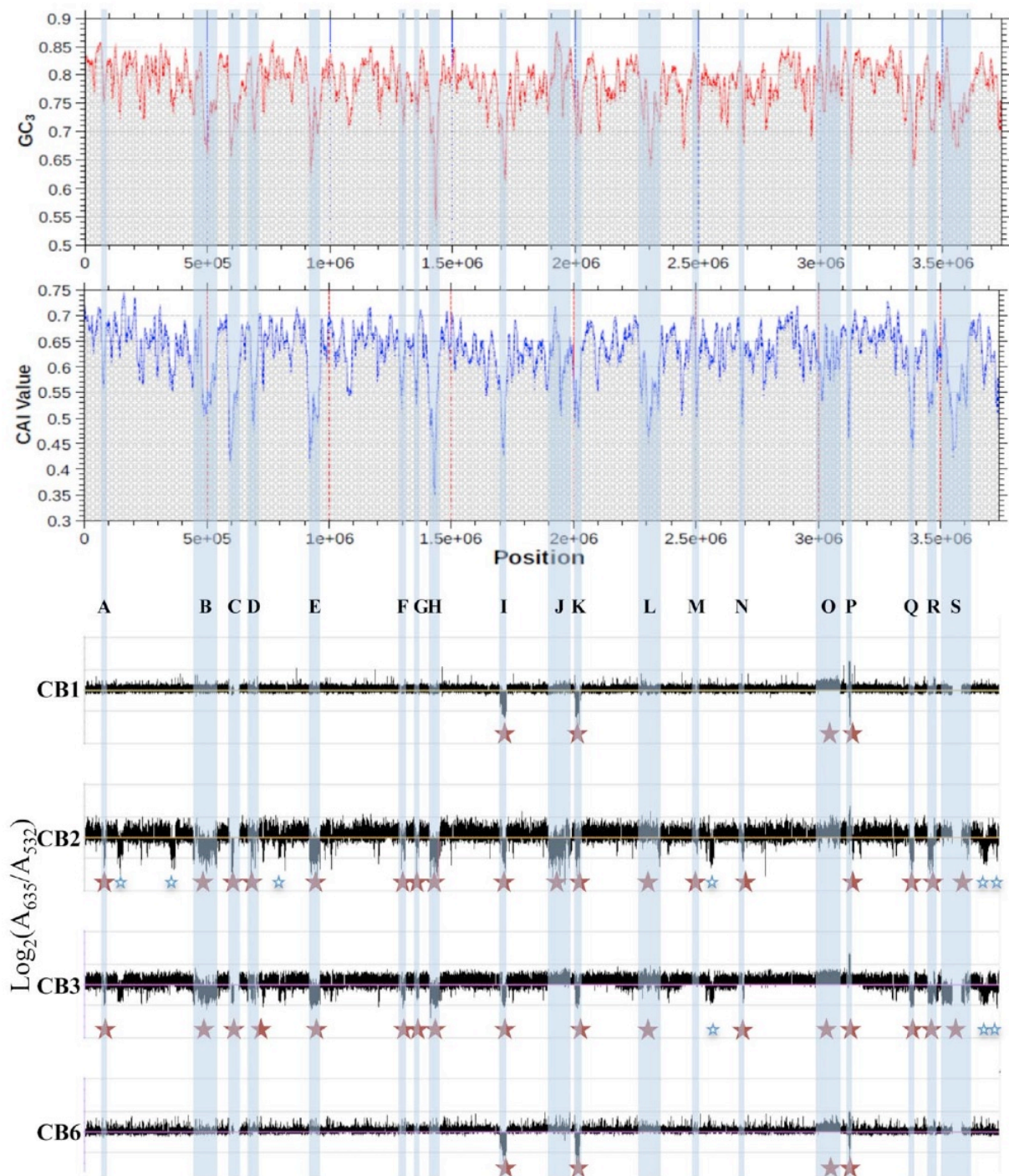


Figure 3. Deleted or duplicated regions in CB strains, obtained by CGH experiments. The codon adaptation index (CAI) and the GC content at the 3rd nucleotide position of each codon (GC_3) of *Thiomonas* sp. 3As is shown in the upper part. Y-axis displays the \log_2 -ratios (Cy5 (measured at 635 nm)/Cy3 (measured at 532 nm)). Red stars: variable regions found in CB strains and corresponding to GELs; Blue stars: other variable regions. doi:10.1371/journal.pgen.1000859.g003

differed from the general synteny observed in the rest of the genome (Table S3), (iii) genes with mobile and extrachromosomal element functions such as those coding for integrases were

localized within these regions, (iv) these regions were present at the 3'-end of tRNA or miscRNA genes, and/or (v) the borders of five deletions were verified in CB strains, by PCR and direct

repeats (10 to 112 bp-long) bordering these GEIs were found (Table S3).

Genetic Content of the 19 Genomic Islands Found in *Thiomonas* sp. 3As Genome

Genes found in the 19 *Thiomonas* sp. 3As GEIs and the synteny they share with genes in other bacteria are shown in Table S3. Interestingly, 70 (76.9%) of the 91 complete and partial ISs identified in the genome were located in genomic islands which represent only 21.5% of the genome (Figure S1A). In addition to the high numbers of ISs found in these GEIs, many hypothetical proteins as well as modification/restriction enzymes were encoded by these regions. In ten GEIs, accessory genes are involved in a particular metabolism such as acetoin, atrazin, benzoate, ethyl tetra-butyl ether (ETBE) hydroxyisobutyrate phenylacetic acid and urea degradation (ThGEI-E, ThGEI-C, ThGEI-S or ThGEI-R), or heavy metal resistance (ThGEI-J, ThGEI-L, ThGEI-O).

Interestingly, several genes found in distinct GEIs shared high amino acid identity (>70%, Figure S2). In addition, 47 genes found in the two regions ThGEI-C and ThGEI-S shared 100% identity. Because of this duplication, a 7 kbp region in ThGEI-S could not be sequenced and this gap may correspond to duplicated

genes of ThGEI-C. These observations suggest that genomic rearrangements occurred between several GEIs. Moreover, several islands seem to be composite, since some fragments of such islands are deleted or duplicated in *Thiomonas* strains. Such composite structure may originate from insertion or excision of DNA elements in these GEIs, which involve integrase or excisionase. This hypothesis is strengthened by the observation that 32 integrases were found in almost all GEIs except for ThGEI-B and ThGEI-R. Some of such integrases are similar to phage integrases. In addition, 2 excisionases are present in ThGEI-H and ThGEI-P and such genes were localized in the vicinity of tRNA, an additional phage-like character.

One GEI, ThGEI-J, contains a prophage region (55.6 kbp) and a cluster of 6 heavy metal resistance genes (39.4 kbp, i.e., *cad*, *cus*, *czc* and *sil* genes involved in resistance to Cd, Cu, Zn, Co and Ag) (Figure 4). The prophage region comprises 27 phage-related genes coding for structural and capsid or tail assembly proteins, replication, lysis and virulence factors. No conserved synteny with any previously described prophage could be observed. However, filamentous phage-like particles with icosahedral symmetry (capsid diameter of approximately 100 nm) and a various length tail (>600 nm), were observed by TEM from *Thiomonas* sp. 3As liquid

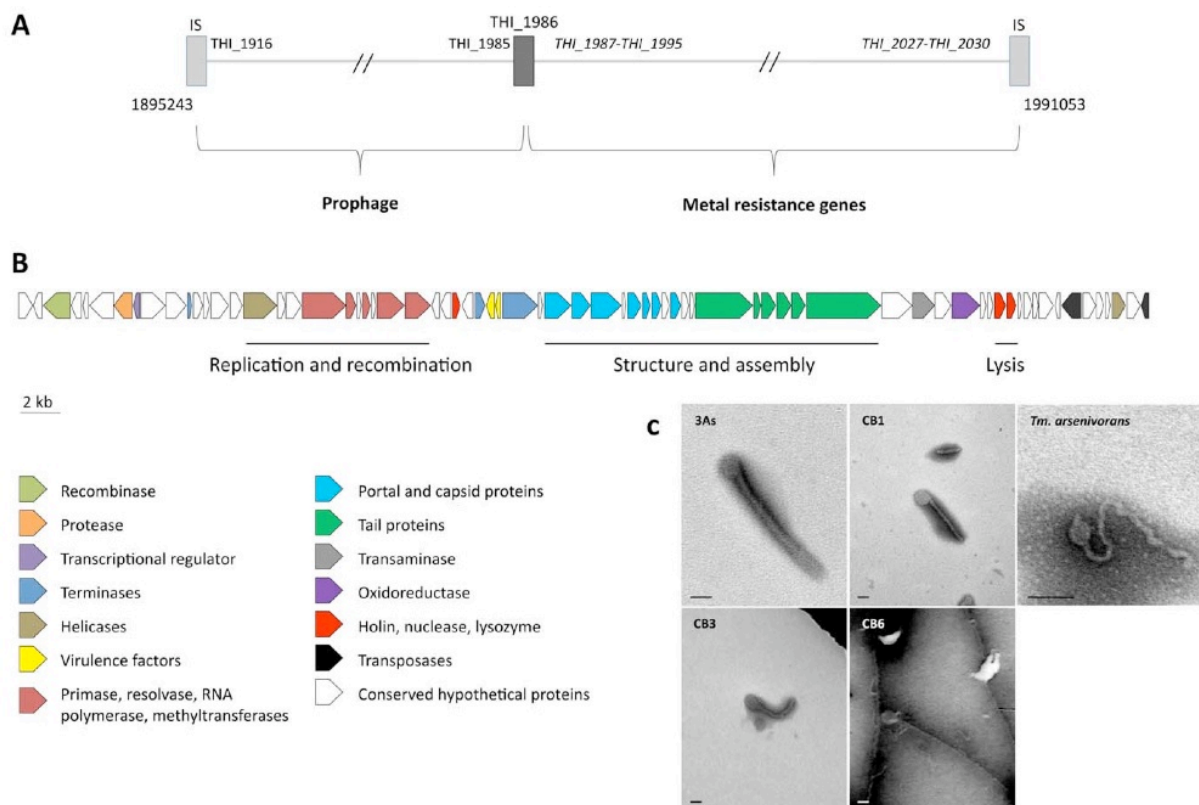


Figure 4. Prophage-like region and phage-like particle. (A) Schematic representation of the prophage-like region (55.6 kbp) and the contiguous metal resistance gene cluster (39.4 kbp). The total region is flanked by two ISs (light grey rectangles), and a partial transposase gene (THI_1986, dark grey rectangle) delimiting the two clusters. Genes in italics are duplicated in the *Thiomonas* sp. 3As genome; (B) organization of the prophage-like region. The 71 putative CDS, colored according to their category, are shown by arrows representing their direction. This cluster corresponds to the unique prophage-like region of the genome, containing notably structural, lysis and virulence associated phage genes; (C) transmission-electron micrographs of tailed bacteriophage-like particles from *Thiomonas* species. No bacteriophage-like particle was observed in *Thiomonas* sp. CB2 culture supernatant. Cultures were amended with mitomycin C (0.5 µg/mL). Bars: 100 nm. The 3As phage-like particle is hypothetically coded by the prophage-like region described above.
doi:10.1371/journal.pgen.1000859.g004

cultures exposed to the phage lytic phase inducer mitomycin C. Similar phage-like particles were observed in growth culture supernatants from CB1, CB3, CB6 and “*Tm. arsenivorans*” (Figure 4) but not from CB2, Ynys1 and *Tm. perometabolis* (data not shown), in agreement with CGH results showing that the ThGEI-J is absent in these strains (Figure 3, Table S3, ArrayExpress database, accession number E-MEXP-2260). These observations suggest that this prophage-like region may be functional in 3As, CB1, CB3, CB6 and “*Tm. arsenivorans*” under stress conditions, resulting in the formation of phage-like particles.

ThGEI-L and ThGEI-O Gene Content and Their Probable Acquisition by HGT

GEIs contribute to the adaptation of microorganisms to their ecological niches and participate in genome plasticity and evolution [14]. Therefore, the environmental conditions may influence the loss or conservation of GEIs. Such hypothesis was checked by searching for genome similarities between strains originated from similar environments, i.e. AMD. To this aim, a hierarchical clustering was established based on genomic comparisons (Figure 2C). Interestingly, the clustering obtained was different from that of the 16S rRNA/*rpoA*-based phylogenetic trees (Figure 2A). Indeed, all strains that originated from AMD heavily loaded with arsenic, i.e. “*Tm. arsenivorans*” and strains 3As, CB1, CB2, CB3 and CB6, grouped together, whereas Ynys1 and *Tm. perometabolis* formed a distinct group. Genes possibly dispensable for AMD survival were therefore searched for and we identified 2541 CDS conserved in all strains originated from AMD, and these CDS may constitute the “AMD” core genome of *Thiomonas*. Interestingly, several genes present in the ThGEI-L and ThGEI-O were conserved in AMD-originated strains but absent in the other strains (Figure 5).

The ThGEI-L carries genes involved in pantothenate and biotin synthesis, and may confer auxotrophy to the strains carrying this island. Moreover, genes encoding Co/Zn/Cd efflux pump were present in this GEI. In addition, this island is particularly rich in proteins with GGDEF and EAL domains. The GGDEF or EAL domain proteins are involved in either synthesis or hydrolysis of bis-(3'-5') cyclic dimeric GMP (c-di-GMP), an ubiquitous second messenger in the bacterial world that regulates cell-surface-associated traits and motility [43,44]. Because of the presence of such genes in the vicinity of 2 genes involved in chemotaxis, this island may be important for *Thiomonas* strains to form biofilm, a cellular process involved in resistance to toxic compounds [37]. Indeed, some of these genes are duplicated in CB2 and CB3 and these two strains were shown to develop better biofilm synthesis capacities (Table S2). This island also carries several genes encoding integrases and components of T4SS, such as *virB1*, *virB4*, *trbBCD*, *traCEFGI*, *mob* and a *pilE*-like gene. The presence of such genes suggests that this island originated from an integrative and conjugative element (ICE) that disseminates via conjugation [45]. These observations suggest that this island may be still mobile.

ThGEI-O contains the *aox* and *ars2* genes (Figure 5). In addition, several other genes were found such as *mer*, *cop*, *cus* and *cad* genes involved in mercury, copper and cadmium resistance, respectively, *cys* involved in sulfate assimilation, and *moe/moa* genes involved in molybdenum cofactor biosynthesis as well as genes involved in exopolysaccharide production. The synteny of the genes found in *Thiomonas* sp. 3As ThGEI-O is not conserved in other arsenic-oxidizing bacteria (Figure 5). Several genes present in this region are duplicated in CB1 and CB6 (i.e. the *cop* and *aox* genes), or in CB3 (i.e. *mer*, *cop*, *cus*, *dsb*, *cys*, *ars*, *moe/moa*, *aox*, and *ptxB* genes). Only a single copy of this region is present in CB2 and

3As. PCR amplification and sequencing revealed that this ThGEI-O island is located in a different genomic region in CB2 as compared to 3As. Moreover, the *aox* and *ars* genes found in the ThGEI-O are duplicated in “*Tm. arsenivorans*” but absent in Ynys1 and *Tm. perometabolis*. Indeed, these two strains were unable to oxidize As(III), their As(III) resistance was lower than that of the other strains, and gene PCR amplification of *aox* and *ars2* failed with DNA extracted from these strains (Table S2). Altogether, the presence of at least one copy of these genes in all six strains isolated from arsenic-rich environments (i.e. 3As, CB strains and “*Tm. arsenivorans*”) suggests that this GEI is of particular importance for the growth of *Thiomonas* strains in their toxic natural environment, AMD.

The evolutionary origin of the ThGEI-L and -O was investigated using two different approaches. First, we performed the phylogenetic analysis of the 196 genes contained in these two islands (Table S4, Table S5). The resulting trees revealed that these genes have very different evolutionary histories suggesting that the formation of ThGEI-L and -O islands occurred through the recruitment of genes from various origins by HGT (Table S4, columns 2–5). Interestingly, the closest homologue of 30/75 and 22/121 3As genes, in ThGEI-L and -O respectively, is found in other *Thiomonas* species (mainly *Tm. intermedia*), suggesting that the formation of these islands occurred prior to the diversification of the *Thiomonas* genus and is thus relatively ancient. This hypothesis is supported by the global correspondence analysis (COA) performed on the entire genome. Our results did not reveal any particular codon usage bias, strengthening the hypothesis that these ThGEIs are ancient in *Thiomonas* genus (Figure S3). This may explain why the major genes of these two islands are present in 3As, CB1, CB2, CB3, CB6 and “*Thiomonas arsenivorans*”, as for example, the *ars2* operon and *aox* genes of the ThGEI-O. The phylogenetic analysis of *aox* genes revealed that all *Thiomonas* sequences grouped together with relationships that are very similar to organism relationships inferred with *rpoA* (Figure S4A and S4B). This indicates that these genes were already present in the *Thiomonas* ancestor and vertically transmitted in this genus, but lost in Ynys1 and *Tm. perometabolis*. The phylogenetic analysis of the *arsB* genes, revealed that all *Thiomonas* sequences found in the ThGEI-O (i.e. *arsB2* from 3As, CB1, CB2, CB3, CB6 and “*Tm. arsenivorans*”), grouped together but not with *arsB1* genes that are part of the core genome of *Thiomonas*. Moreover, the evolutionary histories of these two proteins are different: *ArsB1* proteins belong in a group containing mainly Alpha-Proteobacteria, whereas *ArsB2* seems more closely related to Gamma-Proteobacteria (Figure S4C). These observations revealed that the *ars1* and *ars2* operons were not acquired from the same source or at the same time.

Discussion

The exploration of the *Thiomonas* sp. 3As genome suggests that this strain has a wide range of metabolic capacities at its disposal. Many of them may make this bacterium particularly well suited to survive in its extreme environment, the acidic and arsenic-rich waters draining the Carnoulès mine tailings, as for example biofilm formation and heavy metal resistance. Moreover, some metabolic capacities are unique as compared to another arsenic-resistant bacterium, whose genome has been recently sequenced and annotated, *H. arsenicoxydans*, a strict chemoorganotroph, isolated from activated sludge [38]. The first metabolic idiosyncrasy of *Thiomonas* sp. 3As is its particular carbon and energy metabolic capacities. Indeed, several organic or inorganic electron donors, such as reduced inorganic sulfur compounds [31], could

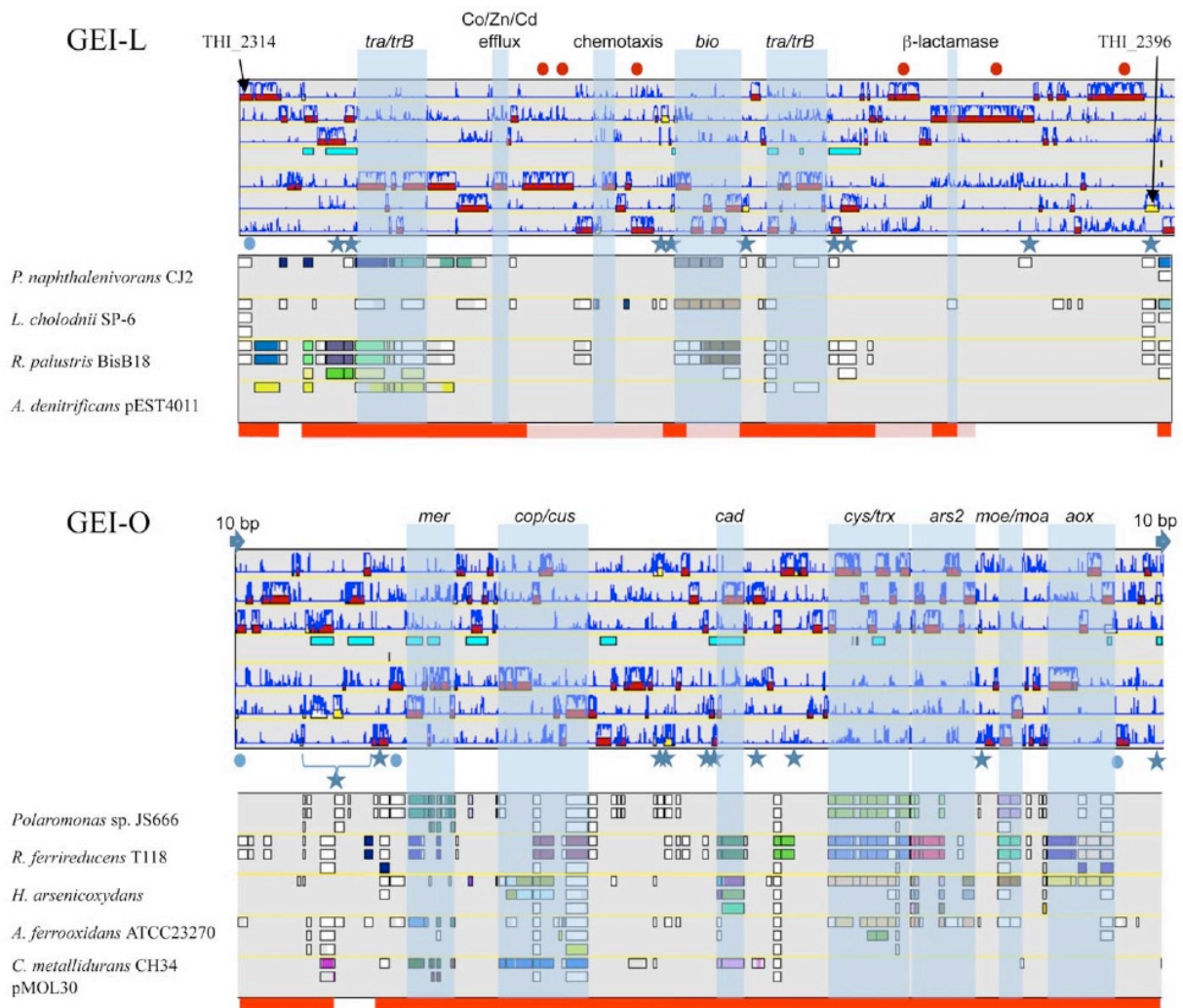


Figure 5. Genomic Island ThGEI-O containing genes involved in arsenic metabolism, and ThGEI-L. The best synteny results of the genes were obtained with genes of *Polaromonas naphthalenivorans* CJ2, *Leptothrix cholodnii* SP-6, *Rhodopseudomonas palustris* BisB18, *Achromobacter denitrificans* (ThGEI-L) or *Polaromonas* sp. JS666, *Rhodobacter ferrireducens*, *H. arsenicoxydans*, *Acidithiobacillus ferrooxidans* ATCC23270, and *Cupriavidus metallidurans* CH34 pMOL30 (ThGEI-O). Blue circles: genes originated from phage; Blue stars: transposases; Blue arrows: direct repeats. The genes conserved in 3As and CB strains or in AMD-originated strains (3As, CB strains and “*Tm. arsenivorans*”), are indicated by pink and red lines, respectively. Red dots indicate proteins with GGDEF and EAL domains. doi:10.1371/journal.pgen.1000859.g005

be used. Second, some *Thiomonas* strains, i.e. CB1, CB3, CB6 and *Tm. arsenivorans*, carry two copies of the *aox* operon. As far as could be ascertained, this is the first example of *aox* gene duplication. Finally, *Thiomonas* sp. 3As is able to grow at pH 3. Several genes potentially involved in acid resistance were found in *Thiomonas* genome. In addition, the Carnoulès toxic environment may cause severe DNA damage in *Thiomonas* sp. 3As, since arsenic is a co-mutagen that inhibits the DNA repair system [41]. DNA repair genes that have been previously shown to be induced in the presence of arsenic in *H. arsenicoxydans* were all found in *Thiomonas* sp. 3As genome, and the expression of *polA* has been shown to be induced in the presence of arsenic [8]. These observations suggest that this bacterium may respond to DNA damage. Nevertheless, we can hypothesize that these stressful conditions may lead to genomic rearrangements in *Thiomonas* genome. This could explain

the important genomic diversity observed among the members of both the 3As species and the *Thiomonas* genus.

At the intra-species level, the dispensable genome defined by comparison of the CB strains with the 3As genome corresponds to 25.3% of *Thiomonas* sp. 3As genome. By comparison, this value is higher than that observed, with the same approach, in other bacteria such as *S. agalactiae* (18%) [46], lower than values calculated in the case of a pathogenic *E. coli* (32.4%) [47], and similar to the value obtained in *Bacillus subtilis* (27%) [48]. The value calculated for *Thiomonas* 3As and CB strains is very high, considering that these strains were isolated from the same site, closely related, and appear to share a recent common ancestor, as illustrated by our phylogenetical analyses. Consequently, we observed that despite strong sequence identities of housekeeping genes such as 16S rRNA or *rpoA*, the whole genome DNA-DNA

hybridization value was relatively low, close to or less than 70%, for strains CB2 and CB3. Conventionally, this should indicate that these bacteria belong to separate evolutionary lineages and must be considered as different species [49]. However, the 16S rRNA-*rpoA* based analysis and CGH experiments revealed that the low DNA-DNA hybridization value correlates with the duplication or absence of several GEIs in these strains. Consequently, we proposed that despite low DNA-DNA hybridization values, these five strains do indeed belong to the same species. Similarly, the DNA-DNA hybridization values obtained with *Thiomonas* sp. 3As as compared to strains Ynys1 and *Tm. perometabolis* were very low, as previously observed [5,10]. Altogether, the great genetic diversity observed in the present study by CGH experiments revealed that DNA-DNA hybridization method may not be appropriate to evaluate evolutionary lineages in *Thiomonas* strains. In this respect the CGH approach seems to be a reliable phylogenetic tool for typing these strains, as suggested in previous studies on other bacteria [47,50].

19 GEIs constitute a large flexible pool of accessory genes that encode adaptive traits. Some of these genes are not required for survival in AMD, since they were not found in all AMD-originated strain genome and correspond therefore to the dispensable gene pool. On the other hand, CGH-based clustering analysis revealed a significant relationship between 3As, CB1-6 and “*Tm. arsenivorans*”, which originate from geographically distinct but similarly arsenic-rich environments. The *Thiomonas* sp. 3As strain and “*Tm. arsenivorans*” form two distinct groups on the basis of phylogenetical, physiological and genetic analyses. Nevertheless, the percent of flexible CDS of *Thiomonas* sp. 3As with “*Tm. arsenivorans*”, is relatively low (18.4%), as compared to the value obtained with Ynys1 and *Tm. perometabolis* (37.9% and 53.6%, respectively). This value obtained with “*Tm. arsenivorans*” was in the same order of magnitude as the value obtained with CB2 and CB3 (23% and 23.6%, respectively). Altogether, 70% of the *Thiomonas* sp. 3As genome was conserved among all strains originated from AMD. Interestingly, two GEIs were conserved or duplicated in all these strains originated from AMD, i.e. ThGEI-O that carries the arsenic-specific operons *ars2* and *aox*, and genes involved in heavy metal resistance, and ThGEI-L that carries several genes involved in heavy metal resistance, biofilm formation and/or motility. Therefore, these GEIs shared by these species are presumably part of the AMD-originated *Thiomonas* core genome. This observation suggests that the acquisition or loss of these GEIs contributes to the evolution of this subgroup of the *Thiomonas* genus and that the evolution of *Thiomonas* strains has been influenced by their similar environments.

Several observations suggest that *Thiomonas* genome evolved by acquiring GEIs through horizontal gene transfer or by genome rearrangement. In the case of two islands, ThGEI-L and -O, an in-deep phylogenetic analysis revealed that these islands have a composite structure probably due to secondary acquisition or losses/rearrangements of some genes. In the case of other GEIs, the existence of HGT is suggested by the fact that genes form syntenic blocks and their GC% were divergent from the rest of the genome. Three mechanisms, i. e. conjugation, transduction and natural transformation, known to be involved in HGT in bacteria [14] may explain GEIs acquisition in *Thiomonas*. One prophage was found in the *Thiomonas* genome and may contribute to horizontal gene transfer, as previously shown in pathogenic bacteria such as *Vibrio cholera*, *Yersinia pseudotuberculosis*, *Bartonella* [51–54], Cyanobacteria [55,56] or for the transfer of pathogenic island from *S. aureus* to *Listeria monocytogenes* [57]. In addition, genes encoding Type IV secretion systems (T4SS) were carried by the pTHI plasmid and the ThGEI-L. It has been recently proposed

that GEI-type T4SS are involved in the propagation of GEIs [45,58]. Therefore, it could also be possible that *Thiomonas* acquired such islands by conjugation. Orthologs of *Neisseria gonorrhoeae* genes involved in natural transformation [59] were also found in *Thiomonas* sp. 3As genome, i.e. the *pil* genes encoding a type IV pili components and *comALMP*. This suggests that *Thiomonas* strains are able to acquire exogenous DNA. Finally, several observations suggest that *Thiomonas* genome has undergone genomic rearrangements contributing to its evolution, as illustrated for the two GEIs, ThGEI-L and -O. Such rearrangements may be promoted by repeat sequences or duplications, that are at the origin of recombination [60]. Indeed, repeats sequences represent 7.62% of the *Thiomonas* sp. 3As genome, and some of the loci found in the GEIs are duplicated with high sequence identities, as ThGEI-C and ThGEI-S that are almost identical. In addition, several IS elements are highly similar, sharing more than 70% nucleotide identity. Interestingly, the majority of the ISs present in the *Thiomonas* sp. 3As genome are found in GEIs. This observation suggests that ISs duplication has played a significant role in both assembly and evolution of these islands, or participated in GEI reshuffling. Altogether, conjugation, transduction, natural transformation and recombination may be at the origin of the high genomic content divergence observed among *Thiomonas* strains.

In conclusion, evidences presented here suggest that *Thiomonas* sp. 3As has acquired some of its particular capacities that contribute to its survival and proliferation in AMD by horizontal gene transfer and genomic rearrangement. Furthermore, these data revealed a high degree of genetic variability within the *Thiomonas* genus, even at the intra-species level. Indeed, the analysis of duplications and deletions of GEIs in several strains revealed the huge significance of these GEIs in the evolution of the *Thiomonas* genus, as well as the influence of the natural environment on the genomic evolution of this extremophile. The majority of intra or inter-species comparisons carried out thus far have concerned pathogens. Our analysis shows that GEIs play also an important role in the evolution of environmental isolates exposed to toxic elements.

Materials and Methods

Bacterial Strains

Thiomonas sp. 3As was obtained from the acidic waters draining the Carnoulès mine tailings, southeastern France [5]. *Thiomonas* strains CB1, CB2, CB3 and CB6 were isolated from the same site: briefly, the isolates were purified by repeated single colony isolation on either R2A medium (Difco; strains CB1, CB2 and CB3) or 100:10 medium ([61]; strain CB6). Physiological, phylogenetic and genetic analyses of these four strains were performed as described previously (Tables S2, S6, [8]). Strains *Thiomonas* Ynys1 [62], *Tm. perometabolis* [63] and “*Tm. arsenivorans*” [9] were cultivated as previously described [8]. DNA-DNA hybridization was carried out as described by [64] under consideration of the modifications described by [65] using a model Cary 100 Bio UV/VIS-spectrophotometer equipped with a peltier-thermostated 6×6 multicell changer and a temperature controller with *in situ* temperature probe (Varian).

DNA Preparation, Sequencing, and Annotation

DNA was extracted and purified from liquid cultures of pure isolates as previously described [8]. The complete genome sequence of *Thiomonas* sp. 3As was determined using the whole-genome shotgun method. Three libraries were constructed, two plasmids and one BAC to order contigs, as previously described [38]. From these libraries, 26,112, 7,680 and 3,840 clones were

end-sequenced, and the assembly was performed with the Phred/Phrap/Consed software package (www.phrap.com), as described previously [38]. An addition of 3,292 sequences was needed for the finishing phase. Coding sequences were predicted as previously described [38]. Putative orthology relationships between two genomes were defined by gene pairs satisfying either the Bidirectional Best Hit criterion or an alignment threshold (at least 40% sequence identity over at least 80% of the length of the smallest protein). These relationships were subsequently used to search for conserved gene clusters (synteny groups) among several bacterial genomes using an algorithm based on an exact graph-theoretical approach [66]. This method allowed for multiple correspondences between genes, detection of paralogy relationships, gene fusions, and chromosomal rearrangements (inversion, insertion/deletion). The 'gap' parameter, representing the maximum number of consecutive genes that are not involved in a synteny group, was set to five. Manual validation of automatic annotations was performed in a relational database (ArsenoScope, <https://www.genoscope.cns.fr/agc/mage/wwwpkgdb/MageHome/index.php?webpage=mage>) using the MaGe web interface [67]. The EMBL (<http://www.ebi.ac.uk/embl>) accession numbers for the genome of *Thiomonas* sp. 3As are FP475956 (chromosome) and FP475957 (plasmid).

Comparative Genome Hybridization (CGH) Array

A custom 385K array for the *Thiomonas* sp. 3As chromosome and plasmid was designed and constructed by NimbleGen Systems. This DNA array encompasses 3,645 CDS of the 3As genome. Probe length was 50 nt and current mean probe spacing was 7 nt. Genomic DNAs from all strains were extracted with the Wizard Genomic DNA Purification Kit (Promega). DNA samples were labeled and purified using the BioPrime Array CGH Genomic Labeling System protocol (Invitrogen). Test (Cy3-labeled) and reference (*Thiomonas* sp. 3As genomic DNA, Cy5-labeled) genomic DNAs were combined (400 pmol fluorescent dye each) and were co-hybridized to the array for 16 h at 42°C in a MAUI Hybridization System (BioMicro System) and slides were washed according to NimbleGen's recommendations. Dye swap experiments were used to compare *Thiomonas* sp. 3As and *Thiomonas* sp. CB1 genomic DNAs. Arrays were scanned with an Axon 4000B scanner. Data were acquired and analyzed using NimbleScan 2.0 and SignalMap 1.9 software (Roche, NimbleGen) and analyzed using the Partek Genomics Suite software (Partek Incorporated, St. Louis, Missouri, U.S.A.). Briefly, log₂-ratios (Cy5/Cy3) were calculated using the segMNT algorithm and gains and losses of genomic material were identified using Partek Genomics Suite Software, as follows: the files were imported and normalized with the qspline normalization [68] by NimbleScan. These data were then imported into the Partek Genomics Suite Soft. The segmentation was performed using the circular binary segmentation algorithm from Olshen *et al.* [69]. Permutations are used to provide the reference distribution to check a second time. 1000 permutations are run using the Partek software. If the resulting p-value is below the threshold ($p = 0.01$), then a breakpoint is added. To verify deletions, PCRs were performed as described in supplementary materials, using primers designed to anneal at the borders of the expected deletions. The CGH data are available in the database ArrayExpress, with the accession number E-MEXP-2260.

Phage Excision and Electron Microscopy

Phage formation was induced by treating exponential cultures with mitomycin C (0.5 µg/mL) for 24 h. The suspension was negatively stained with 16% ammonium molybdate for 10 seconds

and dried over Formvar-coated nickel grids. Grids were examined at 40,000-fold magnification using a Hitachi 600 transmission electron microscope at 75 kV and photographed using a Hamamatsu ORCA-HR camera (Hamamatsu City, Shizuoka, Japan) with the AMT software (Advanced Microscopy Techniques Corp., Danvers, MA).

Phylogenetic and Correspondence Analysis

For each CDS, homologues were searched in NCBI databases. The 300 sequences with the best score were aligned using ClustalW [70]. Alignments were checked by hand and positions with more than 5% of gaps were automatically removed. Neighbor-Joining trees were constructed and analyzed to determine the evolutionary origin of each CDS (Table S4). The correspondence analysis (COA) [71] was performed using the library FactoMineR (<http://factominer.free.fr>) from the statistical package R (<http://www.r-project.org>). For all annotated genes of *Thiomonas* sp. 3As, we determined all the relative synonymous codon usage values [72] obtaining a matrix where the rows represent the genes and the 57 columns are the RSCU values for individual codons. As usual, the 3 TER codons were excluded from the analysis. Codons corresponding to Cystein (TGC/TGT) and the dipeptide of Arginine (AGG/AGA) were also removed from the analysis as they induce systematic artefactual biases [73].

Supporting Information

Figure S1 Circular representations of *Thiomonas* chromosome and pTHI plasmid. Gene organization found in (A) the *Thiomonas* sp. 3As chromosome. The localisation of the 19 GEIs (A-S) is schematized with grey triangles; (B) the plasmid pTHI. Circles display (from the outside): (1) GC percent deviation (GC window - mean GC) in a 1000-bp window; (2) Predicted CDSs transcribed in the clockwise direction (red); (3) Predicted CDSs transcribed in the counterclockwise direction (blue); (4) GC skew (G+C/G-C) in a 1000-bp window; (5) Transposable elements (pink) and pseudogenes (grey).

Found at: doi:10.1371/journal.pgen.1000859.s001 (26.67 MB TIF)

Figure S2 Duplication of genes found in the several GEIs. Only those with high amino acid identities are shown.

Found at: doi:10.1371/journal.pgen.1000859.s002 (0.95 MB TIF)

Figure S3 Factor maps obtained by crossing the first and second axes of the correspondence analysis computed on 3,632 *Thiomonas* sp. 3As genes. For clarity, genes that are not harbored in the 19 GEIs (defined in Table S3) are not represented.

Found at: doi:10.1371/journal.pgen.1000859.s003 (0.94 MB TIF)

Figure S4 Phylogenetic trees of arsenic specific genes compared to *rpoA*. blue: gamma-Proteobacteria; brown: Hydrogenophilales; orange: Methylophilales; light green: Nitrosomonadales; deep green: Rhodocyclales; red: Neisseriales; pink: Burkholderiales. (A) *rpoA*; (B) *aoxB*; (C) *arsB*.

Found at: doi:10.1371/journal.pgen.1000859.s004 (2.19 MB TIF)

Table S1 List of ISs present in *Thiomonas* sp. 3As genome and plasmid.

Found at: doi:10.1371/journal.pgen.1000859.s005 (0.05 MB XLS)

Table S2 Summary of physiological and genetic data obtained from the *Thiomonas* strains used in this study.

Found at: doi:10.1371/journal.pgen.1000859.s006 (0.04 MB DOC)

Table S3 Genomic Islands (GEIs) or islets found in the *Thiomonas* sp. 3As genome.

Found at: doi:10.1371/journal.pgen.1000859.s007 (0.04 MB XLS)

Table S4 Phylogenetic analysis of the 75 genes contained in the ThGEI-L island.

Found at: doi:10.1371/journal.pgen.1000859.s008 (0.04 MB XLS)

Table S5 Phylogenetic analysis of the 121 genes contained in the ThGEI-0 island.

Found at: doi:10.1371/journal.pgen.1000859.s009 (0.05 MB XLS)

Table S6 PCR targets and GenBank Accession IDs of strains used in this study.

Found at: doi:10.1371/journal.pgen.1000859.s010 (0.03 MB DOC)

Acknowledgments

Tm. perometabolis was obtained from the Pasteur Institute, Paris, France. Ynys1 was kindly donated by Kevin Hallberg and D. B. Johnson (Bangor University, North Wales, UK).

Author Contributions

Conceived and designed the experiments: FAP JYC JW CM PNB. Performed the experiments: SK MM VB CGB ME CP. Analyzed the data: FAP SK MM MC VB CBA MB FBB OB JCA SC AHS FH CJ EK AL DL CM DM PO PS DR DS ET SW PNB. Contributed reagents/materials/analysis tools: VB CP ZR GS. Wrote the paper: FAP PNB.

References

- Johnson DB, Hallberg KB (2003) The microbiology of acidic mine waters. *Res Microbiol* 154: 466–473.
- Hallberg KB, Johnson DB (2005) Microbiology of a wetland ecosystem constructed to remediate mine drainage from a heavy metal mine. *Sci Total Environ* 338: 53–66.
- Bruneel O, Personné JC, Casiot C, Leblanc M, Elbaz-Poulichet F, et al. (2003) Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J Appl Microbiol* 95: 492–499.
- Battaglia-Brunet F, Dictor MC, Garrido F, Crouzet C, Morin D, et al. (2002) An arsenic(III)-oxidizing bacterial population: selection, characterization, and performance in reactors. *J Appl Microbiol* 93: 656–667.
- Duquesne K, Lieutaud A, Ratouchniak J, Muller D, Lett MC, et al. (2008) Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. *Environ Microbiol* 10: 228–237.
- Coupland K, Battaglia-Brunet F, Hallberg KB, Dictor MC, Garrido F, et al. (2004) Oxidation of iron, sulfur and arsenic in mine waters and mine wastes: an important role of novel *Thiomonas* spp. In: Tsezos AHM, Remondaki E, eds. *Biohydrometallurgy: a sustainable technology in evolution*. Zografou, Greece: National Technical University of Athens. pp 639–646.
- Moreira D, Amils R (1997) Phylogeny of *Thiobacillus cuprinus* and other mixotrophic thiobacilli: proposal for *Thiomonas* gen. nov. *Int J Syst Bacteriol* 47: 522–528.
- Bryan CG, Marchal M, Battaglia-Brunet F, Klugler V, Lemaitre-Guillier C, et al. (2009) Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse adaptation processes. *BMC Microbiol* 9: 127.
- Battaglia-Brunet F, Joulian C, Garrido F, Dictor MC, Morin D, et al. (2006) Oxidation of arsenite by *Thiomonas* strains and characterization of *Thiomonas arsenivorans* sp. nov. *Antonie van Leeuwenhoek* 89: 99–108.
- Katayama Y, Uchino Y, Wood AP, Kelly DP (2006) Confirmation of *Thiomonas delicata* (formerly *Thiobacillus delicatus*) as a distinct species of the genus *Thiomonas* Moreira and Amils 1997 with comments on some species currently assigned to the genus. *Int J Syst Evol Microbiol* 56: 2553–2557.
- Hallberg KB, Johnson DB (2003) Novel acidophiles isolated from moderately acidic mine drainage waters. *Hydrometallurgy* 71: 139–148.
- Casiot C, Morin G, Juillot F, Bruneel O, Personné JC, et al. (2003) Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* 37: 2929–2936.
- Bentley S (2009) Sequencing the species pan-genome. *Nat Rev Microbiol* 7: 258–259.
- Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, et al. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33: 376–393.
- Maiques E, Ubeda C, Tormo MA, Ferrer MD, Lasa I, et al. (2007) Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. *J Bacteriol* 189: 5608–5616.
- Lesic B, Carniel E (2005) Horizontal transfer of the high-pathogenicity island of *Yersinia pseudotuberculosis*. *J Bacteriol* 187: 3352–3358.
- Sentchilo V, Ravatn R, Werlen C, Zehnder AJ, van der Meer JR (2003) Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. strain B13. *J Bacteriol* 185: 4530–4538.
- Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54: 641–679.
- Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* 11: 472–477.
- Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R (2005) The microbial pan-genome. *Curr Opin Genet Dev* 15: 589–594.
- Achtman M, Wagner M (2008) Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* 6: 431–440.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323: 741–746.
- Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, et al. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci U S A* 102: 13950–13955.
- Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, et al. (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 3: e231. doi:10.1371/journal.pgen.0030231.
- Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, et al. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 311: 1768–1770.
- Costechareyre D, Bertolla F, Nesme X (2009) Homologous recombination in *Agrobacterium*: potential implications for the genomic species concept in bacteria. *Mol Biol Evol* 26: 167–176.
- Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. *Nature* 409: 1092–1101.
- Schluter A, Szczepanowski R, Puhler A, Top EM (2007) Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *FEMS Microbiol Rev* 31: 449–477.
- Nascimento AM, Chartone-Souza E (2003) Operon *mer*: bacterial resistance to mercury and potential for bioremediation of contaminated environments. *Genet Mol Res* 2: 92–101.
- Friedrich CG, Bardschewsky F, Rother D, Quentmeier A, Fischer J (2005) Prokaryotic sulfur oxidation. *Curr Opin Microbiol* 8: 253–259.
- Frigaard NU, Dahl C (2009) Sulfur metabolism in phototrophic sulfur bacteria. *Adv Microb Physiol* 54: 103–200.
- Slonczewski JL, Fujisawa M, Dopson M, Krulwich TA (2009) Cytoplasmic pH measurement and homeostasis in bacteria and archaea. *Adv Microb Physiol* 55: 1–79, 317.
- Baker-Austin C, Dopson M (2007) Life in acid: pH homeostasis in acidophiles. *Trends Microbiol* 15: 165–171.
- Stingl K, Altendorf K, Bakker EP (2002) Acid survival of *Helicobacter pylori*: how does urease activity trigger cytoplasmic pH homeostasis? *Trends Microbiol* 10: 70–74.
- Iyer R, Iverson TM, Accardi A, Miller C (2002) A biological role for prokaryotic ClC chloride channels. *Nature* 419: 715–718.
- Silver S, Phung le T (2005) A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. *J Ind Microbiol Biotechnol* 32: 587–605.
- Harrison JJ, Ceri H, Turner RJ (2007) Multimetal resistance and tolerance in microbial biofilms. *Nat Rev Micro* 5: 928–938.
- Muller D, Médigue C, Koechler S, Barbe V, Barakat M, et al. (2007) A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS Genet* 3: e53. doi:10.1371/journal.pgen.0030053.
- Silver S, Phung LT (2005) Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Appl Environ Microbiol* 71: 599–608.
- Meng YL, Liu Z, Rosen BP (2004) As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*. *J Biol Chem* 279: 18334–18341.
- Hartwig A, Blessing H, Schwerdtle T, Walter I (2003) Modulation of DNA repair processes by arsenic and selenium compounds. *Toxicology* 193: 161–169.
- Schwerdtle T, Walter I, Mackiw I, Hartwig A (2003) Induction of oxidative DNA damage by arsenite and its trivalent and pentavalent methylated metabolites in cultured human cells and isolated DNA. *Carcinogenesis* 24: 967–974.
- Schirmer T, Jenal U (2009) Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol* 7: 724–735.
- Hengge R (2009) Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* 7: 263–273.

45. Burrus V, Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol* 155: 376–386.
46. Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, et al. (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* 99: 12391–12396.
47. Fukiya S, Mizoguchi H, Tobe T, Mori H (2004) Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J Bacteriol* 186: 3911–3921.
48. Earl AM, Losick R, Kolter R (2007) *Bacillus subtilis* genome diversity. *J Bacteriol* 189: 1163–1170.
49. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, et al. (1987) Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* 37: 463–464.
50. Binnewies TT, Motro Y, Hallin PF, Lund O, Dunn D, et al. (2006) Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct Integr Genomics* 6: 165–185.
51. Li M, Kotetishvili M, Chen Y, Sozhamannan S (2003) Comparative genomic analyses of the *Vibrio* pathogenicity island and cholera toxin prophage regions in non-epidemic serogroup strains of *Vibrio cholerae*. *Appl Environ Microbiol* 69: 1728–1738.
52. Derbise A, Chenal-Francois V, Pouillot F, Fayolle C, Prevost MC, et al. (2007) A horizontally acquired filamentous phage contributes to the pathogenicity of the plague *Bacillus*. *Mol Microbiol* 63: 1145–1157.
53. Brussow H, Canchaya C, Hardt WD (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 68: 560–602.
54. Berglund EC, Frank AC, Calteau A, Vinnere Pettersson O, Granberg F, et al. (2009) Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. *PLoS Genet* 5: e1000546. doi:10.1371/journal.pgen.1000546.
55. Dammeyer T, Bagby SC, Sullivan MB, Chisholm SW, Frankenberg-Dinkel N (2008) Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr Biol* 18: 442–448.
56. Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G, et al. (2005) Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ Microbiol* 7: 1505–1513.
57. Chen J, Novick RP (2009) Phage-mediated intergeneric transfer of toxin genes. *Science* 323: 139–141.
58. Juhas M, Crook DW, Hood DW (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol* 10: 2377–2386.
59. Hamilton HL, Dillard JP (2006) Natural transformation of *Neisseria gonorrhoeae*: from DNA donation to homologous recombination. *Mol Microbiol* 59: 376–385.
60. Treangen TJ, Abraham AL, Touchon M, Rocha EP (2009) Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol Rev* 33: 539–571.
61. Schrader JA, Holmes DS (1988) Phenotypic switching of *Thiobacillus ferrooxidans*. *J Bacteriol* 170: 3915–3923.
62. Hallberg KB, Johnson DB (2003) Novel acidophiles isolated from moderately acidic mine drainage waters. *Hydrometallurgy* 71: 139–148.
63. London J, Rittenberg SC (1967) *Thiobacillus perometabolis* nov. sp., a non-autotrophic thiobacillus. *Arch Mikrobiol* 59: 218–225.
64. De Ley J, Cattoir H, Reynaerts A (1970) The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* 12: 133–142.
65. Huss VAR, Festl H, Schleifer KH (1983) Studies on the spectrophotometric determination of DNA hybridization from renaturates rates. *Syst Appl Microbiol* 4: 184–192.
66. Boyer F, Morgat A, Labarre L, Pothier J, Viari A (2005) Syntons, metabolons and interactons: an exact graph-theoretical approach for exploring neighbourhood between genomic and functional data. *Bioinformatics* 21: 4209–4215.
67. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, et al. (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res* 34: 53–65.
68. Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, et al. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* 3: research0048.0041–0048.0016.
69. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5: 557–572.
70. Thompson JD, Plewniak F, Thierry J, Poch O (2000) DbcLustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res* 28: 2919–2926.
71. Benzécri JP (1984) L'analyse des correspondances. Paris: Dunod.
72. Sharp PM, Tuohy TM, Mosurski KR (1986) Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 14: 5125–5143.
73. Perriere G, Thioulouse J (2002) Use and misuse of correspondence analysis in codon usage studies. *Nucleic Acids Res* 30: 4548–4555.

C/ Discussion

Thiomonas 3As possède un chromosome de 3,7 Mb et un plasmide pTHI de 46kb. Le chromosome comprend 3,632 séquences codantes et le contenu en GC (%) est riche sur l'ensemble du génome (64% sur le chromosome, 60% sur le plasmide) (Bryan *et al.*, 2009).

Une analyse fine du génome a permis de dresser un inventaire des potentialités de cette souche. Ainsi, certaines capacités métaboliques sont propres à cette souche par rapport à *H. arsenicoxydans*, on peut citer, les métabolismes du carbone ou énergétique par exemple. Différents donneurs d'électrons organiques ou inorganiques, tels que les composés inorganiques soufrés réduits pourraient donc être utilisés par cette bactérie. Par ailleurs, les gènes impliqués dans les mécanismes décrits précédemment chez *H. arsenicoxydans* tels que la résistance à l'arsenic *via* les gènes *ars*, la formation d'un biofilm, la motilité, la réparation de l'ADN ou encore la réponse au stress oxydatif sont également retrouvés chez cette bactérie.

L'organisation des gènes codant l'arsénite oxydase a été analysée chez *Thiomonas* 3As. Ainsi, les gènes codant cette enzyme sont localisés sur le chromosome et présentent une forte synténie avec ceux de *H. arsenicoxydans*. De plus, les gènes *aoxA* et *aoxB* montrent respectivement 58% et 65% de similarité avec le gène *aoxA* ou *aoxB* de *H. arsenicoxydans*. On retrouve également tout comme chez *H. arsenicoxydans*, un système de régulation à deux composants codé par *aoxRS* ainsi qu'un transporteur du phosphate codé par *pstB* localisés à proximité de l'opéron *aox*. Les gènes *aoxR*, *aoxS* et *pstB* présentent 40%, 32% et 39% de similarité avec leurs homologues chez *H. arsenicoxydans*. Une analyse des domaines a révélé que la protéine AoxR chez *Thiomonas* 3As possède, tout comme *H. arsenicoxydans* (Chapitre 1), un motif de liaison au facteur alternatif σ^{54} . Un gène *rpoN* codant ce facteur a également été localisé sur le chromosome de *Thiomonas* 3As. On retrouve aussi des gènes impliqués dans la biosynthèse d'un cofacteur à molybdoptérine *moeA* et *moeA2*. Ces gènes ont également été identifiés dans le génome de *H. arsenicoxydans*, mais *Thiomonas* 3As est la première souche pour laquelle ils ont été localisés dans l'environnement génétique de l'*aox*. Chez *A. faecalis*, il a été démontré que la grande sous-unité de l'arsénite oxydase AroA possède un noyau molybdène (Ellis *et al.*, 2001). La colocalisation des gènes observée chez *Thiomonas* confirme cette dernière.

Enfin, l'analyse des profils CGH (Comparative Genomic Hybridization) a permis de mettre en évidence que certaines souches de *Thiomonas* (CB1, CB3 et CB6) et *Tm. arsenivorans* possèdent deux copies de l'opéron *aox*. A notre connaissance, il s'agit du premier exemple de duplication de ces gènes au sein d'un même génome.

Dans ce chapitre nous avons donc pu montrer que les souches de *Thiomonas* possèdent des gènes, tels que les opéron *ars* et *aox*, qui contribuent à leur survie et à leur prolifération dans des environnements contaminés par l'arsenic comme les DMA. Malgré leur capacité métabolique particulière vis-à-vis de l'arsénite, les gènes codant l'arsénite oxydase gardent une organisation conservée, par rapport à *H. arsenicoxydans*. Toutefois, on notera la particularité de certaines souches de *Thiomonas* sp. qui possèdent des gènes *moeA* à proximité de l'opéron *aox* et/ou une duplication du cluster *aox*.

Mécanismes globaux du stress arsénié chez
Rhizobium sp. NT-26

A/ Introduction

Dans le chapitre 4, l'analyse des génomes de *Thiomonas* par une approche de génomique descriptive a permis de fournir un aperçu des capacités particulières de ces souches. Néanmoins, l'absence d'outils génétiques disponibles chez cet organisme rend l'étude de son métabolisme de l'arsenic difficile. Aussi, une seconde bactérie chimolithoautotrophe, *Rhizobium* sp. NT-26, a été choisie comme organisme de référence pour ces travaux. En effet, cette bactérie est capable de croître dans un environnement où l'arsénite sert de donneur d'électrons, l'oxygène d'accepteur d'électrons et le dioxyde de carbone (CO₂) ou le bicarbonate (HCO₃⁻) de source de carbone. Il s'agit d'une α -protéobactérie qui appartient aux *Agrobacterium/Rhizobium* et qui a été isolée d'une roche contenant de l'arsénopyrite issue d'une mine d'or du territoire nord-australien (Santini *et al.*, 2000). Elle résiste à de fortes concentrations d'arsenic : 10 mM d'As(III) et à plus de 100 mM d'As(V) (Santini & Vanden Hoven, 2004). Enfin, la disponibilité d'outils génétiques fait de cet organisme un modèle de choix pour l'étude du métabolisme de l'arsenic chez les chimolithotrophes.

Le génome de *Rhizobium* sp. NT-26 a été séquencé, en collaboration avec le Genoscope et j'ai participé à son annotation experte. La première étape a été d'attribuer une fonction au produit des gènes détectés dans le génome. Dans notre cas, la fonction a été attribuée par homologie avec des gènes déjà étudiés, par identification de domaines conservés avec d'autres protéines ou plus occasionnellement par rapport au contexte génétique, à l'aide de l'interface MaGe. Ce processus est essentiel pour la validation des données génomiques. L'objectif a ensuite été d'explorer le génome à la recherche de gènes codant des protéines impliquées dans les mécanismes de résistance à l'arsenic décrits chez *H. arsenicoxydans* et les souches de *Thiomonas* sp. Cette étude a été couplée à des tests physiologiques visant à confirmer les potentialités suggérées par l'analyse du génome.

Toutefois, l'identification d'un cadre ouvert de lecture ne permet pas toujours d'avoir accès à la fonction du gène étudié. De plus, le séquençage du génome fournit uniquement un répertoire de gènes de chaque cellule. La réalité cellulaire, elle, va dépendre de l'emploi de ce répertoire aussi bien en terme de choix des gènes exprimés, de l'accumulation des produits, que de la modulation de la fonctionnalité des protéines. La seconde étape a donc été d'appliquer à *Rhizobium* NT-26, les techniques utilisées précédemment pour l'étude de la réponse au stress arsénié. Pour ce faire, une analyse de l'accumulation des protéines a

été menée par une approche de protéomique différentielle classique couplée à une analyse 2D-DIGE (two-dimensional fluorescence difference gel electrophoresis). En parallèle, une analyse des gènes impliqués dans la régulation de l'activité arsénite oxydase a été menée par une approche de mutagénèse aléatoire.

Un certain nombre de résultats récents tendent à mettre en évidence que l'adaptation d'une cellule au cours d'un processus physiologique, ne passe pas uniquement par « l'allumage » ou « l'extinction » d'une protéine, mais souvent par une modulation plus fine de son accumulation. Aussi, la stratégie de cette étude a été de comparer les protéines présentes dans deux conditions (+ et - As(III)) par une approche de protéomique différentielle par 2D-DIGE. Cette technique est basée sur le marquage direct des lysines des protéines de trois échantillons par des fluorophores, tels que les cyanines dans notre cas (Unlü *et al.*, 1997). Par exemple, l'échantillon en condition témoin est marqué à la cyanine 5 (Cy5TM), l'échantillon en condition de stress (+As(III)), à la cyanine 3 (Cy3TM) et le standard interne, à la cyanine 2 (Cy2TM). Les trois échantillons sont mélangés et une électrophorèse bidimensionnelle est effectuée sur un même gel. L'abondance relative des peptides est comparée sur base de leur intensité de fluorescence. La 2D-DIGE présente l'avantage d'une quantification par analyse d'image (« Typhoon TM »), mais aussi par rapport à une analyse différentielle classique, d'une normalisation simplifiée grâce à l'introduction d'un standard interne.

Ma participation à cette étude a porté sur l'exploitation des données du génome afin de déterminer les potentialités de cet organisme, les tests physiologiques, la mise au point de la protéomique classique et 2D-DIGE et l'analyse de ces données. Actuellement, les données sont complétées par une analyse transcriptomique différentielle. Ces études visent à comprendre si les mécanismes de régulation décrits chez *H. arsenicoxydans* peuvent se transposer à une bactérie chimolithotrophe facultative telle que *Rhizobium* NT-26 sp, mais aussi de déterminer si cette bactérie possède d'autres mécanismes qui lui permettent de coloniser un environnement contaminé par l'arsenic (motilité, biofilm...). Ces travaux dont les données sont présentées ci-dessous, font l'objet d'une publication qui est en cours de rédaction.

B/ Résultats / Discussion

I- Caractéristiques générales du génome de *Rhizobium* NT-26

Rhizobium sp. NT-26 arbore un chromosome de 4,2 Mb et un mégaplasme pNT-26 de 291 kb. Une description de chacun d'entre eux est donnée ci-dessous.

1- Le chromosome de NT-26

Le chromosome contient 4,364 séquences codantes, ce qui représente 90% du génome. Le génome a été comparé aux séquences contenues dans la « Prokaryotic Genome DataBase » (PkgDB) et la banque de données RefSeq (NCBI Reference Sequences). La meilleure syntenie est obtenue avec le chromosome de *Rhizobium leguminosarum*, puisque 51,25 % des CDS de NT-26 sont retrouvés dans des groupes de synténies (syntons) de ce chromosome. Néanmoins, le chromosome circulaire d'*Agrobacterium tumefaciens* C58 est considéré comme étant le plus proche car 67,73% des CDS forment une syntenie avec le chromosome de NT-26 et la taille des syntons est la plus grande. Ces résultats confirment bien son appartenance au groupe des *Agrobacterium/Rhizobium*.

La valeur moyenne du GC (%) est de 62% sur le chromosome. Il existe aussi des régions pour lesquelles le pourcentage en G+C est clairement distinct de la moyenne. Dix régions ont un contenu en G+C plus faible, mais aucune région plus riche en G+C n'a été déterminée. Ces régions ont une taille comprise entre de 7 à 116 kb, leur CAI (**C**odon **A**daptation **I**ndex) est inférieur au CAI moyen et elles contiennent au minimum un segment répété. Enfin, dans certaines de ces régions, des éléments transposables et des gènes codants des ARNt sont présents. Ces caractéristiques suggèrent que ces régions sont des îlots génomiques potentiels (Daubin *et al.*, 2003).

Différents mécanismes décrits précédemment chez *H. arsenicoxydans* et *Thiomonas* sp. sont retrouvés chez cet organisme, parmi lesquels on peut citer : le métabolisme de l'arsenic, la motilité ou la formation de biofilm.

(i) Le métabolisme de l'arsenic

Deux opérons *ars* ont été identifiés dans le génome de NT-26. L'un, *ars1*, est localisé sur le plasmide et le second, *ars2*, sur le chromosome. L'opéron *ars2* est situé dans le même locus que le système Sox (qui sera décrit plus loin). Celui-ci possède un gène additionnel *arsA* qui code une ATPase associée à la pompe d'efflux de l'arsénite. Cette fonction n'est pas indispensable pour la résistance à l'arsenic mais augmente considérablement son efficacité et permet ainsi une expulsion active de l'arsenic (Branco *et al.*, 2008).

(ii) La motilité

Rhizobium NT-26 est dotée de deux flagelles subterminaux. Les gènes impliqués dans la biosynthèse du flagelle sont organisés en un cluster et un certain nombre de ces gènes montre une forte syntenie avec la cascade flagellaire de *Synorhizobium meliloti*. Chez cette dernière, le système à deux composants VisRN constitue l'opéron maître et contrôle les gènes en aval de la cascade. A ce jour, aucun substrat du senseur VisN n'a été identifié. Les gènes *visRN* ainsi que les gènes de classe II tels *flhM* ou *motB* sont retrouvés chez *Rhizobium* NT-26.

Des tests de motilité en présence ou en absence d'As(III) ont été effectués. Une nage de la bactérie a été observée en présence d'As(III) mais en absence de ce métalloïde aucune motilité n'est visible. Ainsi, contrairement à *H. arsenicoxydans* où l'on observe un gain de motilité, ces résultats suggèrent que chez *Rhizobium* NT-26 l'arsénite induit la motilité.

(iii) Le biofilm

Un opéron *big* potentiellement impliqué dans la formation du biofilm a été identifié. Chez *Xyloella fastidiosa* et *Agrobacterium tumefaciens* des opérons homologues à *big* sont induits durant la formation du biofilm ou lors de l'adhérence et une surexpression de celui-ci entraîne une production accrue de biofilm (Barbosa & Benedetti, 2007). A ce jour, mis à part le rôle de BigR comme répresseur transcriptionnel de l'opéron *big*, on ne connaît pas la fonction des produits des gènes *big*.

Par ailleurs, tout comme chez *Thiomonas* (Chapitre 4), l'analyse du génome de *Rhizobium* NT-26 révèle la présence de gènes impliqués dans des métabolismes particuliers. L'exemple de l'oxydation du soufre inorganique est donné ci-dessous. Certains procaryotes sont capables d'oxyder des composés soufrés inorganiques comme le thiosulfate à pH acide, neutre ou alcalin (Dahl, 2008). La voie d'oxydation du thiosulfate est présente chez *Rhizobium* NT-26 et est très similaire à celle de *Pseudaminobacter salicylatoxidans* KTC001 (Lahiri *et al.*, 2006). Sept gènes qui codent cinq protéines clefs du complexe Sox de l'organisme modèle *Paracoccus pantotrophus* ont été retrouvés : *soxXYZABCD*. Les protéines sont les suivantes : SoxXA, SoxYZ, SoxB et SoxCD. Chez *P. pantotrophus*, SoxXA est un cytochrome hétérodimérique de type-c, SoxYZ une protéine hétérodimérique qui établit des liaisons covalentes avec le thiosulfate, SoxB un monomère qui contient deux atomes de manganèse (Rother *et al.*, 2001) et SoxCD un hétérotétramère de deux protéines SoxC qui est une sous-unité contenant un cofacteur à molybdène et SoxD un cytochrome dihème de type c (Wodara *et al.*, 1997). *Rhizobium* sp. NT-26 serait donc capable d'oxyder le thiosulfate selon l'équation suivante : $S-SO_3 + 5H_2O \rightarrow 2SO_4^{2-} + 8e^- + 10H^+$.

Enfin, un opéron *nod* impliqué dans la nodulation a été identifié chez cette bactérie. Chez *Rhizobium* NT-26, tout comme chez *Rhizobium loti*, mais à la différence de *Rhizobium leguminosarum* et *Rhizobium meliloti*, l'information symbiotique serait portée par le chromosome plutôt que le plasmide (Sullivan *et al.*, 1995).

2- Le plasmide de NT-26

Le mégaplasmide contient 326 séquences codantes. La valeur moyenne du GC (%) est de 60%.

(i) La résistance aux métaux et à l'arsenic

Un opéron *mer* codant les protéines MerR, MerT, MerP, MerC, MerA impliquées dans la résistance au mercure, a été identifié. L'opéron *ars1* est situé dans le même locus que l'opéron *mer*. L'arsenic et le mercure peuvent coexister dans les environnements. Aussi il ne paraît pas aberrant de retrouver ces deux mécanismes de résistance au sein d'un même îlot génomique.

Chez *Rhizobium* sp. NT-26, les gènes *aroAB* codant l'arsénite oxydase, sont localisés sur le mégaplasmide à proximité des gènes *aroRS* et en amont d'un gène codant un cytochrome périplasmique *c₅₅₂*. L'arsénite oxydase de *Rhizobium* NT-26 formerait un hétérodimère $\alpha\beta$ localisé dans le périplasme. Par ailleurs, le cytochrome a été caractérisé *in vivo* et *in vitro* comme jouant le rôle d'accepteur d'électrons dans l'oxydation de l'arsénite (Santini et al., 2007), ce qui suggère un rôle énergétique de l'enzyme. Une analyse des domaines de AoxR révèle également un motif de liaison au facteur alternatif σ^{54} chez *Rhizobium* sp. NT-26. Un gène *rpoN* codant ce facteur a d'ailleurs été localisé sur le chromosome.

Tout comme chez *H. arsenicoxydans*, un opéron *pst* codant des protéines impliquées dans le transport du phosphate est localisé à proximité de l'opéron *aro* chez *Rhizobium* NT-26, confirmant cette notion d' « îlot arsenic » chez ce micro-organisme. On notera qu'un gène *moeA* a également été identifié à proximité de l'opéron *aox*, ce qui avait déjà été démontré chez *Thiomonas* 3As. Ce gène présente 35% de similarité avec le gène *moeA* de *Thiomonas* 3As et semblerait être conservé chez les bactéries chimiolithoautotrophes. Néanmoins, à notre connaissance il s'agit de la première bactérie pour laquelle les gènes codant l'arsénite oxydase sont localisés sur un plasmide.

(ii) Système de sécrétion de type IV

Deux opérons *vir* ont été identifiés chez ce micro-organisme. Les protéines Vir sont impliquées dans la formation du système de sécrétion de type IV qui permet la sécrétion d'ADN ou de protéines à travers la membrane par la formation d'un pore dans l'enveloppe cellulaire. Chez *A. tumefaciens* ce système lui permet d'infecter les plantes (Foulongne et al., 2002). Chez NT-26, l'un des opérons ne possède pas les gènes *virB5* nécessaire à la synthèse du pilus, et *virB7*, impliqué dans la formation du canal. Néanmoins, le second opéron situé à environ 140 kb en aval de celui-ci, présente l'ensemble des gènes. Ainsi, la souche NT-26 serait capable d'exprimer un système de sécrétion de type IV.

II- Etude du stress arsénié chez NT-26

L'étude du stress arsénié a été menée par une approche de protéomique 2D-DIGE en présence et en absence d'arsénite. 63 spots ont été repiqués et analysés. L'analyse des résultats a permis de démontrer une accumulation différentielle des protéines énumérées dans la table 1 (page 180).

Sur base de ces données, nous avons pu montrer que tout comme chez *H. arsenicoxydans*, la réponse au stress arsénié est caractérisée par une accumulation de protéines impliquées dans le stress oxydatif, la motilité, le transport du phosphate et le métabolisme arsénié chez cet organisme. Par ailleurs, des protéines impliquées dans des mécanismes particuliers ont aussi été identifiées, tels que l'arsénite oxydase, des protéines impliquées dans la biosynthèse des acides aminés, d'acides gras, divers transporteurs de type ABC. Il semblerait donc y avoir des différences dans les réponses au stress arsénié entre les organismes chimolithoautotrophes facultatifs et les hétérotrophes décrits précédemment. Néanmoins, les résultats de protéome semblent indiquer que l'arsénite oxydase est également soumise à régulation en réponse à l'As(III) chez *Rhizobium* sp. NT-26.

Dans un premier temps, afin d'identifier la concentration d'As(III) permettant une oxydation optimale, un test phénotypique au nitrate d'argent a été effectué à différentes concentrations en As(III) et les diamètres des halos d'oxydation ont été mesurés. Cette technique est largement utilisée dans le cadre de l'étude de l'activité arsénite oxydase et a notamment été utilisée dans les travaux du chapitre 1. Ainsi, un diamètre maximum a été observé pour une concentration de 7,5 mM d'As(III).

En parallèle, une banque de mutants a été construite, puis criblée sur base de leur activité arsénite oxydase. Sur les 6000 clones testés, sept clones ne forment pas de précipité brun et ont été considérés comme des mutants de l'activité arsénite oxydase. Afin de déterminer la nature des gènes impliqués dans l'activité oxydase chez ces sept mutants, leur ADN a été extrait. Il a ensuite été digéré, ligué, amplifié par PCR inverse et le produit a été séquencé.

Pour cinq de ces sept mutants, le transposon a pu être localisé dans les gènes *aroA*, *aroB*, *aroR*, *ptxB* et *moeB*. Les trois premiers sont localisés sur le plasmide et le dernier sur le chromosome. L'exploitation des données accessibles sur la plateforme MaGe (Magnifying Genomes, www.genoscope.cnrs.fr) a permis de déterminer les fonctions associées aux produits des gènes, ainsi que les similarités par rapport à des gènes homologues présents dans d'autres organismes.

Ainsi, les gènes *aroA* et *aroB* codent respectivement pour la grande et la petite sous-unité de l'arsénite oxydase. Le gène *aroA* présente 47% de similarité avec *aoxB* d'*H. arsenicoxydans* et 46% avec *aoxB* de *Thiomonas* 3As. Le gène *aroB* présente quant à lui 45% de similarité avec *aoxA* de *H. arsenicoxydans* et 44% avec *aoxA* de *Thiomonas* 3As.

Le gène *aroR* code le régulateur du système à deux composants *aroRS* localisé en amont de l'opéron *aro*. Il présente 38% de similarité avec *aoxR* de *H. arsenicoxydans* et 44% avec *aoxR* de *Thiomonas* 3As.

Le gène *ptxB* code une protéine impliquée dans la liaison du phosphite. Il présente 44% et 40% de similarité respectivement avec *ptxB* de *H. arsenicoxydans* et *Thiomonas* 3As.

Le gène *moeB* code une protéine impliquée dans la synthèse d'un cofacteur à molybdoptérine. Il présente 42% de similarité avec *thiF* de *H. arsenicoxydans* et 44% avec *moeB* de *Thiomonas* 3As. Par ailleurs, une analyse des domaines de cette protéine révèle notamment la présence d'un motif caractéristique de MoeB.

L'approche génétique a permis de dresser un inventaire au moins partiel des mécanismes qui régissent le contrôle de l'oxydation de l'As(III) chez *Rhizobium* sp. NT-26. En effet, trois groupes de gènes bien individualisés ont pu être caractérisés :

(i) Premièrement, les gènes ayant un rôle dans la structure de l'arsénite oxydase elle-même, ce qui inclut les gènes *aroA* et *aroB*.

(ii) Ensuite, un gène impliqué dans la régulation des gènes codant pour l'arsénite oxydase : *aroR*.

Les gènes *aroAB* et *aroR* ont déjà été caractérisés lors d'études de l'oxydation bactérienne de l'As(III), notamment chez *H. arsenicoxydans* dans le chapitre 1, en accord avec le modèle conceptuel de la régulation de l'opéron *aox* faisant notamment intervenir le système à deux composants *aoxRS* (*aroRS*).

(iii) Enfin, un dernier groupe inclut les gènes non encore identifiés dans les études précédentes : *ptxB* et *moeB*.

Le gène *ptxB* est fréquemment retrouvé dans l'environnement génétique des gènes codant pour l'arsénite oxydase, notamment chez *H. arsenicoxydans* et *Thiomonas* 3As. Chez *Achromobacter* sp. SY8, ce gène est dénommé *aoxX* et est cotranscrit avec les gènes *aoxRS*. La protéine AoxX serait impliquée directement ou indirectement dans la perception de l'As(III) chez ce micro-organisme. Dans ce chapitre, nous avons montré pour la première fois que le gène *ptxB* est indispensable pour établir une activité arsénite oxydase chez *Rhizobium* sp. NT-26.

Le gène *moeB* code pour un cofacteur à molybdoptérine. Dans le chapitre 1, nous avons montré qu'un transposon localisé dans un gène codant un transporteur du molybdène (*modB* et *modC*) affecte l'activité arsénite oxydase chez *H. arsenicoxydans*. Néanmoins, l'activité est restaurée par ajout de molybdène dans le milieu de culture. Nous

avons donc testé cette hypothèse pour le mutant dont l'insertion est localisée dans le gène *moeB*. Les résultats obtenus montrent qu'aucune activité n'est observée même après ajout de molybdène. Il s'agit de la première étude qui permet de caractériser le rôle de ce gène dans l'oxydation de l'As(III).

La protéine MoeB serait donc indispensable à l'activité arsénite oxydase, malgré la présence des gènes impliqués dans le transport du molybdène dans le génome de *Rhizobium* sp. NT-26. La grande sous-unité AroA ayant un noyau molybdène, il n'est pas suprenant que ce gène soit nécessaire à l'activité.

Afin de préciser le rôle de l'arsénite oxydase chez une bactérie chimolithoautotrophe, la résistance à l'As(III) des différents mutants a été testée. Les résultats obtenus montrent que les mutants ne sont pas plus sensibles à cet élément que la souche sauvage. Les opérons *ars* seraient donc suffisants pour résister à l'As(III) chez *Rhizobium* sp. NT-26, tout comme chez *H. arsenicoxydans*. Par ailleurs, en condition d'autotrophie, les résultats préliminaires semblent indiquer que les mutants de l'arsénite oxydase sont incapables de croître avec l'As(III) comme seule source d'électrons. Enfin, chez *H. arsenicoxydans* une relation entre l'oxydation de l'arsénite et la motilité avait été démontrée (Chapitres 1 à 3). Ce phénotype a donc également été testé chez *Rhizobium* sp. NT-26, mais aucune différence de nage n'a été observée entre les mutants et la souche sauvage.

Chez *Rhizobium* sp. NT-26 l'arsénite oxydase serait donc uniquement dédiée à sa croissance et serait soumise à une régulation faisant intervenir AoxR tout comme chez *H. arsenicoxydans*, mais aussi des régulateurs non encore identifiés en réponse au stress arsénié, tels que MoeB et PtxB.

Table 1 : Protéines accumulées différemment en réponse à l'As(III) chez *Rhizobium* sp. NT-26. Les protéines identifiées par protéomique classique ou par 2D-DIGE sont représentés respectivement en noir et en vert. Les protéines induites apparaissent en gras.

Label	Gene	Spot number	Function	Theoretical		
				Fold Change	pI	Molecular weight (kDa)
A. CELL ENVELOPE AND CELLULAR PROCESSES						
20364		3	Putative oligoendopeptidase F	2,5	5,4	70
20398	<i>cbbK</i>	13	Phosphoglycerate kinase	2,2	4,9	42
20401	<i>cbbT</i>	1	Transketolase	3,4	5,5	70
30042	<i>ilvD</i>	3	Dihydroxy-acid dehydratase	2,5	5,3	67
30043	<i>chvE</i>	44	Multiple sugar binding periplasmic receptor	3,7	4,5	38
30111	<i>murE</i>	7	UDP-N-acetylmuramoylalanyl-D-glutamate 2,6-diaminopimelate ligase	2,8	5,6	51
30112	<i>murF</i>	7	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase	2,8	5,7	51
30119	<i>ddl</i>	19	D-alanine-D-alanine ligase	2,9	5,5	33
50172	<i>lpxA</i>	34	Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase	2,9	6	29
50480	<i>pyrC</i>	59	Dihydroorotase (DHOase)	2,4	5	46
50531	<i>clpP</i>	22	ATP-dependent Clp protease proteolytic subunit (Endopeptidase Clp)	4,3	5,6	23
50618	<i>purN</i>	34	Phosphoribosylglycinamide formyltransferase (GART) (GAR transformylase) (5'-phosphoribosylglycinamide transformylase)	2,9	5,9	24
50650	<i>gmk</i>	37	Guanylate kinase	2,3	5,6	25
60206	<i>sodB</i>	22, 30	Superoxide dismutase, iron/manganese cofactor	3	5,8	23
70289	<i>tldD</i>	11	Peptidase TldD	3,7	5,5	50
70336	<i>purD</i>	10	Phosphoribosylamine-glycine ligase	6,4	4,9	44
70389	<i>groEL</i>	4	60 kDa chaperonin (Protein Cpn60) (GroEL protein)	2,3	5	58
70573	<i>fla</i>	25	Flagellin	9,7	4,2	40
70575	<i>flaA</i>	17, 25	Flagellin A	7,2	4,2	41
70615	<i>hisD</i>	10	Histidinol dehydrogenase	6,4	4,9	45
70617	<i>murA</i>	62	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Enolpyruvate transferase 1) (UDP-N-acetylglucosamine enolpyruvyl transferase 1) (EPT 1)	-2,1	5,5	46
90077		6	Putative peptidase, M20 family	3,3	4,8	64
90090	<i>dnaK</i>	2	Chaperone protein DnaK	2,5	4,9	68
p10020	<i>aroA</i>	24, 26, 29	Arsenite oxidase large subunit AroA; [3Fe-4S] cluster, Molybdopterin cofactor-binding active site	10	5,6	93
p10108	<i>arsC1</i>	23	Arsenate reductase (Arsenical pump modifier)	22	5,1	16
p10112	<i>arsH</i>	34	NADPH-dependent FMN reductase; ArsH-like protein	2,9	5,9	26
50634	<i>ndk</i>	43	Nucleoside diphosphate kinase	-3,7	5,5	15
80028	<i>pyrC</i>	56	Dihydroorotase (DHOase)	-2,6	5,6	38
80043	<i>dapD</i>	55	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (Tetrahydronicolinate N-succinyltransferase) (THP succinyltransferase) (Tetrahydropicolinate succinylase)	-2,6	5,3	30
80216	<i>deoB</i>	62	Phosphopentomutase	-2,1	5,4	43
100022	<i>pckA</i>	45, 52	Phosphoenolpyruvate carboxykinase [ATP] (PEP carboxykinase) (Phosphoenolpyruvate carboxylase) (PEPCK)	-3	5,3	58
100183	<i>dgoK</i>	57	2-dehydro-3-deoxygalactonokinase protein	-2,5	5,4	33
100646	<i>idhA</i>	60	Inositol 2-dehydrogenase	-2,3	5,7	35
100696	<i>hadH</i>	40	3-hydroxyacyl-CoA dehydrogenase	-2,3	5,8	26
B. TRANSPORT AND BINDING PROTEIN						
20095		5	Putative periplasmic binding ABC transporter protein, probable sugar binding precursor, putative signal peptide	3,1	4,8	64
20210		5, 42	Putative ABC transporter substrate binding protein	3,5	4,7	59

20216		36, 48	ABC transporter	-2,8		
20230		5, 42	Putative oligopeptide ABC transporter	3,5	4,7	62
40191		17	Putative ABC transporter (substrate-binding protein)	4,6	5	39
70514	<i>fbpA</i>	17	Iron(III) ABC transporter, periplasmic iron(III)-binding protein	4,6	4,8	38
80161		18	Putative periplasmic mannitol-binding protein; putative Trap-T transport system, DctP subunit	2	6,1	40
90245	<i>frcB</i>	44	ABC transporter, substrate binding protein (sugar)	3,7	4,6	35
100649		25	Putative rhizopine-binding protein, mocB	9,7	4,2	34
100653		44	Putative ribose periplasmic binding protein	3,7	4,5	36
p10014	<i>pstS2</i>	44	Phosphate ABC transporter periplasmic binding protein	3,7	4,5	37
20032		54	Putative TRAP dicarboxylate transporter	-2,6	4,5	35
20541	<i>livK</i>	54	High affinity branched chain amino acid ABC transporter	-2,6	4,6	39
50080		49	ABC transporter	-2,9	4,5	28
50771		45	Putative oligopeptide ABC transporter (substrate-binding protein)	-3,1	5,8	61
100263	<i>dppA</i>	46, 47, 48	Periplasmic dipeptide transport protein precursor (Dipeptide-binding protein) (DBP) (ABC transporter)	-3	4,7	59
100362		55	Putative ABC transporter, periplasmic substrate-binding protein	-2,6	4,7	33

C. INFORMATION AND REGULATION PATHWAY

20255	<i>gltX</i>	7	Glutamyl tRNA synthetase	2,8	5,7	54
50202	<i>serS</i>	9	Seryl tRNA synthetase	4,3	5,4	47
50218	<i>valS</i>	24	Valyl tRNA synthetase	11	5	109
60080		42	Putative serine protease do-like	3,9	4,7	54
70394	<i>hisS</i>	6	Histidyl-tRNA synthetase	3,3	5,2	58
80152	<i>pheT</i>	23	Phenylalanyl-tRNA synthetase beta chain (Phenylalanine--tRNA ligase beta chain)	11	5,5	86
90183	<i>pnp</i>	3	Polyribonucleotide nucleotidyl-transferase (PNPase)	2,5	5,2	78
100069	<i>leuS</i>	24	Leucine tRNA synthetase	11	5,5	98
100707	<i>ctpA</i>	13	Carboxy-terminal-processing protease precursor (C-terminal-processing protease)	2,2	5	47
40116	<i>alaS</i>	32	Alanyl-tRNA synthetase	-5,8	4,3	95
50467	<i>gatA</i>	45, 52	Glutamyl-tRNA(Gln) amidotransferase subunit A (Glu-ADT subunit A)	-3	5,5	53
70270		51	Peptidase, family M16 protein	-2,8	5,3	48
70671	<i>dctP</i>	55	TRAP transporter, DctP-like periplasmic substrate-binding subunit precursor	-2,6	5,4	36
20357		23	Putative 5'nucleotidase protein	22	4,5	69
50303	<i>tufB</i>	12, 59	Elongation factor Tu	21	5,3	43

D. METABOLISM

Intermediary metabolism

20077	<i>doxF</i>	7	Salicyladehyde dehydrogenase	2,8	5,7	51
20087	<i>exoB</i>	18	UDP glucose 4-epimerase	2	5,4	36
20567	<i>pydA</i>	6	Dihydropyrimidine dehydrogenase	3,3	5,1	48
20763		15	Putative proline dipeptidase, PepQ	3,3	5,4	42
30098	<i>pfp</i>	14	Pyrophosphate fructose-6-Phosphate-1-phosphotransferase	9,2	5,7	44
30226		20	Putative glycerophosphoryl diester phosphodiesterase	4,8	5,5	26
70666	<i>uxuB</i>	8	UxuB/UxaB, Fructuronate reductase/Tagaturonate reductase (Mannonate oxidoreductase/Altronate oxidoreductase/Mannitol dehydrogenase)	5	5,5	51
80139		14	Peptidase M29, aminopeptidase II	9,2	5,7	45
80218	<i>upp</i>	21, 37	Uracil phosphoribosyltransferase	2,5	5,8	23
20231	<i>hipO</i>	60	Hippurate hydrolase	-2,3	5,8	41
20264	<i>xylA</i>	63	D-xylose isomerase	-2	5,3	49
20307		56	Putative glycerate dehydrogenase (gyaR-like)	-2,6	5,6	35
20397	<i>fdaA</i>	58	Putative fructose-bisphosphate aldolase class 1, Fda	-2,5	5,7	72
20400	<i>cbbG</i>	16	Glyceraldehyde 3-Phosphate dehydrogenase	-4,5	6,1	36
30075		56	Putative Iron containing alcohol dehydrogenase	-2,6	5,6	41

Energy metabolism

20122	<i>aspC</i>	12	Aspartate aminotransferase	39,6	5,4	44
30198	<i>paaB</i>	34	Enoyl-CoA hydratase/isomerase PhaB	2,9	5,8	28
50165	<i>eno</i>	13	Enolase	2,2	4,8	45

50515	<i>nuoE</i>	12	NADH quinone oxidoreductase subunit E	39,6	5,6	34
70358	<i>atpE</i>	37	ATP synthase subunit C, membrane-bound, F0 sector; DCCD-binding	2,3	4,9	76
80026	<i>frk</i>	15	Fructokinase	3,3	5,4	34
100264		9	Putative aminobenzoyl-glutamate utilization protein B, abgB	4,3	5,3	51
100284	<i>phbA</i>	14	Acetyl-CoA acetyltransferase with thiolase domain (Acetoacetyl-CoA thiolase)	3,3	5,8	41
100476	<i>etfB</i>	34	Electron transfer flavoprotein beta-subunit (Beta-ETF) (Electron transfer flavoprotein small subunit) (ETFSS)	2,9	7,9	27
100235	<i>nirK</i>	12	Nitrite reductase, copper-containing	39,6	5,4	41
100347	<i>acnA</i>	33	Aconitate hydratase	-3,1	5,2	95
100411	<i>atpA</i>	41	ATP synthase subunit alpha, membrane-bound, F1 sector	-2	5,9	55
100413	<i>atpD</i>	36	ATP synthase subunit beta, membrane-bound, F1 sector	-2,6		
100427		32, 33	Putative carbon monoxide dehydrogenase large subunit, coxL-like	-4,5	5,3	84
100428		40	CoxM-like protein, putative carbon monoxide dehydrogenase medium subunit	-2,3	5,8	28
100475	<i>etfA</i>	55	Electron transfer flavoprotein alpha-subunit (Alpha-ETF) (Electron transfer flavoprotein large subunit) (ETFLS)	-2,6	4,5	31
Aa biosynthesis						
20563		11	Putative N-carbamoyl L-cysteine amidohydrolase, putative AtcC	3,7	5	45
20804	<i>carA</i>	59	Carbamoyl phosphate synthetase glutamine amidotransferase small subunit	2,4	5,2	43
100029	<i>ahcY</i>	9	Adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase)	4,3	5,3	51
100088	<i>asd</i>	12	Aspartate semialdehyde dehydrogenase	39,6	5,5	38
100090	<i>leuB</i>	13	3 isopropylmalate dehydrogenase	2,2	4,9	40
100368	<i>argG</i>	12	Argininosuccinate synthase	39,6	5,6	45
100470	<i>lysA</i>	7	Diaminopimelate decarboxylase	2,8	5,7	46
100712	<i>proA</i>	14	Gamma-glutamyl phosphate reductase	9,2	5,9	45
20780	<i>aatA</i>	62	Aspartate aminotransferase	-2,1	5,7	43
50273	<i>ilvD</i>	38	Dihydroxyaciddehydratase DAD	-2,3	5,3	65
100039	<i>trpB</i>	62	Tryptophane synthase beta chain	-2,1	5,6	44
100055	<i>aroE</i>	57	Shikimate dehydrogenase	-2,5	5,7	30
Biosynthesis of cofactors, prosthetic groups, and carriers						
20734	<i>hemF</i>	19	Coproporphyrinogen III oxidase	2,9	5,4	35
50076	<i>nadE</i>	4	Glutamine-dependent NAD(+) synthetase (NAD(+) synthase [glutamine-hydrolyzing])	2,3	5	61
100430	<i>dxr</i>	14	1-deoxy-D-xylulose 5-phosphate reductoisomerase	3,3	5,2	42
30174	<i>pdxJ</i>	57	Pyridoxine 5'-Phosphate synthase	-2,5	5,3	27
50152	<i>ispDF</i>	60	IspD/IspF bifunctional enzyme [Includes: 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (4-diphosphocytidyl-2C- methyl-D-erythritol synthase) (MEP cytidyltransferase) (MCT); 2-C- methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECPS) (MECDP-synthase)]	-2,3	5,8	42
70067	<i>pdxA</i>	16	4-hydroxythreonine 4Phosphate dehydrogenase	-4,5	6,1	33
Fatty acids and phospholids metabolism						
40030	<i>fabF</i>	11	3-oxoacyl-[acyl-carrier-protein] synthase II (Beta-ketoacyl-ACP synthase II)	3,7	5	45
80030		37	Putative short-chain dehydrogenase/reductase SDR	2,3	5,6	25
80197		21	Putative 3-oxoacyl-(Acyl carrier protein) reductase	3	6,6	27
100146	<i>scoA</i>	20	Succinyl CoA 3ketoacidcoenzymeA transferase subunit A	4,8	5,4	25
50574	<i>fabH</i>	28	3-oxoacyl-[acyl-carrier-protein] synthase III (Beta-ketoacyl-ACP synthase III)(KAS III)	-3,2	5,4	34
50653	<i>fabF</i>	50	3-oxoacyl synthase II	-2,8	5,3	44
70296	<i>fabI</i>	58	Enoyl-[acyl-carrier-protein] reductase (NADH)	-2,5	5,6	29
90177	<i>fabA</i>	35	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (Beta-hydroxydecanoyl thioester dehydrase)	-2,7	5,6	19
90178	<i>fabB</i>	62	3-oxoacyl-[acyl-carrier-protein] synthase I (Beta- ketoacyl-ACP synthase I) (KAS I)	-2,1	5,4	43

100107		45, 52	Fatty acid CoA ligase	-3	5,4	55
E. TRANSCRIPTION						
20386	<i>suhB</i>	34	Inositol monophosphatase	2,9	6,1	27
50611	<i>rnd</i>	14	Ribonuclease D	9,2	5,8	42
F. UNKNOWN FUNCTION						
20090		12	Conserved exported protein of unknown function, saccharopine dehydrogenase family	39,6	5,4	45
20270		15	Conserved hypothetical protein putative xylose isomerase-like	3,3	5,3	39
20319		15	Conserved hypothetical protein, putative luciferase-like	3,3	5,4	37
30223		19	Conserved protein of unknown function	2,9	5,6	32
40078		19	Conserved hypothetical protein, putative signal peptide protein	2,9	5,7	34
50469		40	Conserved protein of unknown function, putative metal-dependent hydrolase	-2,3	5,3	25
90056		3	Conserved exported protein of unknown function, putative periplasmic binding protein	2,5	5,4	70
90145		54	Conserved hypothetical protein	-2,6	4,5	32
100075		20	Conserved hypothetical protein	4,8	5,2	26
20150		54	Conserved exported protein of unknown function	-2,6		
20394		57	Conserved hypothetical protein	-2,5	5,7	72
30009		58	Putative beta lactamase like protein	-2,5	5,3	28
70092		54	Conserved exported protein of unknown function	-2,6	4,8	35
90217		56	Conserved hypothetical protein, putative methionine synthase	-2,6	5,5	41

Conclusion générale et perspectives

Les micro-organismes sont capables de s'adapter à une large gamme de conditions environnementales. De nombreuses bactéries peuvent notamment coloniser des environnements contaminés par l'arsenic, où elles participent activement au cycle biogéochimique de ce métalloïde. Les formes inorganiques : arsénite, As(III), et arséniate, As(V), sont les plus abondantes mais aussi les plus toxiques. De nombreuses études ont porté sur les mécanismes de réduction de l'arséniate, aussi bien au niveau enzymatique que génétique. En revanche, peu de travaux ont été consacrés à l'oxydation de l'As(III) et aux mécanismes impliqués dans la réponse au stress arsénié. Ce travail de thèse a porté sur les mécanismes moléculaires du stress arsénié chez *Herminiimonas arsenicoxydans*.

D'une manière générale, la disponibilité du génome complet de cet organisme a facilité les analyses du protéome, la mise au point d'amorces ciblant certaines catégories de gènes (q-RT-PCR) ou le dessin d'oligonucléotides couvrant l'ensemble du génome (transcriptome). Aussi, le séquençage d'un génome permet d'avoir accès au répertoire de gènes d'un organisme, d'émettre des hypothèses quant à ses capacités métaboliques, mais il constitue surtout le point de départ des approches de génomiques fonctionnelles qui visent à comprendre l'utilisation de ce répertoire.

A l'issue de cette thèse, les travaux réalisés ont permis de caractériser au moins partiellement la réponse bactérienne au stress arsénié et plus particulièrement la régulation de l'arsénite oxydase. En effet, les travaux ont débuté avec l'étude des mécanismes moléculaires impliqués dans la régulation de l'activité arsénite oxydase (Chapitre 1). Cette approche a permis de caractériser plusieurs régulateurs de l'arsénite oxydase en réponse au stress arsénié. Toutefois, d'autres gènes peuvent être soumis à régulation après exposition à ce métalloïde. Aussi, dans un second temps nous sommes intéressés à la réponse globale des bactéries face à un stress arsénié (Chapitre 2). Des intervenants impliqués dans des mécanismes variés ont pu être caractérisés dans cette partie, ce qui nous a permis de montrer que la réponse ne se limite pas à la seule activité arsénite oxydase, mais est pléiotrope et complexe chez cet organisme. Cette étude illustre la diversité des mécanismes de réponse aux stress décrits dans la partie introductive, en particulier la diversité d'adaptation d'un micro-organisme face à un même stress. Toutefois aucune protéine ou gène codant l'arsénite oxydase n'ont été retrouvés lors de ces travaux (Chapitre 2), alors que dans le chapitre 1 nous avons montré qu'elle est régulée en réponse à l'As(III). Ces résultats montrent bien que l'étude d'un stress ne peut se limiter à la

caractérisation de la réponse à un seul instant *t*. Ainsi, le chapitre 3 a eu pour objectif d'étudier la cinétique d'expression des gènes en réponse au stress arsénié. Cette analyse a permis de déterminer que la réponse au stress arsénié implique deux phases clefs chez *H. arsenicoxydans* : (i) Une réponse précoce (15 min) qui est caractérisée par l'induction de gènes impliqués dans le stress oxydatif, le métabolisme du soufre, les protéines chaperons, et (ii) une réponse tardive (8 h) qui est caractérisée par l'induction de gènes impliqués dans la motilité, le transport du phosphate et le métabolisme de l'arsenic. La phase précoce a été qualifiée de processus général d'adaptation et la phase tardive de processus spécifique. Enfin, ces données couplées à des analyses physiologiques ont permis de proposer un lien entre l'arsenic et la motilité.

Les travaux effectués lors de ma thèse ont porté majoritairement sur *H. arsenicoxydans*. Toutefois, les processus d'oxydation de l'arsénite concernent d'autres micro-organismes. Aussi, l'analyse de ces mécanismes a été étendue à des bactéries chimiolithoautotrophes : *Thiomonas* sp. et *Rhizobium* NT-26.

Les souches *Thiomonas* sp. et l'organisation de leurs opérons *aox* ont été décrites dans le chapitre 4. Ces bactéries présentent des capacités métaboliques intéressantes telles que l'utilisation de sources de carbone variées et la production d'énergie, mais aucun outil génétique n'a pu être mis au point chez ces organismes. Aussi, l'étude du métabolisme arsénié a été menée chez une bactérie pour laquelle des outils génétiques sont disponibles : *Rhizobium* sp. NT-26. Des études récentes ont montré que cette bactérie peut utiliser l'oxydation de l'As(III) pour son métabolisme énergétique. Dans un premier temps, son génome a été séquencé et annoté de manière experte. Ensuite, dans la perspective d'étudier la réponse au stress arsénié chez des bactéries ayant ce métabolisme particulier, les travaux se sont organisés autour de deux axes menés en parallèle : une approche de mutagénèse aléatoire et une approche de génomique descriptive et fonctionnelle. L'analyse du génome et des données protéomiques semble indiquer que certains mécanismes tels que la résistance à l'arsenic *via ars*, la motilité et la formation de biofilm sont communs à *H. arsenicoxydans*, et que l'arsénite oxydase serait aussi soumise à régulation. L'étude des mécanismes moléculaires impliqués dans la régulation de l'arsénite oxydase a donc également été effectuée chez ce micro-organisme.

Ainsi, la caractérisation des souches de *Thiomonas* et de *Rhizobium* sp. NT-26 complète le repertoire des souches capables d'oxyder l'arsénite déjà décrites dans la littérature. Toutefois, d'autres mécanismes tels que le métabolisme du carbone, le

métabolisme énergétique sont spécifiques à ces micro-organismes et la nodulation, le système de sécrétion de type IV sont spécifiques à *Rhizobium* NT-26. Par ailleurs, la présence d'un gène *moeA* dans l'environnement génétique de l'opéron *aox* chez *Rhizobium* NT-26 et *Thiomonas* sp. semble être spécifique aux bactéries chimiolithoautotrophes. Enfin, les protéines AoxR, mais aussi MoeB et PtxB ont été montrées comme étant impliquées dans la régulation de l'activité arsénite oxydase chez *Rhizobium* sp. NT-26.

Ce travail de thèse a traité de différents aspects qui étendent les connaissances que l'on avait jusqu'ici de la réponse au stress arsénié chez les micro-organismes. Les résultats obtenus permettent de suggérer un certain nombre de directions qui pourront être développées dans le cadre de cette thématique dans le futur.

A courts et moyens termes, il s'agira de compléter l'étude menée chez *Rhizobium* sp. NT-26, par les données de transcriptome et de mutagenèse aléatoire. L'identification des mutants déficients dans l'activité arsénite oxydase a été effectuée par un criblage phénotypique. C'est pourquoi, afin de confirmer ou non l'effet des mutations sur l'activité arsénite oxydase, la spéciation de l'As devra être déterminée par HPLC couplée à l'ICP-AES pour les différents mutants. Les résultats devront ensuite être validés par une approche de complémentation des gènes d'intérêts. Cette dernière permettra de déterminer si la perte de l'oxydation est bien due au gène caractérisé ou à un effet polaire d'un gène en aval.

Il s'agira également de préciser les rôles de DnaJ, MoeB et PtxB dans la régulation de l'activité arsénite oxydase chez *Herminiimonas arsenicoxydans* et/ou *Rhizobium* sp. NT-26. D'une manière générale, l'ensemble des régulateurs n'a probablement pas été intégralement caractérisé par le criblage de la banque. En effet, le clonage des gènes *aoxABCD* de *H. arsenicoxydans* chez *E. coli* ne permet pas d'établir une activité arsénite oxydase malgré l'expression des gènes. Il reste donc encore à caractériser des partenaires nécessaires à cette activité. Une approche biochimique de purification par affinité de type Tap-Tag devrait permettre d'identifier les partenaires de l'arsénite oxydase.

A plus long terme, une approche basée sur la spectrométrie de masse de type NIMS (Nanostructure-initiator Mass Spectrometry) devrait permettre de caractériser les protéines interagissant avec l'arsenic. Enfin, il sera également intéressant d'étendre l'étude du métabolisme arsénié à des bactéries anaérobies telle que la souche *Desulfotomaculum hydrothermale* qui a également été séquencée par le Genoscope dans le cadre d'une collaboration avec l'équipe. En effet, cette souche sulfato-réductrice, à l'inverse de celles

étudiées jusqu'ici, est capable d'utiliser l'As(V) comme accepteur terminal d'électrons en anaérobiose en présence de diverses sources de carbone.

Ainsi, à terme ces travaux, reposant sur des approches variées, devraient conduire à la caractérisation moléculaire des mécanismes de régulation de l'expression des gènes impliqués dans l'oxydation de l'arsénite, mais devrait aussi permettre d'appréhender l'implication de l'arsenic dans le métabolisme énergétique, les processus de résistance et de détoxification.

Le site de Carnoulès est une mine de plomb, de zinc et d'argent au pied du Massif Central dans le Gard qui a été exploitée jusqu'en 1963. Les déchets des minéraux extraits (stériles) sont aujourd'hui stockés sur le site et contiennent de fortes quantités de sulfure de fer, de plomb et d'arsenic (Morin *et al.*, 2003). Les concentrations en arsenic sur ce site sont parmi les plus importantes au monde, jusqu'à 12 000 mg/L (Koffi *et al.*, 2003). L'action combinée de l'air et des eaux de pluie influence l'oxydation des affleurements sulfurés, ce qui conduit à une production acide et une dissolution des métaux présents. Les eaux acides de percolation se jettent ensuite dans le Reigous : une source d'eau qui prend naissance sous les stériles. Ce ruisseau est chargé de divers métaux en particulier une très forte concentration en arsenic dissous de 100 à 350 mg/L, en fonction de la saison. On observe une atténuation biologique naturelle le long du Reigous. Les micro-organismes présents participent activement à cette bioremédiation, soit en participant à l'oxydation de l'arsénite en arséniate, soit en participant à la co-précipitation de l'arséniate avec les ions ferriques Fe(III) (Koffi *et al.*, 2003 ; Duquesne *et al.*, 2008). En effet, à pH acide, l'As(V) s'absorbe facilement aux ions Fe(III), ce qui influence fortement l'immobilisation de l'élément. Le site semble présenter une faible diversité bactérienne (due au faible pH, à la toxicité des métaux dissous et au nombre limité d'accepteurs et de donneurs d'électrons). La population dominante est composée de *Gallionella ferruginea* qui est capable d'oxyder le fer, d'*Actinobacteria* Fe-oxydante et d' *Acidithiobacillus ferrooxidans* Fe-oxydante et sulfato-réductrice. On retrouve également une autre bactérie sulfato-réductrice, *Desulfobacterium*, une bactérie arsénite-oxydante, *Thiomonas*, des archées et des eucaryotes comme *Euglena mutabilis* (Bruneel *et al.*, 2008).

L'exploration de ce site a débuté par l'analyse des bactéries du genre *Thiomonas* (Chapitre 4 et Annexe 1). Toutefois, dans l'environnement, les micro-organismes se retrouvent dans des communautés parfois très complexes, dans lesquelles des espèces très éloignées interagissent. Cette étude ne peut donc se résumer à la simple étude de ces souches, même si leur métabolisme semble jouer un rôle important. L'étude des communautés microbiennes par une approche métagénomique, présente une réelle avancée pour comprendre la complexité de ce type d'environnement et les relations/interactions bactériennes au sein de cette niche écologique (Tyson *et al.*, 2004). La métagénomique est une analyse génomique culture indépendante d'une communauté microbienne (Riesenfeld *et al.*, 2004). Cette approche a pour objectif une analyse phylogénétique ou de déterminer les potentialités génétiques des populations microbiennes d'un écosystème donné (Riesenfeld *et al.*, 2004 ; Handelsman, 2004). Comme exemple

d'étude de la potentialité, on peut citer le séquençage à grande échelle mené sur les micro-organismes de différents océans (Rusch *et al.*, 2007), ou sur la mer Sargasse (Tyson *et al.*, 2004). Néanmoins, la reconstruction des pseudo-génomés n'est pas très aisée, même avec des outils informatiques puissants et des données de séquences importantes. Ainsi, les analyses génomiques préalables de souches isolées du milieu d'intérêt (notamment *Thiomonas*) ont aidé à la reconstruction des pseudo-génomés.

Sur la base des fonctions putatives assignées aux gènes, il a été possible de reconstruire *in silico* les capacités métaboliques de la communauté. Il existe des bases de données spécialisées dans cette démarche comme KEGG (Kyoto Encyclopedia Of Genes and Genomes, <http://www.genome.ad.jp/kegg/>), ERGO (<http://ergo.integratedgenomics.com/ERGO/>) ou MetaCyc (<http://metacyc.org/>). Celles-ci ont ensuite été (in-)validées par une analyse métabolomique.

L'étude environnementale de Carnoules a permis de mettre en évidence les interactions trophiques des bactéries présentes dans cet environnement extrême et de souligner le rôle clef de la souche arsénite-oxydante *Thiomonas* (Annexe 2).

Des bactéries modifient leur génome pour survivre à l'arsenic

[Fermer la fenêtre](#)

Institut des sciences biologiques

Parutions

Des bactéries modifient leur génome pour survivre à l'arsenic

En remaniant leur génome, les bactéries du genre *Thiomonas* seraient capables de survivre dans des environnements extrêmement contaminés ! C'est ce que les chercheurs du laboratoire de Génétique Moléculaire, Génomique et Microbiologie (UMR7156 CNRS, Strasbourg) ont conclu de leurs travaux publiés dans la revue *PLoS Genetics* le 26 février 2010.

Au sein de l'Université de Strasbourg et plus précisément du laboratoire Génétique Moléculaire, Génomique et Microbiologie (GMGM), les travaux de l'équipe « Ecophysiologie Moléculaire des Procaryotes », dirigée par P. Bertin, concernent principalement les métabolismes microbiens impliqués dans la résistance et la détoxification d'éléments comme l'arsenic.

Dans le cadre du réseau de recherche CNRS « Métabolisme de l'Arsenic chez les Microorganismes » (GDR2909), des approches de génomique descriptive et fonctionnelle, combinée à des méthodes génétiques et biochimiques, sont mises en oeuvre pour décrire la physiologie de microorganismes isolés d'environnements naturels ou anthropisés. Des outils de génomique environnementale ont été récemment développés afin de caractériser des communautés complexes et identifier, y compris chez des micro-organismes non cultivés, de nouvelles fonctions potentiellement impliquées dans la remédiation de sites contaminés.

Alors que les souches d'*Hermiimonas* précédemment étudiées semblent inféodées à quelques niches écologiques, les bactéries du genre *Thiomonas* sont présentes dans la plupart des environnements contaminés par l'arsenic. Afin de comprendre ce qui leur permet de survivre dans ces milieux, le génome de l'une d'entre elles a été analysé en détail. Il a ensuite été comparé à d'autres souches de *Thiomonas*, originaires du même site ou provenant d'environnements différents et appartenant ou non à la même espèce selon les critères classiques de phylogénie.

Les travaux ont pu montrer que ces bactéries possèdent un pool de gènes variables leur permettant de survivre et de coloniser efficacement un environnement toxique. Les chercheurs expliquent cette acquisition de résistance en partie par la présence de phages (virus de bactéries) dans les milieux étudiés. En effet, par leur rôle dans le transfert de gènes de résistance et dans les événements de réarrangements du génome, les phages sont des acteurs probables de l'adaptation de la souche à son environnement.

L'analyse approfondie des données génomiques suggère également que l'environnement spécifique de ces microorganismes influence l'évolution de leur génome.

Tous ces phénomènes conduisent à une importante variabilité génétique, y compris chez des bactéries très proches phylogénétiquement, et élargissent ainsi la vision relativement restreinte de l'espèce bactérienne.

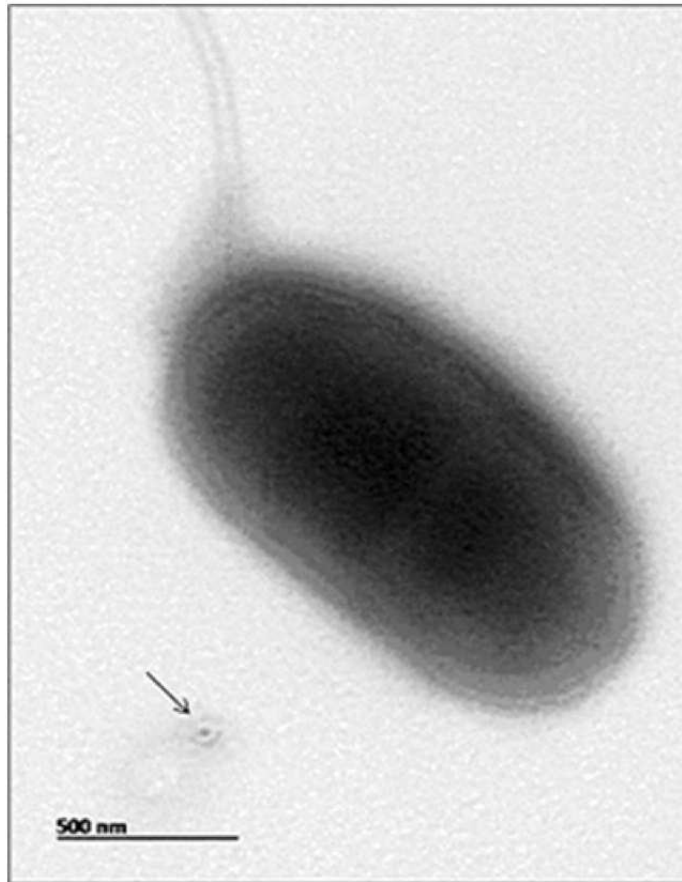


Figure : Cellule de *Thiomonas* et son phage (flèche) en microscopie électronique à transmission. Le phage serait impliqué dans le transfert de gènes de résistance et dans les événements de réarrangement du génome permettant à la souche de s'adapter à un environnement toxique particulier.

🔍 En savoir plus

- *Structure, function and evolution of the Thiomonas spp. genome*
Arsène-Ploetze F, Koechler S, Marchal M, Coppée J-Y, Chandler M, Bonnefoy V, Brochier-Armanet C, Barakat M, Barbe V, Battaglia-Brunet F, Bruneel O, Bryan C, Cleiss-Arnold J, Cruveiller S, Erhardt M, Heinrich-Salmeron A, Hommais F, Jouliau C, Krin E, Lieutaud A, Lièvreumont D, Michel C, Muller D, Ortet P, Proux C, Siguier P, Roche D, Rouy Z, Salvignol G, Slyemi D, Talla E, Weiss S, Weissenbach J, Médigue C, Bertin P
Public Library of Science (PLoS) Genetics
Published online February 26, 2010, doi:10.1371/journal.pgen.1000859

✉ Contact chercheur

- **Philippe Bertin**
Génétique Moléculaire, Génomique et Microbiologie (GMGM)
UMR7156 CNRS/Université Strasbourg 1
Institut de Botanique
28 Rue Goethe
67083 Strasbourg Cedex
Métabolisme de l'Arsenic chez les Microorganismes
GDR2909 CNRS - <http://gdr2909.u-strasbg.fr/gdr2909/>

Diversity of trophic interactions inside an arsenic-rich microbial ecosystem

Diversity of trophic interactions inside an arsenic-rich microbial ecosystem

Philippe N. Bertin^{a1}, Audrey Heinrich-Salmeron^a, Eric Pelletier^{b,c,d}, Florence Goulhen-Chollet^a, Florence Arsène-Ploetze^a, Sébastien Gallien^f, Alexandra Calteau^{c,d,e}, David Vallenet^{c,d,5}, Corinne Casiot^g, Béatrice Chane-Woon-Ming^{c,d,e}, Ludovic Giloteaux^h, Mohamed Barakatⁱ, Violaine Bonnefoy^j, Odile Bruneel^g, Michael Chandler^k, Jessica Cleiss-Arnold^a, Robert Duran^h, Françoise Elbaz-Poulichet^g, Nuria Fonknechten^{b,c,d}, David Halter^a, Béatrice Lauga^h, Marie Marchal^a, Damien Mornico^{c,d,e}, Philippe Ortetⁱ, Eduardo P. C. Rocha^l, Christine Schaeffer^f, Patricia Siguier^k, Emmanuel Talla^j, Adam Alexander Thil Smith^{c,d,e}, Sara Vieira-Silva^l, Alain Van Dorsselaer^f, Jean Weissenbach^{b,c,d}, Claudine Médigue^{c,d,e}, Denis Le Paslier^{b,c,d}.

^aGénétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS & UdS, Strasbourg, France; ^bCEA, DSV, IG, Genoscope, Laboratoire de Métagénomique des Procaryotes, France; ^cCNRS UMR8030, Evry, France; ^dUEVE, Université d'Evry, France; ^eCEA, DSV, IG, Genoscope, Laboratoire de Génomique Comparative, France; ^fLaboratoire de Spectrométrie de Masse Bio-Organique, Institut Pluridisciplinaire Hubert Curien, UMR7178 CNRS & UdS, Strasbourg, France; ^gLaboratoire Hydrosociences Montpellier, UMR 5569, Montpellier, France; ^hEnvironnement et Microbiologie, UMR5254 CNRS & UPPA, Institut Pluridisciplinaire de Recherche sur l'Environnement et les Matériaux, Pau, France; ⁱLaboratoire d'Ecologie Microbienne de la Rhizosphère et d'Environnements Extrêmes, UMR6191 CNRS, CEA & Université Aix-Marseille II, Saint-Paul-lez-Durance, France; ^jLaboratoire de Chimie Bactérienne, UPR9043 CNRS et Université de la Méditerranée, Marseille, France; ^kLaboratoire de Microbiologie et Génétique Moléculaires, UMR5100 CNRS, Toulouse, France; ^lInstitut Pasteur, Microbial Evolutionary Genomics, CNRS, URA2171, Paris, France

¹To whom correspondence should be addressed. E-mail : philippe.bertin@unistra.fr
Phone: +33 (0)3 68 85 18 18; Fax: +33 (0)3 68 85 20 28

Key words: microbial genomics / extreme environments / uncultured microorganisms / bacterial phylum / syntrophy

Classification: Biological Sciences, Environmental Sciences and Ecology

Microorganisms play a crucial role in the biogeochemical cycles of elements, including toxic compounds originating from human activities. The complete understanding of those metabolic processes requires however the deciphering of both the structure and the function of microbial communities. Until recently, such an endeavour has been largely hampered by a lack of genomic data on uncultivated microorganisms. Using metagenomic and metaproteomic approaches, we demonstrated here that an environment highly contaminated with arsenic and toxic metals is dominated by seven microbial strains whose genomes were reconstructed. Most of them represent yet uncultivated bacteria, including microorganisms belonging to a novel bacterial phylum. In depth analysis of the (proteo)genomics data unravelled some unsuspected capabilities of those microorganisms, in particular regarding not only sulphur, iron or arsenic oxidation, but also organic compound or cofactor metabolism. From these results, we built an integrated model of the various interactions that may promote microbial growth inside the arsenic-rich ecosystem. These metabolisms, including those of the uncultivated *Candidatus Fodinabacter communificans*, may be of prime importance in the efficient biomineralization of such anthropized ecosystems.

Introduction

Microorganisms are involved in biogeochemical nutrient cycles and play therefore a crucial role in the biosphere (1, 2). Although microbial ecosystems constitute a major reservoir of our planet's genetic biodiversity, it is generally recognized that most of the microorganisms present in the environment are not accessible by culture-dependent techniques. However, the recent development of genomics has given an unprecedented opportunity to gain insight into the structure and the function of microbial communities (3-5). Indeed, environmental genomics has extended their analysis far beyond the sole taxonomic studies, giving rise to an integrated picture of ecosystems, including those considered as extreme. Such a comprehensive view may be relevant to understand the impact of human activities on biosphere. Among the most toxic anthropogenic contaminants, arsenic, in the As(III) form (arsenite), may have been an important mineral substrate providing chemolithotrophic organisms with energy in the environments harbouring the first forms of life on earth (6). Nowadays, high arsenic contents are encountered in mine drainage waters, where the metalloid is usually associated with sulphur, iron and other metals (7). Few forms of life are known to thrive in these acid waters, as compared with neutral waters (8).

The former mine of Carnoulès, Gard (France), provides an outstanding example of such an extreme environment. The sulphurous wastes contain As-rich pyrite and the leached waters are the source of a small stream called the Reigous that contains between 50 and 350 mg.l⁻¹ of soluble arsenic, mainly in the form of As(III) (Table S1). However, although the arsenic levels remain still high, this concentration decreases by 95% between the source of the Reigous and its confluence with the river Amous, 1.5 km downstream. This natural process of remediation is mainly due to microbial metabolism, leading to the oxidation of iron and arsenic into Fe(III) and As(V), and their co-precipitation of with sulphur (9-11).

Despite many recent works, few of these microbes have been isolated (11-13). In the present study, we used a multidisciplinary approach to get an integrated view of the microorganisms present and their role in the complex metabolic processes at work.

Results and Discussion

Genome Reconstruction and Phylogeny. DNA recovered from the upper zone of sediments was sequenced and assembled in 7 major bins, using a combination of 16S rRNA gene sequences, GC%, mean coverage of the various contigs, mean polymorphism frequency and similarity to already sequenced genomes, including *Acidithiobacillus ferrooxidans* and *Thiomonas* sp. 3As (<https://www.genoscope.cns.fr/agc/mage/arsenoscope>). Even though the assembling of sequences from closely related organisms cannot be excluded, each supercontig group was considered to correspond to the genome of one single major organism (Table S2, and Figs. S1, S2). We evaluated the origin of bins by a phylogenetic analysis of 16S rRNA sequences obtained from both a clone library and metagenome sequencing (Fig. 1). We also made a phylogenomic analysis of the 7 bins using 27 universal marker genes (individually or combined) supposed to be sufficiently conserved to build a tree of life (14). Sequence analyses revealed a dominance of bacteria belonging to the Proteobacteria phylum (β - and γ -Proteobacteria with 38 and 23%, respectively). Remaining clones belong to Firmicutes (4%) and Acidobacteria (3%), and to a lesser extent to Spirochaetes (1%) and α -Proteobacteria (1%). This suggests that the microbial community at Carnoulès includes a few other bacterial phyla in addition to those represented by the 7 major genomes, and few Archaea (9). More precisely, and except for CARN4 whose missing rRNA sequence precluded its 16S-based classification, and CARN6 where no marker was found, the candidates CARN2, CARN3 and CARN5 were affiliated to the β -Proteobacterium *Thiomonas* genus, the Acidobacteria clade and to the γ -Proteobacterium *At.*

ferrooxidans, respectively (Table S3). In addition, the comparison of CARN7 16S rRNA gene with sequences in the RDP database suggests that this organism is phylogenetically related to the *Gallionella* β -Proteobacterium. All phylogenetic results for the bins were supported, at least in part, by an analysis of gene order conservation known to be correlated with evolutive distances (15). Remarkably, the candidates corresponding to the CARN1 and CARN4 bins showed an important phylogenetic relationship and may represent 2 subpopulations (Figs. S1, S2), as recently observed in other contexts (5). Those bins did not correspond to any known taxonomic class and, according to the metabolic properties identified in the present study (see below), this new genus was herein named *Candidatus Fodinabacter communicans* (from fodina, mine and communicare, share), in accordance with the recommendations for incompletely characterized microorganisms (16). Interestingly, similar 16S rRNA sequences present in databases have already been identified from acid mine environments, suggesting that these uncultivated bacteria are widespread in such ecosystems.

The size of the 7 bins ranged from 1.5 Mbp to more than 4.0 Mbp and their GC content ranged approximately from 52 to 65% (Table S4). DNA sequencing led to a low coverage for some contigs, in particular those of CARN3, which suggests that the size of the corresponding genome is underestimated. In addition, the comparison of the CARN5 data with those of the 2 already available *At. ferrooxidans* genomes revealed the existence of at least one large gap in CARN5 and ATCC 53993, as compared to ATCC 23270. Similar results were obtained by 454 pyrosequencing, which supports the lack of any major bias resulting from the cloning procedure for genomic library construction (Fig. S3). In ATCC 23270, this region contains a duplication of most tRNA encoding genes, several genes involved in the synthesis of membrane sugars and multiple transposases suggesting a probable acquisition by horizontal gene transfer.

Arsenic Metabolism, Stress Response and Regulation. We investigated the fitness of the bacterial community, in particular regarding arsenic, by exploring *in silico* the gene content of each bin. In parallel, we analyzed the proteome synthesized *in situ*, which allowed the reliable identification of more than 500 unique proteins (Table S5AB). Arsenic toxicity results from various biological effects, including oxidative stress and DNA damage. As expected, both metagenomic and/or metaproteomic data showed that all Carnoulès strains can express functions protecting against oxidative stress, including those coding for catalase, superoxide dismutase, glutathion peroxidase or thioredoxin/thioredoxin reductase. Most bins also contain genes coding for the bacterioferritin comigratory protein, which protects cells against toxic hydroxyl radicals. Multiple genes coding for enzymes involved in DNA recombination and repair were also identified and were recently shown to be important for viability in the presence of arsenic (17). In addition, the genome of each bin carries at least one *ars* operon conferring resistance to arsenic by the means of an ArsB efflux pump (Fig. 2). Surprisingly, the 3 operons identified in *Thiomonas* sp. CARN2 were located in the vicinity of 3 *aoxAB* operons, suggesting the existence of 3 distinct arsenic islands in this strain (Fig. S5). One of those islands was flanked by multiple transposases (Table S6), suggesting a possible acquisition by horizontal gene transfer. Arsenic methylation may also be considered as a resistance mechanism (18), and *arsM* genes coding for an arsenite S-adenosylmethyltransferase were identified, in particular in the CARN6 bin. This observation is in agreement with the presence of methylated forms of arsenic in Carnoulès sediments, i.e. monomethylarsonate and dimethylarsinate (7.10^{-4} mg MMA and 3.10^{-4} mg DMA, respectively, per mg dry weight). Finally, because of the structural similarity between phosphate and As(V), arsenic metabolizing strains may preferentially transport phosphate via the specific Pst phosphate transport system rather than the Pit general transport

mechanism to reduce the entry of As(V) (17, 19). No Pit gene was identified in any bin while all strains contain a Pst system (Table S5B), further supporting this hypothesis.

Due to their role in biofilm formation, motility and aggregation have been often demonstrated to be major determinants for survival in toxic environments, including arsenic (17, 19). Except for *Acidithiobacillus* sp. CARN5, most bins contain multiple genes required for flagellum biosynthesis and motility (Fig. 2). Moreover, *Thiomonas* sp., *Acidithiobacillus* sp. and *Gallionella*-related strains also carry Type 4 pilus operons, which play an important role in twitching motility and adhesion. Finally, *Acidithiobacillus* sp. CARN5 contains several genes involved in capsular biosynthesis, e.g. *kdtA*, *crd* and *gmd*, suggesting the possible formation of a biofilm. Such adaptive traits typically depend on multiple regulatory mechanisms that allow bacteria to respond to a wide panel of stimuli, including high concentrations of iron and arsenic. In this respect, all bins encode nucleoid-associated proteins such as HU, IHF, FIS or H-NS (20), which play a major role in both the structure and the function of chromosomal DNA. Multiple regulators, e.g. various alternative sigma factors and regulatory proteins belonging to the two-component systems or those involved in cell-to-cell signaling, were also identified in all strains (Table S7).

Energy Metabolism. A wide diversity of bioenergetic electron chains may be needed to accommodate the fluctuations in the various physico-chemical parameters. Accordingly, several terminal oxidases might be operative in all strains of the Carnoulès microorganisms, e.g. the cytochrome oxidase *cox* and *cta* operons and the *cyo* and *cyd* operons encoding quinol oxidases. In addition, several operons involved in anaerobic respiration were present in the bin of *Thiomonas* sp. (CARN2), in particular those involved in the metabolism of nitrate and nitrite, such as *ntr*, *nar*, and *nas*, or the use of fumarate as an electron acceptor under anaerobic conditions (Fig. 2). More importantly, possible inorganic electron donors

may comprise sulphur compounds and arsenite. This was attested by the synthesis by *Thiomonas* sp. of proteins such as *sor*, *sox*, *tetH* and *sqr* as well as *aox* involved in the oxidation of reduced inorganic sulphur compounds, e.g. sulfide, sulphur, sulphite, thiosulphate and tetrathionate, and As(III), respectively. Similarly, the CARN7 bin harbours a *dsr*ABEFHCMKLJOPN operon. The DsrA protein showed 72% similarity with ortholog in the sulphur-oxidising *Thiobacillus denitrificans*. This suggests that the *Gallionella*-related strain is also able to use sulphur compounds as electron donor in its energy metabolism, as suggested in *Gallionella ferruginea* (21). In addition, the *Acidithiobacillus* sp. CARN5 bin expressed the sole RusA protein (Table S5B). This rusticyanin is involved in electron transport with iron used as an energy source, which suggests a major role of CARN5 in the iron oxidation observed on the study site (11, 22). All these metabolisms may play a key role in the natural bioremediation observed at Carnoulès. Finally, the oxidation of H₂ to protons may also be a source of energy due to the presence in CARN2, CARN5 and CARN6 of several hydrogenase-encoding genes. In contrast, no respiratory arsenate reductase, which allows anaerobic respiration of As(V), was identified in any of the 7 bins.

Trophic Interactions. The existence of metabolic exchanges regarding carbon and nitrogen may be of prime importance inside a microbial community. As expected from a partly oligotrophic environment and suggested by the predicted slow growth rate of each bin (Table S8), we identified several enzymes responsible for autotrophic metabolism (Fig. 2). Carbon fixation may depend on genes coding for ribulose 1,5-biphosphate carboxylase/oxygenase (involved in Calvin cycle), carboxysome structural proteins and carbon monoxide dehydrogenase (involved in acetyl-coenzyme A synthesis) that were identified in and expressed by *Thiomonas* sp. (CARN2), *Acidithiobacillus* sp. (CARN5) or *Gallionella*-related (CARN7) strains. On the other hand, CARN5 and CARN7 were also

able to fix nitrogen by the means of *nif* genes, which encode a nitrogenase. Nevertheless, mixotrophic or heterotrophic metabolism may also exist in Carnoulès, suggesting that microorganisms recycle some organic compounds released by others. For example, CARN2, CARN5 and CARN7 encode lactate, formate or succinate dehydrogenases, while *fucP* and *exuT* genes coding for L-fucose and hexuronate transporters, respectively are present in CARN3. In addition, the CARN6 bin carries the cellulase encoding gene *bczS* and the alpha-amylase *amyM* gene, suggesting that this strain is able to metabolize more complex carbohydrates. Remarkably, the CARN1 and CARN4 bins of *Candidatus* F. communificans carry the arginase-encoding gene *rocF* involved in urea biosynthesis, while a complete urease-encoding *ure* operon was present in *Thiomonas* sp. CARN2 strain. This observation is in agreement with the identification of urease activity in various *Thiomonas* strains (23) and further supports a metabolic interaction between the Carnoulès microorganisms. Similarly, several enzymes required for amino acid transport and metabolism were identified in the bins lacking the carbon and nitrogen fixation genetic determinants. In particular, the CARN1 and CARN4 bins of *Candidatus* F. communificans, which were shown to express most of the identified proteins (Table S5B), encode multiple peptidases and also contain *liv* and *opp* genes involved in branched amino-acid and oligopeptide transport, respectively. Genes involved in fermentation were also identified in both bins, e.g. those of the lysine pathway which converts lysine into butyrate, acetate and ammonia, principally in obligate anaerobic bacteria (24). Finally, the CARN1 and CARN4 bins carry genes involved in cobalamin (vitamin B12) synthesis (25) while cobalamin transporters are present in the genome of CARN2 and CARN5, as observed in already sequenced related bacteria (<https://www.genoscope.cns.fr/agc/mage/arsenoscope>). This biosynthesis may also be useful to eucaryotes, in particular *Euglena* sp. present on the study site (26) whose cell cycle is known to require cobalamin (27, 28). Remarkably, iron oxidation by *At. ferrooxidans* was

shown to be strongly increased in the presence of cobalamin (Fig. S6), which suggests that such a metabolic cooperation may be of importance in the natural remediation observed at Carnoulès.

These observations were further supported by a Multiple Factor Analysis (MFA) of both metagenomic and metaproteomic data on a reaction frequency matrix, suggesting an important and specific role for each organism inside the Carnoulès community. Indeed, factorial planes segregated CARN2 from the other bins (Fig. 3A) and separated CARN1-CARN4 from CARN5-CARN7. From the variable classification results, several clusters of reactions were then associated to bin groups. These include, for example, Cluster 4 linked to CARN2 (energy metabolism, inorganic nutrient metabolism and arsenic detoxification) (Table S9A); Cluster 1 groups reactions common to CARN2 and CARN5 (Calvin-Benson-Bassham cycle and urea degradation pathways); Cluster 5 includes reactions related to lysine fermentation and other amino acid degradation pathways in CARN1 and CARN4, and also gathers reactions related to cobalamin biosynthesis. In addition, CARN3 was opposed to CARN6 (Fig. 3B), revealing CARN6-specific reactions (Clusters 5 and 6) that include those involved in purine degradation and cellulose metabolism (Table S9B). Interestingly, no such correlation was observed between our data and those from the acidic AMD biofilm (29), which suggests that the diversity of these two ecosystems differ markedly and highlights the metabolic complementarity inside the Carnoulès community (Fig. S7).

Conclusion

Taken together, these environmental genomics approaches allowed us to inventory the multiple metabolic pathways in the ecosystem under study and to visualize several activities expressed *in situ* (Fig. 2 and Table S5B). More importantly, our observations revealed the diversity of trophic interactions that may exist inside a microbial community and allow

microorganisms, including yet uncultivated bacteria, to efficiently colonize an arsenic-rich environment. These processes include not only the fixation of inorganic carbon and nitrogen but also the use or the recycling of other mineral and organic resources such as arsenic, iron, sulphur, urea, vitamins and amino acids. Such an equilibrium between organic and inorganic compound metabolism may provide all partners with nutrients. In this respect, the role of *Thiomonas* sp. in arsenic oxidation associated with other metabolic activities, including sulphur oxidation by the *Gallionella*-related strain and iron oxidation by *Acidithiobacillus* sp., seems to be of prime importance in the partial but efficient natural remediation of the Carnoulès contaminated ecosystem. Other microorganisms, in particular *Candidatus* F. communificans, may recycle organic compounds released by others or provide them with essential cofactors. Our observations therefore support an indirect but crucial role of this first representative of a novel bacterial phylum in the bioremediation process of the arsenic-rich ecosystem. In the future, similar strategies will lead to important advances in microbial ecology, revealing what was recently regarded as impossible to explore.

Materials and Methods

Sampling and Chemical Analysis. Five cm deep white sediments covering the bottom of the Reigous creek were collected at the station called COWG (13), and stored at -80°C. The main physico-chemical parameters (pH, T°, dissolved oxygen...) were determined in the field and arsenic speciation, Fe(II) and sulfate analyses were performed as previously described (9). A pH value of roughly 3.5 was measured, and arsenic, iron and sulfate concentrations reached 87, 625 and 3,209 mg.l⁻¹, respectively (Table S1).

DNA Isolation and Sequencing. DNA was extracted from the cellular fraction either directly using the UltraClean Soil DNA Isolation Kit (MoBio Laboratories Inc., Carlsbad, CA, USA), or after separation of microbial cells by Nycodenz gradient, using the Wizard

Genomic DNA Extraction kit (Promega, U.S.A.), and stored at -20°C. DNA was then fragmented by nebulisation and fragments ranging from 3 and 5 kb were used to construct a genomic library. DNA inserts were sequenced from both ends using chain-terminator, or by GS-FLEX pyrosequencing using standard procedures (see *SI Text*).

Bioinformatics. Contigs were assembled (see *SI Text*) and integrated into the MicroScope platform (<http://www.genoscope.cns.fr/agc/mage/carnoulescope>) for the prediction of coding sequences, followed by automatic and expert annotation (30). Molecular phylogenies were inferred using 16S rRNA gene library and 27 marker genes chosen from the reference gene set proposed to reconstruct the tree of life (see *SI Text*). Minimum generation times for the super-contigs were predicted from the intensity of the codon usage bias in highly expressed genes, as previously described (31). The metabolic network was predicted by the “Pathway Tools” software (32) using MetaCyc (33) as a reference pathway database and then analysed by a Multiple Factor Analysis (MFA) (see *SI Text*).

Protein Extraction and Mass Spectrometry Identification. Proteins extracted from bacterial cells were separated by SDS-PAGE, stained with Coomassie Brilliant Blue R-250, and *in gel* digested, as previously described (19). The resulting peptide extracts were analyzed by nanoLC-MS/MS, and data were analyzed using the MASCOT 2.2.0. algorithm (Matrix Science, London, UK) to search against a target-decoy protein database (see *SI Text*).

Bacterial Culture and Growth Medium. Iron oxidation was tested with *At. ferrooxidans* B5 strain isolated from Carnoulès (13) and grown at 30°C for 5 days on 100:10 solid medium (34) supplemented or not with 1 mg.L⁻¹ cobalamin (Sigma). Cultures were grown at 30°C and shaken at 140 rpm. Iron oxidation was followed by the appearance of an orange precipitate.

REFERENCES

1. Haferburg G & Kothe E (2007) Microbes and metals: interactions in the environment. *J. Basic Microbiol.* 47:453–467.
2. Falkowski PG, Fenchel T, & Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* 320:1034-1039.
3. Allen EE & Banfield JF (2005) Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* 3:489-498.
4. Bertin PN, Médigue C, & Normand P (2008) Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. *Microbiology* 154:347-359.
5. Wilmes P, Simmons SL, Deneff VJ, & Banfield JF (2009) The dynamic genetic repertoire of microbial communities. *FEMS Microbiol. Rev.* 33(1):109-132.
6. Lebrun E, *et al.* (2003) Arsenite oxidase, an ancient bioenergetic enzyme. *Mol Biol Evol* 20(5):686-693.
7. Vaughan DJ (2006) Arsenic. *Elements* 2(2):71-75.
8. Johnson DB & Hallberg KB (2003) The microbiology of acidic mine waters. *Res. Microbiol.* 154(7):466-473.
9. Bruneel O, *et al.* (2008) Archaeal diversity in a Fe-As rich acid mine drainage at Carnoulès (France). *Extremophiles* 12(4):563-571.
10. Casiot C, *et al.* (2003) Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res.* 37:2929-2936.
11. Duquesne K, *et al.* (2003) Immobilization of arsenite and ferric iron by *Acidithiobacillus ferrooxidans* and its relevance in acid mine drainage. *Appl. Env. Microbiol.* 69:6165-6173.
12. Duquesne K, *et al.* (2008) Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. *Environ. Microbiol.* 10:228-237.
13. Bruneel O, *et al.* (2003) Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J. Appl. Microbiol.* 95:492-499.
14. Ciccarelli FD, *et al.* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311(5765):1283-1287.
15. Huynen MA & Bork P (1998) Measuring genome evolution. *Proc. Natl. Acad. Sci. U S A.* 95(11):5849-5856.
16. Murray RGE & Stackebrandt E (1995) Taxonomic note: Implementation of the provisional status *Candidatus* for incompletely described procaryotes. *Int. J. Syst. Bacteriol.* 45(1):186-187.
17. Muller D, *et al.* (2007) A tale of two oxydation states : bacterial colonization of arsenic-rich environment. *PLoS Genet.* 3(4):e53: 518-530.
18. Qin J, *et al.* (2006) Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* 103(7):2075-2080.
19. Weiss S, *et al.* (2009) Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data. *Biochimie* 91:192-203.
20. Tendeng C & Bertin PN (2003) H-NS in Gram-negative bacteria: A family of multifaceted proteins. *Trends Microbiol.* 11(11):511-518.
21. Lutters-Czekalla S (1990) Lithoautotrophic growth of the iron bacterium *Gallionella ferruginea* with thiosulfate or sulfide as energy source. *Arch. Microbiol.* 154(5):417-421.

22. Morin G, *et al.* (2003) Bacterial formation of tooeleite and mixed arsenic(III) or arsenic(V)-iron(III) gels in the Carnoulès acid mine drainage, France. A XANES, XRD, and SEM study. *Environ. Sci. Technol.* 37:1705-1712.
23. Arsène-Ploetze F, *et al.* (2010) Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet* 6(2):e1000859.
24. Kreimeyer A, *et al.* (2007) Identification of the last unknown genes in the fermentation pathway of lysine. *J. Biol. Chem.* 282(10):7191-7197.
25. Escalante-Semerena JC (2007) Conversion of cobinamide into adenosylcobamide in Bacteria and Archaea. *J. Bacteriol.* 189(13):4555-4560.
26. Casiot C, Bruneel O, Personné JC, Leblanc M, & Elbaz-Poulichet F (2004) Arsenic oxidation and bioaccumulation by the acidophilic protozoan, *Euglena mutabilis*, in acid mine drainage (Carnoules, France). *Sci. Total Environ.* 320:259-267.
27. Olaveson MM & Stokes PM (1989) Responses of the acidophilic alga *Euglena mutabilis* (Euglenophyceae) to carbon enrichment at pH 3. *J. Phycol.* 25:529-539.
28. Bré MH, El Ferjani E, & Lefort-Tran M (1981) Sequential protein-dependent steps in the cell cycle initiation and completion of division in vitamin B12 replenished *Euglena gracilis*. *Protoplasma* 108:301-318.
29. Tyson GW, *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428(6978):37-43.
30. Vallenet D, *et al.* (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* 34(1):53-65.
31. Vieira-Silva S & Rocha EPC (2009) The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet.* 6:e1000808.
32. Karp P, Paley S, & Romero P (2002) The Pathway Tools Software. *Bioinformatics* 18(Suppl. 1):S1-S8.
33. Caspi R, *et al.* (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases *Nucleic Acids Res.* 36(1):D623-D631.
34. Schrader JA & Holmes DS (1988) Phenotypic switching of *Thiobacillus ferrooxidans*. *J. Bacteriol.* 170(9):3915-3923.

ACKNOWLEDGEMENTS. A.H.-S., J.C.-A. and D.H. were supported by a grant from the French Ministry of Education and Research and M.M. and F.G.-C. by a grant from the Agence Nationale de la Recherche (ANR). Financial support came from the Université de Strasbourg (UdS), the Consortium National de Recherche en Génomique (CNRG), the Centre National de la Recherche Scientifique (CNRS) and the Agence Nationale de la Recherche (ANR) in the frame of the RARE and MicroScope projects. ISfinder is supported by the CNRS and has received some support from the ARC. This work was done in the frame of the « Groupement de Recherche - Métabolisme de l’Arsenic chez les Microorganismes (GDR2909-CNRS) » (<http://gdr2909.u-strasbg.fr>).

Author contributions: P.B. coordinated the project, contributed to the analysis of genome information and wrote the manuscript, A.H.-S. and F.G.-C. contributed to metagenomic and proteomic data analysis and to the writing of the manuscript, E.P. performed genome binning and polymorphism analysis, F.A.-P. initiated the experimental work and contributed to proteomic experiments, B.C.W.M. and A.C. performed phylogenetic analyses, L.G., B.L. and R.D. constructed the 16S rDNA environmental library and contributed to data analysis, D.V., D.M., A.A.T.S, S.V.-S. and E.P.C.R. performed the statistical analysis of metabolic data, S.G. and A.V.D. contributed to proteomic data analysis, C.C. and F.E.-P. performed physico-chemical analyses, M.M. and D.H. performed physiology experiments, O.B., J.C.-A., E.T., N.F., V.B., P.O., M.B., P.S., and M.C. contributed to genome annotation and/or data analysis, C.M. coordinated software development for genome analysis (MaGe), J.W. and D.L.P. contributed to metagenome assembling and to the writing of the manuscript.

Data deposition : DNA sequences were submitted to GenBank and were given the following ProjectIDs : 38045 - Carnoules metagenome study (top level) ; 41535 - Carnoules metagenome study, bin 1 ; 41537 - Carnoules metagenome study, bin 2 ; 41539 - Carnoules metagenome study, bin 3 ; 41541 - Carnoules metagenome study, bin 4 ; 41543 - Carnoules metagenome study, bin 5 ; 41545 - Carnoules metagenome study, bin 6 ; 41549 - Carnoules metagenome study, bin 7.

Figure legends

Fig. 1. Phylogenetic tree representing the taxonomic affiliation of the Carnoulès community microorganisms. The 16S rRNA gene sequences were obtained from DNA sediments after PCR amplification (clones CG-X) or metagenomic sequencing (CARN bins, except for CARN4, see text). A total of 759 positions in the final dataset were obtained after all positions containing gaps and missing data were eliminated. The distances were computed using the Maximum Composite Likelihood method and the tree was inferred using the Neighbor-Joining method. The scale bar corresponds to 0.05 substitutions per site. Percentages of 1000 bootstrap resamplings that supported the branching orders in each analysis are shown above or near the relevant nodes. Bootstrap values are shown for branches with more than 50% bootstrap support.

Fig. 2. Model of the Carnoulès bacterial community highlighting the major functions identified by metagenome sequencing or metaproteome characterization. These activities include carbon and nitrogen fixation, energy metabolism, flagellum and capsule biosynthesis, amino acid transport and degradation, detoxification and stress response, arsenic and iron metabolism. The possible interactions between these microorganisms or with other chemical or biological components present on the study site are indicated by arrows. CARN bins are numbered from 1 to 7.

Fig. 3. Multiple Factorial Analysis of the 7 Carnoulès bins, performed on a two-dimensional matrix combining bins and enzymatic reactions, respectively. To highlight possible metabolic distinctions between bins, 3 axes (F1 to F3) capturing the highest dispersion in the resulting dot cloud were selected; they represent more than half the total dispersion. Colored

lines represent the vectors corresponding to the enzymatic reaction frequencies, the concentric disks differentiating those identified in the metaproteomic data. The reaction vectors were then hierarchically clustered, which led to 7 and 9 classes (indicated by the colors) for the first and second factorial planes, respectively. The corresponding functions are listed in Table S9AB.

Figure 2

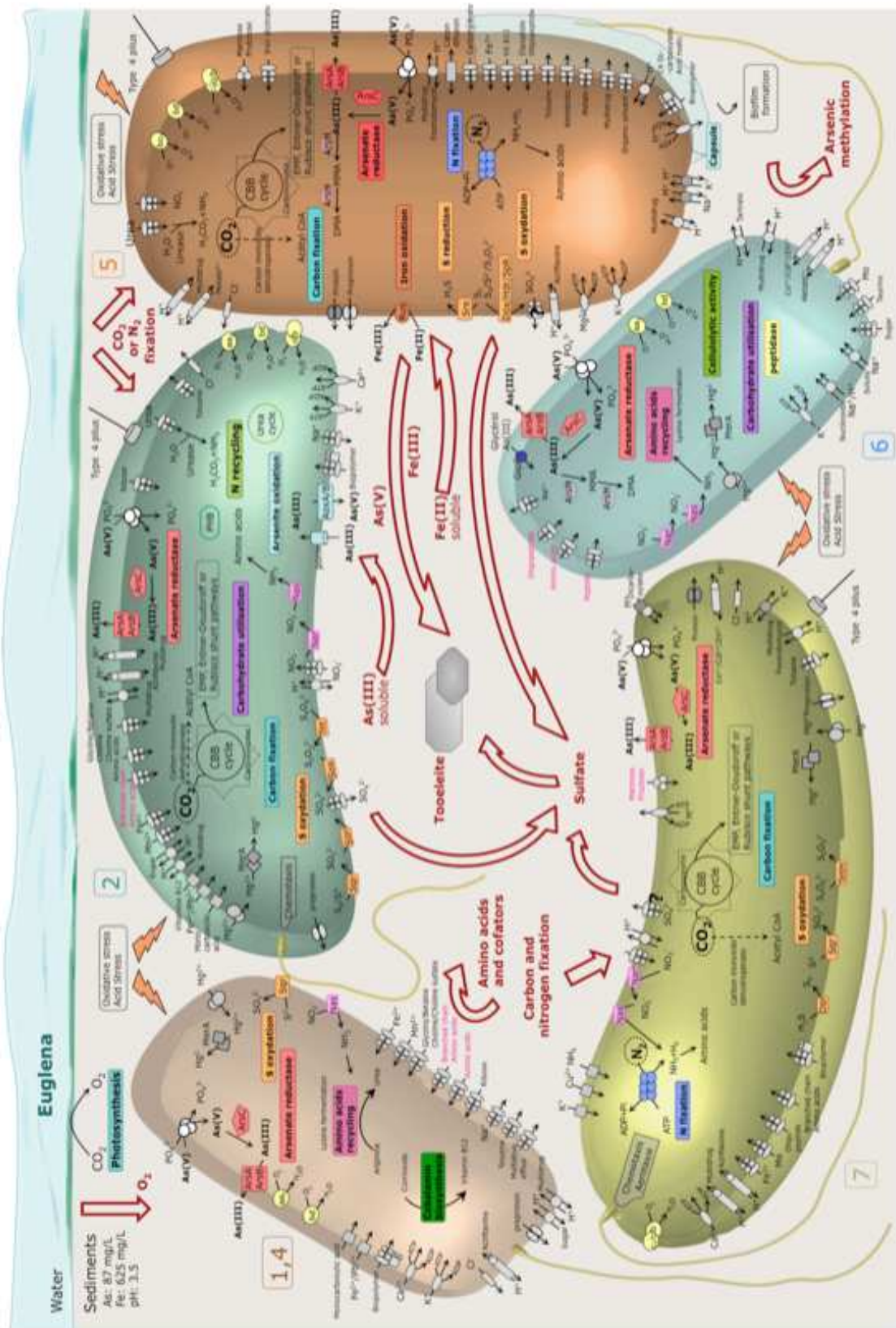
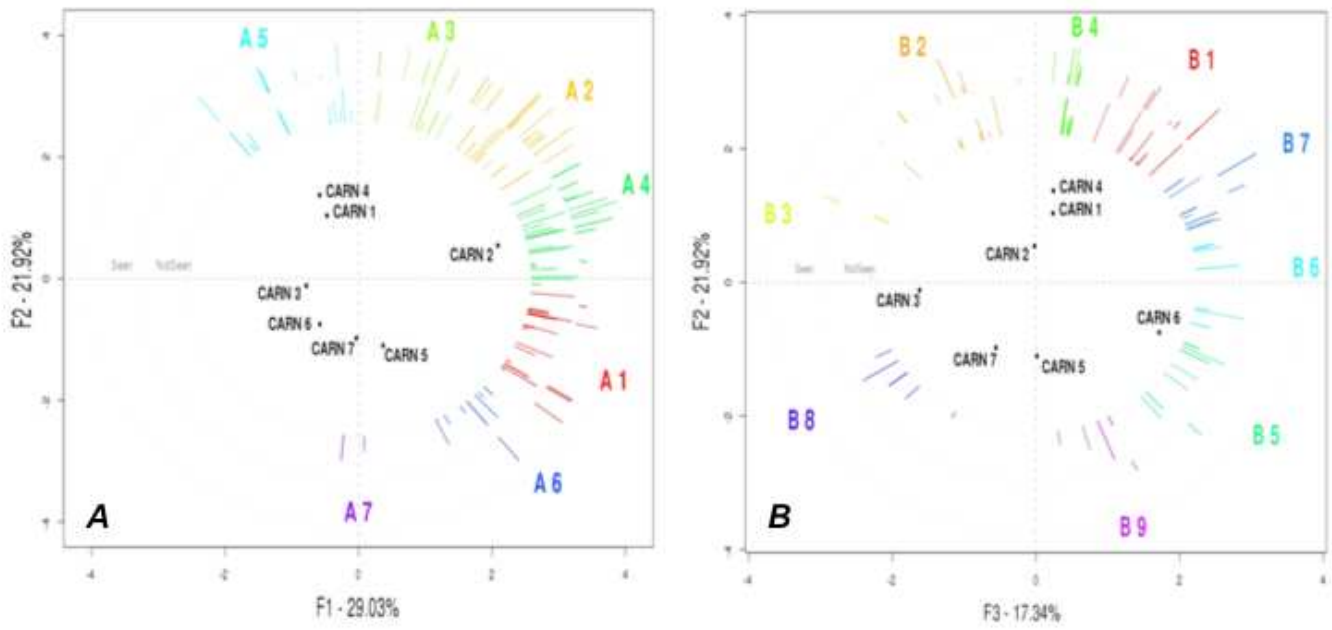


Figure 3

Dans ce chapitre, je présenterai l'expérience que j'ai eue dans l'enseignement en tant que monitrice. Je développerai ensuite ce qu'a représenté pour moi l'enseignement au cours de ces trois années.

Suite au concours de l'École Doctorale, j'ai eu l'opportunité d'obtenir un poste de monitrice à l'Université De Strasbourg. J'ai intégré l'équipe pédagogique de «Biologie» en Faculté de Psychologie en travaux pratiques (TP)/travaux dirigés (TD) pour les licences 1^{ère} ou 2^{ème} année. Ces enseignements ont été réalisés sous la tutelle de Jean-Christophe Cassel.

Au travers de différents TP et TD intégrés et complémentaires, cette unité d'enseignement a pour but de sensibiliser et de présenter aux étudiants en Psychologie les notions de base de la Biologie : un individu (qui devient un patient en Psychologie clinique) est un organisme qui provient d'une cellule unique au départ, qui va se diviser et se spécialiser, et possède des caractères qui lui sont propres. Au cours de mon monitorat j'ai assuré les TP de division cellulaire, d'évolution du système nerveux, d'histologie du système nerveux et les TD de génétique. De façon plus détaillées les notions abordées sont les suivantes :

- La génétique : Transmission des caractères, maladies autosomales, maladies liées aux chromosomes sexuels.
- La division cellulaire : Edification d'un organisme, mitose, méiose, gamétogénèse. Observations microscopiques des différents stades de la division cellulaire.
- Le système nerveux : Le système nerveux central, le système nerveux périphérique, les maladies liées à une dégénérescence du système nerveux. Evolution : céphalisation, spécialisation.

Ma charge d'enseignement se résume à une séance pour chaque TP ou TD. Mon travail d'encadrement a consisté à superviser les étudiants de 20 groupes. Pour les étudiants en 2^{ème} année de Licence, j'étais aussi en charge de leur évaluation. Ici, la principale difficulté se situe dans l'homogénéité des séances, en effet lorsqu'un TP est évalué par un contrôle terminal il s'agit d'enseigner la même chose à tous les étudiants.

Ces trois années passées à faire de la recherche et de l'enseignement ont été particulièrement enrichissantes tant au niveau personnel que professionnel. Il est vrai que le temps passé à enseigner est du temps que l'on ne peut pas accorder au travail de recherche à proprement parler. Je souhaite, tout d'abord, remercier mes directeurs de

thèse qui m'ont permis de m'investir dans cette activité tout en conciliant le travail de thèse. Ce temps passé en dehors du laboratoire est l'occasion de diversifier ses activités, d'avoir des contacts humains extérieurs et différents de ceux du laboratoire et encourage à être organisée et efficace.

Les expériences d'enseignement en génétique et en division cellulaire m'ont permis de mettre à profit les connaissances acquises durant mon cursus universitaire. A l'inverse, pour les enseignements d'histologie et d'évolution du système nerveux j'ai pu approfondir mes connaissances. Cette approche sans pré-requis a été un bon exercice pédagogique.

Durant mon monitorat, j'ai également suivi de nombreuses formations organisées par le CIES (Centre d'Initiation à l'Enseignement Supérieur). J'ai ainsi participé à une formation sur la gestion du trac, le placement de la voix, la lecture rapide, la pédagogie ...

Le monitorat est un autre moyen pour passer « de l'autre côté de la barrière », apprendre à devenir autonome, à prendre la parole en public, à prendre des responsabilités et à partager et proposer des idées.

D'un point de vue plus personnel, j'ai pris beaucoup de plaisir à enseigner. Lorsque certains étudiants quittent le cours en ayant appris quelque chose est extrêmement gratifiant. L'enseignement et la recherche me semblent complémentaires car ils permettent une dynamique entre l'acquisition des connaissances et sa transmission. Cette expérience m'a confortée dans l'objectif d'orienter ma vie professionnelle vers une carrière dans l'enseignement.

A

- Ades, S. E., (2004) Control of the alternative sigma factor σ E in *Escherichia coli*. *Current Opinion in Microbiology* **7**: 157-162.
- Akiyama, Y., A. Kihara, H. Tokuda & K. Ito, (1996) FtsH (HflB) is an ATP-dependent protease selectively acting on SecY and some other membrane proteins. *Journal of Biological Chemistry* **271**: 31196-31201.
- Alba, B. M., J. A. Leeds, C. Onufryk, C. Z. Lu & C. A. Gross, (2002) DegS and YaeL participate sequentially in the cleavage of RseA to activate the σ E-dependent extracytoplasmic stress response. *Genes and Development* **16**: 2156-2168.
- Anderson, G. L., J. Williams & R. Hille, (1992) The purification and characterization of arsenite oxydase from *Alcaligenes faecalis*, a molybdenum-containing hydroxylase. *J. Biol. Chem.* **267**: 23674-23682.
- Anilkumar, G., R. Srinivasan, S. P. Anand & P. Ajitkumar, (2001) Bacterial cell division protein FtsZ is a specific substrate for the AAA family protease FtsH. *Microbiology* **147**: 516-517.
- Aposhian, H. V. & M. M. Aposhian, (2006) Arsenic toxicology: Five questions. *Chemical Research in Toxicology* **19**: 1-15.
- Armitage, J. P., (2006) Bacterial behavior. The prokaryotes. Martin Dworkin, ed. *Springer, Chapter 1.5* 117-121.
- Arnér, E. S. J. & A. Holmgren, (2000) Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry* **267**: 6102-6109.
- Arsène-Ploetze, F., S. Koechler, M. Marchal, J. Y. Coppée, M. Chandler, V. Bonnefoy, C. Brochier-Armanet, M. Barakat, V. Barbe, F. Battaglia-Brunet, O. Bruneel, C. G. Bryan, J. Cleiss-Arnold, S. Cruveiller, M. Erhardt, A. Heinrich-Salmeron, F. Hommais, C. Jouliau, E. Krin, A. Lieutaud, D. Lièremont, C. Michel, D. Muller, P. Ortet, C. Proux, P. Siguier, D. Roche, Z. Rouy, G. Salvignol, D. Slyemi, E. Talla, S. Weiss, J. Weissenbach, C. Médigue & P. N. Bertin, (2010) Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genetics* **6**.
- Azam, T. A. & A. Ishihama, (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *Journal of Biological Chemistry* **274**: 33105-33113.

B

- Barbosa, R. L. & C. E. Benedetti, (2007) BigR, a transcriptional repressor from plant-associated bacteria, regulates an operon implicated in biofilm growth. *Journal of Bacteriology* **189**: 6185-6194.
- Battaglia-Brunet, F., C. Jouliau, F. Garrido, M. C. Dictor, D. Morin, K. Coupland, D. Barrie Johnson, K. B. Hallberg & P. Baranger, (2006) Oxidation of arsenite by *Thiomonas* strains and characterization of *Thiomonas arsenivorans* sp. nov. *Antonie van*

- Leeuwenhoek, International Journal of General and Molecular Microbiology* **89**: 99-108.
- Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill & D. A. Hopwood, (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141-147.
- Bertani, D., A. B. Oppenheim & F. Narberhaus, (2001) An internal region of the RpoH heat shock transcription factor is critical for rapid degradation by the FtsH protease. *FEBS Letters* **493**: 17-20.
- Biers, E. J., S. Sun & E. C. Howard, (2009) Prokaryotic genomes and diversity in surface ocean waters: Interrogating the global ocean sampling metagenome. *Applied and Environmental Microbiology* **75**: 2221-2229.
- Booth, I. R., (2002) Stress and the single cell: Intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. *International Journal of Food Microbiology* **78**: 19-30.
- Bourret, R. B., K. A. Borkovich & M. I. Simon, (1991) Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annual Review of Biochemistry* **60**: 401-441.
- Branco, R., A. P. Chung & P. V. Morais, (2008) Sequencing and expression of two arsenic resistance operons with different functions in the highly arsenic-resistant strain *Ochrobactrum tritici* SCII24T. *BMC Microbiology* **8**.
- Branco, R., R. Francisco, A. P. Chung & P. V. Morais, (2009) Identification of an *aox* system that requires cytochrome c in the highly arsenic-resistant bacterium *Ochrobactrum tritici* SCII24. *Applied and Environmental Microbiology* **75**: 5141-5147.
- Brown, M. H., I. T. Paulsen & R. A. Skurray, (1999) The multidrug efflux protein NorM is a prototype of a new family of transporters [2]. *Molecular Microbiology* **31**: 394-395.
- Brown, N. L., J. V. Stoyanov, S. P. Kidd & J. L. Hobman, (2003) The MerR family of transcriptional regulators. *FEMS Microbiology Reviews* **27**: 145-163.
- Browning, D. F., C. M. Beatty, A. J. Wolfe, J. A. Cole & S. J. W. Busby, (2002) Independent regulation of the divergent *Escherichia coli* *nrfA* and *acsP1* promoters by a nucleoprotein assembly at a shared regulatory region. *Molecular Microbiology* **43**: 687-701.
- Browning, D. F. & S. J. W. Busby, (2004) The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**: 57-65.

- Browning, D. F., J. A. Cole & S. J. W. Busby, (2000) Suppression of FNR-dependent transcription activation at the *Escherichia coli nir* promoter by Fis, IHF and H-NS: Modulation of transcription by a complex nucleo-protein assembly. *Molecular Microbiology* **37**: 1258-1269.
- Bruneel, O., N. Pascault, M. Egal, C. Bancon-Montigny, M. S. Goni-Urriza, F. Elbaz-Poulichet, J. C. Person" & R. Duran, (2008) Archaeal diversity in a Fe-As rich acid mine drainage at Carnoulès (France). *Extremophiles* **12**: 563-571.
- Bruneel, O., J. C. Personné, C. Casiot, M. Leblanc, F. Elbaz-Poulichet, B. J. Mahler, A. Le Flèche & P. A. D. Grimont, (2003) Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *Journal of Applied Microbiology* **95**: 492-499.
- Bryan, C. G., M. Marchal, F. Battaglia-Brunet, V. Kugler, C. Lemaitre-Guillier, D. Lièvreumont, P. N. Bertin & F. Arsène-Ploetze, (2009) Carbon and arsenic metabolism in *Thiomonas* strains: Differences revealed diverse adaptation processes. *BMC Microbiology* **9**.
- Buck, M., M. T. Gallegos, D. J. Studholme, Y. Guo & J. D. Gralla, (2000) The bacterial enhancer-dependent σ_{54} ($\sigma(N)$) transcription factor. *Journal of Bacteriology* **182**: 4129-4136.
- Busby, S. & R. H. Ebright, (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**: 742-746.
- Busby, S. & R. H. Ebright, (1997) Transcription activation at class II CAP-dependent promoters. *Molecular Microbiology* **23**: 853-859.
- Busenlehner, L. S., M. A. Pennella & D. P. Giedroc, (2003) The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiology Reviews* **27**: 131-143.
- Busscher, H. J., R. Bos & H. C. Van der Mei, (1995) Initial microbial adhesion is a determinant for the strength of biofilm adhesion. *FEMS Microbiology Letters* **128**: 229-234.
- Butcher, B. G., S. M. Deane & D. E. Rawlings, (2000) The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. *Applied and Environmental Microbiology* **66**: 1826-1833.
- Butcher, B. G. & D. E. Rawlings, (2002) The divergent chromosomal ars operon of *Acidithiobacillus ferrooxidans* is regulated by an atypical ArsR protein. *Microbiology* **148**: 3983-3992.
- C**
- Cai, L., C. Rensing, X. Li & G. Wang, (2009) Novel gene clusters involved in arsenite oxidation and resistance in two arsenite oxidizers: *Achromobacter* sp. SY8 and *Pseudomonas* sp. TS44. *Applied Microbiology and Biotechnology* **83**: 715-725.

- Cannon, W., F. Claverie-Martin, S. Austin & M. Buck, (1994) Identification of a DNA-contacting surface in the transcription factor sigma-54. *Molecular Microbiology* **11**: 227-236.
- Canovas, D., I. Cases & V. De Lorenzo, (2003) Heavy metal tolerance and metal homeostasis in *Pseudomonas putida* as revealed by complete genome analysis. *Environmental Microbiology* **5**: 1242-1256.
- Caramori, T., D. Barillà, C. Nessi, L. Sacchi & A. Galizzi, (1996) Role of FlgM in σ D-dependent gene expression in *Bacillus subtilis*. *Journal of Bacteriology* **178**: 3113-3118.
- Cases, I. & V. de Lorenzo, (2005) Promoters in the environment: Transcriptional regulation in its natural context. *Nature Reviews Microbiology* **3**: 105-118.
- Cezairliyan, B. O. & R. T. Sauer, (2007) Inhibition of regulated proteolysis by RseB. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 3771-3776.
- Charon, N. W. & S. F. Goldstein, (2002) Genetics of motility and chemotaxis of a fascinating group of bacteria: The spirochetes. In: Annual Review of Genetics. pp. 47-73.
- Chen, C. M., T. K. Misra, S. Silver & B. P. Rosen, (1986) Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. *Journal of Biological Chemistry* **261**: 15030-15038.
- Chen, C. M., H. L. T. Mobley & B. P. Rosen, (1985) Separate resistances to arsenate and arsenite (antimonate) encoded by the arsenical resistance operon of R factor R773. *Journal of Bacteriology* **161**: 758-763.
- Chen, Y. & B. P. Rosen, (1997) Metalloregulatory properties of the ArsD repressor. *Journal of Biological Chemistry* **272**: 14257-14262.
- Choy, H. & S. Adhya, (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology*. (ed. Neidhardt, F.C.) (**ASM Press, Washington**): 1287-1299.
- Costa De Oliveira, R., J. Laval & S. Boiteux, (1987) Induction of SOS and adaptive responses by alkylating agents in *Escherichia coli* mutants deficient in 3-methyladenine-DNA glycosylase activities. *Mutation Research* **183**: 11-20.
- D**
- Dahl, C. C., G. Friedrich., (2008) Microbial sulfur metabolism.
- Danielli, A., S. Romagnoli, D. Roncarati, L. Costantino, I. Delany & V. Scarlato, (2009) Growth phase and metal-dependent transcriptional regulation of the *fecA* genes in *Helicobacter pylori*. *Journal of Bacteriology* **191**: 3717-3725.
- Daubin, V., E. Lerat & G. Perrière, (2003) The source of laterally transferred genes in bacterial genomes. *Genome biology* **4**.

- Del Razo, L. M., B. Quintanilla-Vega, E. Brambila-Colombres, E. S. Calderón -Aranda, M. Manno & A. Albores, (2001) Stress proteins induced by arsenic. *Toxicology and Applied Pharmacology* **177**: 132-148.
- Dey, S. & B. P. Rosen, (1995) Dual mode of energy coupling by the oxyanion-translocating ArsB protein. *Journal of Bacteriology* **177**: 385-389.
- Dickinson, D. A. & H. J. Forman, (2002) Cellular glutathione and thiols metabolism. *Biochemical Pharmacology* **64**: 1019-1026.
- Dixon, R., (2002) in *Signals, Switches, Regulons, and Cascades*. (eds Hodgson, D.A & Thomas C.M.): 212-230.
- Dove, S. L., S. A. Darst & A. Hochschild, (2003) Region 4 of σ as a target for transcription regulation. *Molecular Microbiology* **48**: 863-874.
- Duquesne, K., S. Lebrun, C. Casiot, O. Bruneel, J. C. Personné, M. Leblanc, F. Elbaz-Poulichet, G. Morin & V. Bonnefoy, (2003) Immobilization of Arsenite and Ferric Iron by *Acidithiobacillus ferrooxidans* and Its Relevance to Acid Mine Drainage. *Applied and Environmental Microbiology* **69**: 6165-6173.
- Duquesne, K., A. Lieutaud, J. Ratouchniak, D. Muller, M. C. Lett & V. Bonnefoy, (2008) Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: From the strain isolation to the gene study. *Environmental Microbiology* **10**: 228-237.
- Duquesne, K., A. Lieutaud, J. Ratouchniak, A. Yarz-bal & V. Bonnefoy, (2007) Mechanisms of arsenite elimination by *Thiomonas* sp. isolated from Carnoulès acid mine drainage. *European Journal of Soil Biology* **43**: 351-355.
- E**
- Ebright, R. H., (2000) RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *Journal of Molecular Biology* **304**: 687-698.
- Ellis, P. J., T. Conrads, R. Hille & P. Kuhn, (2001) Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 Å and 2.03 Å. *Structure* **9**: 125-132.
- England, J. C. & J. W. Gober, (2001) Cell cycle control of cell morphogenesis in *Caulobacter*. *Current Opinion in Microbiology* **4**: 674-680.
- Erill, I., S. Campoy & J. Barbé, (2007) Aeons of distress: An evolutionary perspective on the bacterial SOS response. *FEMS Microbiology Reviews* **31**: 637-656.
- Errington, J., (1993) *Bacillus subtilis* sporulation: Regulation of gene expression and control of morphogenesis. *Microbiological Reviews* **57**: 1-33.

F

- Fernández De Henestrosa, A. R., T. Ogi, S. Aoyagi, D. Chafin, J. J. Hayes, H. Ohmori & R. Woodgate, (2000) Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Molecular Microbiology* **35**: 1560-1572.
- Foulongne, V., S. Michaux-Charachon, D. O'Callaghan & M. Ramuz, (2002) Type IV secretion systems and bacterial virulence. *Systèmes de sécrétion des protéines de type IV et virulence bactérienne* **18**: 439-447.
- Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek & C. A. Phillips, (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397-403.
- Frees, D., F. K. Vogensen & H. Ingmer, (2003) Identification of proteins induced at low pH in *Lactococcus lactis*. *International Journal of Food Microbiology* **87**: 293-300.
- Fridovich, I., (1998) Oxygen toxicity: A radical explanation. *Journal of Experimental Biology* **201**: 1203-1209.
- Fu, J., A. L. Gnatt, D. A. Bushnell, G. J. Jensen, N. E. Thompson, R. R. Burgess, P. R. David & R. D. Kornberg, (1999) Yeast RNA polymerase II at 5 Å resolution. *Cell* **98**: 799-810.
- Fujimoto, D. F., R. H. Higginbotham, K. M. Sterba, S. J. Maleki, A. M. Segall, M. S. Smeltzer & B. K. Hurlburt, (2009) *Staphylococcus aureus* SarA is a regulatory protein responsive to redox and pH that can support bacteriophage lambda integrase-mediated excision/ recombination. *Molecular Microbiology* **74**: 1445-1458.

G

- Gailer, J., (2007) Arsenic-selenium and mercury-selenium bonds in biology. *Coordination Chemistry Reviews* **251**: 234-254.
- Genevaux, P., C. Georgopoulos & W. L. Kelley, (2007) The Hsp70 chaperone machines of *Escherichia coli*: A paradigm for the repartition of chaperone functions. *Molecular Microbiology* **66**: 840-857.
- Gil, F., I. Hernández-Lucas, R. Polanco, N. Pacheco, B. Collao, J. M. Villareal, G. Nardocci, G. Calva & C. P. Saavedra, (2009) SoxS regulates the expression of the *Salmonella enterica* serovar Typhimurium *ompW* gene. *Microbiology* **155**: 2490-2497.
- Gong, Y., X. L. Tian, T. Sutherland, G. Sisson, J. Mai, J. Ling & Y. H. Li, (2009) Global transcriptional analysis of acid-inducible genes in *Streptococcus mutans*: Multiple two-component systems involved in acid adaptation. *Microbiology* **155**: 3322-3332.
- Gourse, R. L., W. Ross & T. Gaal, (2000) UPs and downs in bacterial transcription initiation: The role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular Microbiology* **37**: 687-695.

- Grainger, D. C., D. Hurd, M. D. Goldberg & S. J. W. Busby, (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Research* **34**: 4642-4652.
- Grall, N., J. Livny, M. Waldor, M. Barel, A. Charbit & K. L. Meibom, (2009) Pivotal role of the *Francisella tularensis* heat-shock sigma factor RpoH. *Microbiology* **155**: 2560-2572.
- Green, H. H., (1918) Isolation and description of a bacterium causing oxidation of arsenite to arsenate in cattle-dipping baths. *Rep Dir Vet Res S Afr* **6**: 593-599.
- Griffith, K. L., M. M. Fitzpatrick, E. F. Keen Iii & R. E. Wolf Jr, (2009) Two Functions of the C-Terminal Domain of *Escherichia coli* Rob: Mediating "Sequestration-Dispersal" as a Novel Off-On Switch for Regulating Rob's Activity as a Transcription Activator and Preventing Degradation of Rob by Lon Protease. *Journal of Molecular Biology* **388**: 415-430.
- Griffith, K. L. & R. E. Wolf Jr, (2004) Genetic evidence for pre-recruitment as the mechanism of transcription activation by SoxS of *Escherichia coli*: The dominance of DNA binding mutations of SoxS. *Journal of Molecular Biology* **344**: 1-10.
- Gross, C. A., M. Lonetto and R. Losick., (1992) Bacterial sigma factors. *In: Transcription Regulation. Cold Spring Harbor Laboratory Press.*
- Guelfo, J. R., A. Rodriguez-Rojas, I. Matic & J. s. Blázquez, (2010) A MATE-Family Efflux Pump Rescues the *Escherichia coli* 8-Oxoguanine-Repair-Deficient Mutator Phenotype and Protects Against H₂O₂ Killing. *PLoS Genet* **6**: e1000931.

H

- Hakenbeck, R. & J. B. Stock, (1996) Analysis of two-component signal transduction systems involved in transcriptional regulation. *In: Methods in Enzymology.* pp. 281-300.
- Handelsman, J., (2004) Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68**: 669-685.
- Harrison, J. J., H. Ceri & R. J. Turner, (2007) Multimetal resistance and tolerance in microbial biofilms. *Nature Reviews Microbiology* **5**: 928-938.
- Harshey, R. M., (2003) Bacterial Motility on a Surface: Many Ways to a Common Goal. *In: Annual Review of Microbiology.* pp. 249-273.
- He, Z., A. Zhou, E. Baidoo, Q. He, M. P. Joachimiak, P. Benke, R. Phan, A. Mukhopadhyay, C. L. Hemme, K. Huang, E. J. Alm, M. W. Fields, J. Wall, D. Stahl, T. C. Hazen, J. D. Keasling, A. P. Arkin & J. Zhou, (2010) Global transcriptional, physiological and metabolite analyses of the responses of *Desulfovibrio vulgaris* Hildenborough to salt adaptation. *Applied and Environmental Microbiology* **76**: 1574-1586.
- Hecker, M., W. Schumann & U. Völker, (1996) Heat-shock and general stress response in *Bacillus subtilis*. *Molecular Microbiology* **19**: 417-428.

- Hedges, R. W. & S. Baumberg, (1973) Resistance to arsenic compounds conferred by a plasmid transmissible between strains of *Escherichia coli*. *Journal of Bacteriology* **115**: 459-460.
- Helmann, J. D., (1991) Alternative sigma factors and the regulation of flagellar gene expression. *Molecular Microbiology* **5**: 2875-2882.
- Helmann, J. D., (1999) Anti-sigma factors. *Current Opinion in Microbiology* **2**: 135-141.
- Herman, C., S. Prakash, C. Z. Lu, A. Matouschek & C. A. Gross, (2003) Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. *Molecular Cell* **11**: 659-669.
- Herrera, M. C., T. Krell, X. Zhang & J. L. Ramos, (2009) PhhR Binds to Target Sequences at Different Distances with Respect to RNA Polymerase in Order to Activate Transcription. *Journal of Molecular Biology* **394**: 576-586.
- Hlaváček, O. & L. Váchová, (2002) ATP-dependent proteinases in bacteria. *Folia Microbiologica* **47**: 203-212.
- Hoven, R. N. V., J. M. Santini & I. C. A. Streimann, (2004) Prokaryotic arsenic metabolism - Roles in arsenic cycling and bioremediation. In: Australasian Institute of Mining and Metallurgy Publication Series. pp. 93-103.
- Hümpel, A., S. Gebhard, G. M. Cook & M. Berney, (2010) The SigF regulon in *Mycobacterium smegmatis* reveals roles in adaptation to stationary phase, heat, and oxidative stress. *Journal of Bacteriology* **192**: 2491-2502.

I

- Iannino, F., R. A. Ugalde & N. Iñón de Iannino, (2008) Characterization of *Brucella abortus* sigma factor σ_{54} (rpoN): Genetic complementation of *Sinorhizobium meliloti* ntrA mutant. *Microbial Pathogenesis* **45**: 394-402.
- Imlay, J. A. & S. Linn, (1988) DNA damage and oxygen radical toxicity. *Science* **240**: 1302-1309.
- Ishihama, A., (2000) Functional modulation of *Escherichia coli* RNA polymerase. In: Annual Review of Microbiology. pp. 499-518.

J

- Jakobsen, J. S., L. Jelsbak, R. D. Welch, C. Cummings, B. Goldman, E. Stark, S. Slater & D. Kaiser, (2004) σ_{54} enhancer binding proteins and *Myxococcus xanthus* fruiting body development. *Journal of Bacteriology* **186**: 4361-4368.
- Jang, H. J., C. Nde, F. Toghrol & W. E. Bentley, (2009) Global transcriptome analysis of the *Mycobacterium bovis* BCG response to sodium hypochlorite. *Applied Microbiology and Biotechnology* **85**: 127-140.
- Jarrell, K. F. & M. J. McBride, (2008) The surprisingly diverse ways that prokaryotes move. *Nature Reviews Microbiology* **6**: 466-476.

Jubete, Y., M. R. Maurizi & S. Gottesmanf, (1996) Role of the heat shock protein DnaJ in the Lon-dependent degradation of naturally unstable proteins. *Journal of Biological Chemistry* **271**: 30798-30803.

K

Kalir, S., J. McClure, K. Pabbaraju, C. Southward, M. Ronen, S. Leibler, M. G. Surette & U. Alon, (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292**: 2080-2083.

Karata, K., T. Inagawa, A. J. Wilkinson, T. Tatsuta & T. Ogura, (1999) Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. Site-directed mutagenesis of the ATP-dependent protease FtsH. *Journal of Biological Chemistry* **274**: 26225-26232.

Kashyap, D. R., L. M. Botero, W. L. Franck, D. J. Hassett & T. R. McDermott, (2006) Complex regulation of arsenite oxidation in *Agrobacterium tumefaciens*. *Journal of Bacteriology* **188**: 1081-1088.

Keyer, K. & J. A. Imlay, (1996) Superoxide accelerates DNA damage by elevating free-iron levels. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 13635-13640.

Kiliç, N. K., A. Stensballe, D. E. Otzen & G. Dönmez, (2009) Proteomic changes in response to chromium(VI) toxicity in *Pseudomonas aeruginosa*. *Bioresource technology* **101**: 2134-2140.

Koechler, S., J. Cleiss-Arnold, C. Proux, O. Sismeiro, M. A. Dillies, F. Goulhen-Chollet, F. Hommais, D. Lièvre-mont, F. Arsène-Ploetze, J. Y. Coppée & P. N. Bertin, (2010) Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. *BMC Microbiology* **10**.

Koffi, K., M. Leblanc, H. Jourde, C. Casiot, S. Pistre, P. Gouze & F. Elbaz-Poulichet, (2003) Reverse Oxidation Zoning in Mine Tailings Generating Arsenic-rich Acidic Waters (Carnoulès, France). *Mine Water and the Environment* **22**: 7-14.

Kültz, D., (2005) Molecular and evolutionary basis of the cellular stress response. In: *Annual Review of Physiology*. pp. 225-257.

L

Lahiri, C., S. Mandal, W. Ghosh, B. Dam & P. Roy, (2006) A novel gene cluster *soxSRT* is essential for the chemolithotrophic oxidation of thiosulfate and tetrathionate by *Pseudaminobacter salicylatoxydans* KCT001. *Current Microbiology* **52**: 267-273.

Landini, P., (2009) Cross-talk mechanisms in biofilm formation and responses to environmental and physiological stress in *Escherichia coli*. *Research in Microbiology* **160**: 259-266.

Langley, S. & T. J. Beveridge, (1999) Metal binding by *Pseudomonas aeruginosa* PAO1 is influenced by growth of the cells as a biofilm. *Canadian Journal of Microbiology* **45**: 616-622.

Little, J. W. & D. W. Mount, (1982) The SOS regulatory system of *Escherichia coli*. *Cell* **29**: 11-22.

Liu, S. X., M. Athar, I. Lippai, C. Waldren & T. K. Hei, (2001) Induction of oxyradicals by arsenic: Implication for mechanism of genotoxicity. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 1643-1648.

M

Ma, Q. & T. K. Wood, (2009) OmpA influences *Escherichia coli* biofilm formation by repressing cellulose production through the CpxRA two-component system. *Environmental Microbiology* **11**: 2735-2746.

Magariyama, Y., S. Sugiyama, K. Muramoto, Y. Maekawa, I. Kawagishi, Y. Imae & S. Kudo, (1994) Very fast flagellar rotation [9]. *Nature* **371**: 752.

Mandal, B. K. & K. T. Suzuki, (2002) Arsenic round the world: A review. *Talanta* **58**: 201-235.

Martin, R. G., W. K. Gillette, N. I. Martin & J. L. Rosner, (2002) Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in *Escherichia coli*. *Molecular Microbiology* **43**: 355-370.

Martinez, J. L., M. B. Sánchez, L. Martinez-Solano, A. Hernandez, L. Garmendia, A. Fajardo & C. Alvarez-Ortega, (2009) Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiology Reviews* **33**: 430-449.

Martinez-Antonio, A. & J. Collado-Vides, (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology* **6**: 482-489.

Mascher, T., J. D. Helmann & G. Uden, (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiology and Molecular Biology Reviews* **70**: 910-938.

Matsuo, T., J. Chen, Y. Minato, W. Ogawa, T. Mizushima, T. Kuroda & T. Tsuchiya, (2008) SmdAB, a heterodimeric ABC-type multidrug efflux pump, in *Serratia marcescens*. *Journal of Bacteriology* **190**: 648-654.

McCarter, L. L., (2006) Regulation of flagella. *Current Opinion in Microbiology* **9**: 180-186.

McLeod, S. M., S. E. Aiyar, R. L. Gourse & R. C. Johnson, (2002) The C-terminal domains of the RNA polymerase [alpha] subunits: contact site with fis and localization during co-activation with CRP at the *Escherichia coli proPP2* promoter. *Journal of Molecular Biology* **316**: 517-529.

McLeod, S. M. & R. C. Johnson, (2001) Control of transcription by nucleoid proteins. *Current Opinion in Microbiology* **4**: 152-159.

Meng, Y. L., Z. Liu & B. P. Rosen, (2004) As(III) and Sb(III) Uptake by GlpF and Efflux by ArsB in *Escherichia coli*. *Journal of Biological Chemistry* **279**: 18334-18341.

- Merino, S., J. G. Shaw & J. M. Tomás, (2006) Bacterial lateral flagella: An inducible flagella system. *FEMS Microbiology Letters* **263**: 127-135.
- Merrick, M. J., (1993) In a class of its own - The RNA polymerase sigma factor σ_{54} ($\sigma(N)$). *Molecular Microbiology* **10**: 903-909.
- Messens, J., G. Hayburn, A. Desmyter, G. Laus & L. Wyns, (1999) The essential catalytic redox couple in arsenate reductase from *Staphylococcus aureus*. *Biochemistry* **38**: 16857-16865.
- Michel, C., M. Jean, S. Coulon, M. C. Dictor, F. Delorme, D. Morin & F. Garrido, (2007) Biofilms of As(III)-oxidising bacteria: Formation and activity studies for bioremediation process development. *Applied Microbiology and Biotechnology* **77**: 457-467.
- Morin, G., F. Juillot, C. Casiot, O. Bruneel, J. C. Personné, F. Elbaz-Poulichet, M. Leblanc, P. Ildelfonse & G. Calas, (2003) Bacterial formation of tooeleite and Mixed Arsenic(III) or Arsenic(V) - Iron(III) gels in the carnoulès acid mine drainage, France. A XANES, XRD, and SEM study. *Environmental Science and Technology* **37**: 1705-1712.
- Morita, Y., A. Kataoka, S. Shiota, T. Mizushima & T. Tsuchiya, (2000) NorM of *Vibrio parahaemolyticus* is an Na⁺-driven multidrug efflux pump. *Journal of Bacteriology* **182**: 6694-6697.
- Morita, Y., K. Kodama, S. Shiota, T. Mine, A. Kataoka, T. Mizushima & T. Tsuchiya, (1998) NorM, putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy* **42**: 1778-1782.
- Morris, L., W. Cannon, F. Claverie-Martin, S. Austin & M. Buck, (1994) DNA distortion and nucleation of local DNA unwinding within sigma-54 ($\sigma(N)$) holoenzyme closed promoter complexes. *Journal of Biological Chemistry* **269**: 11563-11571.
- Muller, C., I. S. Bang, J. Velayudhan, J. Karlinsey, K. Papenfort, J. Vogel & F. C. Fang, (2009) Acid stress activation of the σE stress response in *Salmonella enterica* serovar Typhimurium. *Molecular Microbiology* **71**: 1228-1238.
- Muller, D., D. Lièvreumont, D. D. Simeonova, J. C. Hubert & M. C. Lett, (2003) Arsenite oxidase *aox* genes from a metal-resistant beta-proteobacterium. *Journal of Bacteriology* **185**: 135-141.
- Muller, D., C. Medigue, S. Koechler, V. Barbe, M. Barakat, E. Talla, V. Bonnefoy, E. Krin, F. Arsene-Ploetze, C. Carapito, M. Chandler, B. Cournoyer, S. Cruveiller, C. Dossat, S. Duval, M. Heymann, E. Leize, A. Lieutaud, D. Lièvreumont, Y. Makita, S. Mangenot, W. Nitschke, P. Ortet, N. Perdrial, B. Schoepp, P. Siguier, D. D. Simeonova, Z. Rouy, B. Segurens, E. Turlin, D. Vallenet, A. Van Dorsselaer, S. Weiss, J. Weissenbach, M. C. Lett, A. Danchin & P. N. Bertin, (2007) A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS genetics* **3**.
- Müller, J., S. Oehler & B. Müller-Hill, (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *Journal of Molecular Biology* **257**: 21-29.

Müller-Hill, B., (1998) Some repressors of bacterial transcription. *Curr. Opin. Microbiol.* **1**.

Mulvey, M. R. & P. C. Loewen, (1989) Nucleotide sequence of *katF* of *Escherichia coli* suggests KatF protein is a novel σ E transcription factor. *Nucleic Acids Research* **17**: 9979-9991.

N

Newman, D. K., T. J. Beveridge & F. M. M. Morel, (1997) Precipitation of arsenic trisulfide by *Desulfotomaculum auripigmentum*. *Applied and Environmental Microbiology* **63**: 2022-2028.

Neyt, C., M. Iriarte, V. H. Thi & G. R. Cornelis, (1997) Virulence and arsenic resistance in *Yersiniae*. *Journal of Bacteriology* **179**: 612-619.

Nies, D. H. & S. Silver, (1995) Ion efflux systems involved in bacterial metal resistances. *Journal of Industrial Microbiology* **14**: 186-199.

Ninfa, A. J. & B. Magasanik, (1989) Covalent modification of the *glnG* product, NRI, by the *glnL* product, NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli*. *Proc Natl Acad Sci U S A* **83**: 5909–5913.

O/P

O'Rourke, K. P., J. D. Shaw, M. W. Pesesky, B. T. Cook, S. M. Roberts, J. P. Bond & G. A. Spatafora, (2010) Genome-wide characterization of the SloR metalloregulome in *Streptococcus mutans*. *Journal of Bacteriology* **192**: 1433-1443.

Onnis-Hayden, A., H. Weng, M. He, S. Hansen, V. Ilyin, K. Lewis & A. Z. Gu, (2009) Prokaryotic real-time gene expression profiling for toxicity assessment. *Environmental Science and Technology* **43**: 4574-4581.

Ordonez, E., M. Letek, N. Valbuena, J. A. Gil & L. M. Mateos, (2005) Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032. *Applied and Environmental Microbiology* **71**: 6206-6215.

Oremland, R. S., S. E. Hoefft, J. M. Santini, N. Bano, R. A. Hollibaugh & J. T. Hollibaugh, (2002) Anaerobic oxidation of arsenite in Mono Lake water and by a facultative, arsenite-oxidizing chemoautotroph, strain MLHE-1. *Applied and Environmental Microbiology* **68**: 4795-4802.

Oremland, R. S., T. R. Kulp, J. S. Blum, S. E. Hoefft, S. Baesman, L. G. Miller & J. F. Stolz, (2005) Microbiology: A microbial arsenic cycle in a salt-saturated, extreme environment. *Science* **308**: 1305-1308.

Oremland, R. S. & J. F. Stolz, (2003) The ecology of arsenic. *Science* **300**: 939-944.

Osborne, F. H. & H. L. Ehrlich, (1976) Oxidation of arsenite by a soil isolate of *Alcaligenes*. *Journal of Applied Bacteriology* **41**: 295-305.

- Paez-Espino, D., J. Tamames, V. De Lorenzo & D. Canovas, (2009) Microbial responses to environmental arsenic. *BioMetals* **22**: 117-130.
- Page, W. J., E. Kwon, A. S. Cornish & A. E. Tindale, (2003) The *csbX* gene of *Azotobacter vinelandii* encodes an MFS efflux pump required for catechol siderophore export. *FEMS Microbiology Letters* **228**: 211-216.
- Pao, S. S., I. T. Paulsen & M. H. Saier Jr, (1998) Major facilitator superfamily. *Microbiology and Molecular Biology Reviews* **62**: 1-34.
- Parkinson, J., (1993) Signal transduction schemes of bacteria. *Cell*. **73**: 857-871.
- Parkinson, J. S. & E. C. Kofoid, (1992) Communication modules in bacterial signaling proteins. *Annual Review of Genetics* **26**: 71-112.
- Partridge, J. D., R. K. Poole & J. Green, (2007) The *Escherichia coli yhjA* gene, encoding a predicted cytochrome c peroxidase, is regulated by FNR and OxyR. *Microbiology* **153**: 1499-1507.
- Patel, P. C., F. Goulhen, C. Boothman, A. G. Gault, J. M. Charnock, K. Kalia & J. R. Lloyd, (2007) Arsenate detoxification in a Pseudomonad hypertolerant to arsenic. *Archives of Microbiology* **187**: 171-183.
- Paulsen, I. T., R. A. Skurray, R. Tam, M. H. Saier Jr, R. J. Turner, J. H. Weiner, E. B. Goldberg & L. L. Grinius, (1996) The SMR family: A novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. *Molecular Microbiology* **19**: 1167-1175.
- Peeters, E., A. Sass, E. Mahenthalingam, H. Nelis & T. Coenye, (2010) Transcriptional response of *Burkholderia cenocepacia* J2315 sessile cells to treatments with high doses of hydrogen peroxide and sodium hypochlorite. *BMC Genomics* **11**.
- Pérez-Rueda, E. & J. Collado-Vides, (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Research* **28**: 1838-1847.
- Phillips, S. E. & M. L. Taylor, (1976) Oxidation of arsenite to arsenate by *Alcaligenes faecalis*. *Applied and Environmental Microbiology* **32**: 392-399.
- Pi, J., H. Yamauchi, Y. Kumagai, G. Sun, T. Yoshida, H. Aikawa, C. Hopenhayn-Rich & N. Shimojo, (2002) Evidence for induction of oxidative stress caused by chronic exposure of Chinese residents to arsenic contained in drinking water. *Environmental Health Perspectives* **110**: 331-336.
- Plumbridge, J., (2002) Regulation of gene expression in the PTS in *Escherichia coli*: The role and interactions of Mlc. *Current Opinion in Microbiology* **5**: 187-193.
- Poelarends, G. J., P. Mazurkiewicz & W. N. Konings, (2002) Multidrug transporters and antibiotic resistance in *Lactococcus lactis*. *Biochimica et Biophysica Acta - Bioenergetics* **1555**: 1-7.
- Porter, S. L. & J. P. Armitage, (2004) Chemotaxis in *Rhodobacter sphaeroides* requires an atypical histidine protein kinase. *Journal of Biological Chemistry* **279**: 54573-54580.

Putrinš, M., H. Ilves, L. Lilje, M. Kivisaar & R. Hörak, (2010) The impact of ColRS two-component system and TtgABC efflux pump on phenol tolerance of *Pseudomonas putida* becomes evident only in growing bacteria. *BMC Microbiology* **10**.

R

Rabilloud, T., (2002) Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics* **2**: 3-10.

Raivio, T. L. & T. J. Silhavy, (2001) Periplasmic stress and ECF sigma factors. In: Annual Review of Microbiology. pp. 591-624.

Rappas, M., D. Bose & X. Zhang, (2007) Bacterial enhancer-binding proteins: unlocking sigma54-dependent gene transcription. *Current Opinion in Structural Biology* **17**: 110-116.

Reid, A. N., R. Pandey, K. Palyada, L. Whitworth, E. Doukhanine & A. Stintzi, (2008) Identification of *Campylobacter jejuni* genes contributing to acid adaptation by transcriptional profiling and genome-wide mutagenesis. *Applied and Environmental Microbiology* **74**: 1598-1612.

Rhine, E. D., E. Garcia-Dominguez, C. D. Phelps & L. Y. Young, (2005) Environmental microbes can speciate and cycle arsenic. *Environmental Science and Technology* **39**: 9569-9573.

Rhine, E. D., C. D. Phelps & L. Y. Young, (2006) Anaerobic arsenite oxidation by novel denitrifying isolates. *Environmental Microbiology* **8**: 899-908.

Rhodium, V. A., W. C. Suh, G. Nonaka, J. West & C. A. Gross, (2006) Conserved and variable functions of the σ^E stress response in related genomes. *PLoS Biology* **4**: 0043-0059.

Riesenfeld, C. S., P. D. Schloss & J. Handelsman, (2004) Metagenomics: Genomic analysis of microbial communities. In: Annual Review of Genetics. pp. 525-552.

Rosen, B. P., (1999) Families of arsenic transporters. *Trends in Microbiology* **7**: 207-212.

Rother, D., H. J. Henrich, A. Quentmeier, F. Bardischewsky & C. G. Friedrich, (2001) Novel genes of the *sox* gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in *Paracoccus pantotrophus* GB17. *Journal of Bacteriology* **183**: 4499-4508.

Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcov \geq n, V. Souza, G. Bonilla-Rosso, L. E. Eguarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Birmingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier & J. C. Venter, (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biology* **5**: 0398-0431.

S

- Salmassi, T. M., K. Venkateswaren, M. Satomi, K. H. Nealson, D. K. Newman & J. G. Hering, (2002) Oxidation of arsenite by *Agrobacterium albertimagni*, AOL15, sp. nov., isolated from Hot Creek, California. *Geomicrobiology Journal* **19**: 53-66.
- Sanders, J. W., K. J. Leenhouts, A. J. Haandrikman, G. Venema & J. Kok, (1995) Stress response in *Lactococcus lactis*: Cloning, expression analysis, and mutation of the lactococcal superoxide dismutase gene. *Journal of Bacteriology* **177**: 5254-5260.
- Santini, J. M., L. I. Sly, R. D. Schnagl & J. M. Macy, (2000) A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: Phylogenetic, physiological, and preliminary biochemical studies. *Applied and Environmental Microbiology* **66**: 92-97.
- Santini, J. M., L. I. Sly, A. Wen, D. Comrie, P. De Wulf-Durand & J. M. Macy, (2002) New arsenite-oxidizing bacteria isolated from Australian gold mining environments - Phylogenetic relationships. *Geomicrobiology Journal* **19**: 67-76.
- Santini, J. M. & R. N. Vanden Hoven, (2004) Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26. *Journal of Bacteriology* **186**: 1614-1619.
- Savel'ev, A. S., L. A. Novikova, I. E. Kovaleva, V. N. Luzikov, W. Neupert & T. Langer, (1998) ATP-dependent proteolysis in mitochondria: m-AAA protease and PIM1 protease exert overlapping substrate specificities and cooperate with the mtHsp70 system. *Journal of Biological Chemistry* **273**: 20596-20602.
- Schröder, I. D., S. Gunsalus, RP., (1993) Activation of the *Escherichia coli* nitrate reductase (*narGHJI*) operon by NarL and Fnr requires integration host factor. *Journal of Biological Chemistry* **15**: 771-774.
- Shi, J., A. Vlamis-Gardikas, F. Öslund, A. Holmgren & B. P. Rosen, (1999) Reactivity of glutaredoxins 1, 2, and 3 from *Escherichia coli* shows that glutaredoxin 2 is the primary hydrogen donor to ArsC-catalyzed arsenate reduction. *Journal of Biological Chemistry* **274**: 36039-36042.
- Shin, M., S. Kang, S. J. Hyun, N. Fujita, A. Ishihama, P. Valentin-Hansen & H. E. Choy, (2001) Repression of *deoP2* in *Escherichia coli* by CytR: Conversion of a transcription activator into a repressor. *EMBO Journal* **20**: 5392-5399.
- Shingler, V., (1996) Signal sensing by σ_{54} -dependent regulators: Derepression as a control mechanism. *Molecular Microbiology* **19**: 409-416.
- Silver, S. & L. T. Phung, (1996) Bacterial heavy metal resistance: New surprises. In: Annual Review of Microbiology. pp. 753-789.
- Silver, S. & L. T. Phung, (2005) A bacterial view of the periodic table: Genes and proteins for toxic inorganic ions. *Journal of Industrial Microbiology and Biotechnology* **32**: 587-605.

- Silver, S., T. Phung Le and B. P. Rosen, (2002) Arsenic metabolism : Resistance, Reduction and oxidation in T.F. Williams. *Environmental Chemistry of arsenic* Marcel Dekker, Inc, New-York; Basel.
- Skovierova, H., G. Rowley, B. Rezuchova, D. Homerova, C. Lewis, M. Roberts & J. Kormanec, (2006) Identification of the σ E regulon of *Salmonella enterica* serovar Typhimurium. *Microbiology* **152**: 1347-1359.
- Soutourina, O. A. & P. N. Bertin, (2003) Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiology Reviews* **27**: 505-523.
- Srinivasan, V. B., G. Rajamohan & W. A. Gebreyes, (2009) Role of AbeS, a novel efflux pump of the SMR family of transporters, in resistance to antimicrobial agents in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy* **53**: 5312-5316.
- Stadtman, E. R., (2004) Cyclic oxidation and reduction of methionine residues of proteins in antioxidant defense and cellular regulation. *Archives of Biochemistry and Biophysics* **423**: 2-5.
- Stock, J. B., A. J. Ninfa & A. M. Stock, (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiological Reviews* **53**: 450-490.
- Stock, J. B., A. M. Stock & J. M. Mottonen, (1990) Signal transduction in bacteria. *Nature* **344**: 395-400.
- Storz, G. & J. A. Imlay, (1999) Oxidative stress. *Current Opinion in Microbiology* **2**: 188-194.
- Streit, W. R. & R. A. Schmitz, (2004) Metagenomics - The key to the uncultured microbes. *Current Opinion in Microbiology* **7**: 492-498.
- Studholme, D. J. & R. Dixon, (2004) In silico analysis of the σ 54-dependent enhancer-binding proteins in *Pirellula* species strain 1. *FEMS Microbiology Letters* **230**: 215-225.
- Sullivan, J. T., H. N. Patrick, W. L. Lowther, D. B. Scott & C. W. Ronson, (1995) Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 8985-8989.

T

- Tan, K., S. Clancy, M. Borovilos, M. Zhou, S. Hörer, S. Moy, L. L. Volkart, J. Sassoon, U. Baumann & A. Joachimiak, (2009) The mannitol operon repressor MtlR belongs to a new class of transcription regulators in bacteria. *Journal of Biological Chemistry* **284**: 36670-36679.
- Taylor, M., R. Butler, S. Chambers, M. Casimiro, F. Badii & M. Merrick, (1996) The RpoN-box motif of the RNA polymerase sigma factor σ (N) plays a role in promoter recognition. *Molecular Microbiology* **22**: 1045-1054.

- Tebbutt, J., V. A. Rhodius, C. L. Webster & S. J. W. Busby, (2002) Architectural requirements for optimal activation by tandem CRP molecules at a class I CRP-dependent promoter. *FEMS Microbiology Letters* **210**: 55-60.
- Teitzel, G. M. & M. R. Parsek, (2003) Heavy metal resistance of biofilm and planktonic *Pseudomonas aeruginosa*. *Applied and Environmental Microbiology* **69**: 2313-2320.
- Tian, T., X. G. Wu, H. M. Duan & L. Q. Zhang, (2010) The resistance-nodulation-division efflux pump EmhABC influences the production of 2,4-diacetylphloroglucinol in *Pseudomonas fluorescens* 2P24. *Microbiology* **156**: 39-48.
- Tseng, T. T., K. S. Gratwick, J. Kollman, D. Park, D. H. Nies, A. Goffeau & M. H. Saier Jr, (1999) The RND permease superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *Journal of Molecular Microbiology and Biotechnology* **1**: 107-125.
- Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar & J. F. Banfield, (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.

U/V/W

- Unlü, M., M. E. Morgan & J. S. Minden, (1997) Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* **18**: 2071-2077.
- Van der Veen, S., S. van Schalkwijk, D. Molenaar, W. M. de Vos, T. Abee & M. H. J. Wells-Bennik, (2010) The SOS response of *Listeria monocytogenes* is involved in stress resistance and mutagenesis. *Microbiology* **156**: 374-384.
- Vaughan, D. J., (2006) Arsenic. *Elements* **2**: 71-75.
- Villaescusa, I. & J. C. Bollinger, (2008) Arsenic in drinking water: Sources, occurrence and health effects (a review). *Re-views in Environmental Science and Biotechnology* **7**: 307-323.
- Vogel, J. & K. Papenfort, (2006) Small non-coding RNAs and the bacterial outer membrane. *Current Opinion in Microbiology* **9**: 605-611.
- Wade, J. T., T. A. Belyaeva, E. I. Hyde & S. J. W. Busby, (2002) A simple mechanism for co-dependence on two activators at an *Escherichia coli* promoter. *EMBO Journal* **20**: 7160-7167.
- Wadhams, G. H. & J. P. Armitage, (2004) Making sense of it all: Bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* **5**: 1024-1037.
- Wadhams, G. H., A. C. Martin & J. P. Armitage, (2000) Identification and localization of a methyl-accepting chemotaxis protein in *Rhodobacter sphaeroides*. *Molecular Microbiology* **36**: 1222-1233.

- Wadhams, G. H., A. V. Warren, A. C. Martin & J. P. Armitage, (2003) Targeting of two signal transduction pathways to different regions of the bacterial cell. *Molecular Microbiology* **50**: 763-770.
- Wai, S. N., Y. Mizunoe, A. Takade, S.-I. Kawabata & S.-I. Yoshida, (1998) *Vibrio cholerae* O1 Strain TSI-4 Produces the Exopolysaccharide Materials That Determine Colony Morphology, Stress Resistance, and Biofilm Formation. *Applied and Environmental Microbiology* **64**: 3648-3655.
- Wang, L., B. Jeon, O. Sahin & Q. Zhang, (2009) Identification of an arsenic resistance and arsenic-sensing system in *Campylobacter jejuni*. *Applied and Environmental Microbiology* **75**: 5064-5073.
- Weiss, S., C. Carapito, J. Cleiss, S. Koechler, E. Turlin, J. Y. Coppee, M. Heymann, V. Kugler, M. Stauffert, S. Cruveiller, C. Médigue, A. Van Dorsselaer, P. N. Bertin & F. Arsène-Ploetze, (2009) Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data. *Biochimie* **91**: 192-203.
- Wodara, C., F. Bardischewsky & C. G. Friedrich, (1997) Cloning and characterization of sulfite dehydrogenase, two c-type cytochromes, and a flavoprotein of *Paracoccus denitrificans* GB17: Essential role of sulfite dehydrogenase in lithotrophic sulfur oxidation. *Journal of Bacteriology* **179**: 5014-5023.
- Wösten, M. M. S. M., (1998) Eubacterial sigma-factors. *FEMS Microbiology Reviews* **22**: 127-150.
- Wu, J. & B. P. Rosen, (1993) Metalloregulated expression of the *ars* operon. *Journal of Biological Chemistry* **268**: 52-58.
- Wysocki, R., P. Bobrowicz & S. Ulaszewski, (1997) The *Saccharomyces cerevisiae* ACR3 gene encodes a putative membrane protein involved in arsenite transport. *Journal of Biological Chemistry* **272**: 30061-30066.
- X/Y/Z**
- Xu, C., W. Shi & B. P. Rosen, (1996a) The chromosomal *arsR* gene of *Escherichia coli* encodes a trans-acting metalloregulatory protein. *Journal of Biological Chemistry* **271**: 2427-2432.
- Xu, C., W. Shi & B. P. Rosen, (1996b) Identification of the As(III)/Sb(III) binding domain and the dimerization of the ArsR repressor. *FASEB Journal* **10**.
- Yang, J., S. Rawat, T. L. Stemmler & B. P. Rosen, (2010) Arsenic binding and transfer by the ArsD As(III) metallochaperone. *Biochemistry* **49**: 3658-3666.
- Ye, J., H. C. Yang, B. P. Rosen & H. Bhattacharjee, (2007) Crystal structure of the flavoprotein ArsH from *Sinorhizobium meliloti*. *FEBS Letters* **581**: 3996-4000.
- Yorimitsu, T. & M. Homma, (2001) Na⁺-driven flagellar motor of *Vibrio*. *Biochimica et Biophysica Acta - Bioenergetics* **1505**: 82-93.

- Young, K. D., (2006) The selective value of bacterial shape. *Microbiology and Molecular Biology Reviews* **70**: 660-703.
- Yura, T., H. Nagai & H. Mori, (1993) Regulation of the heat-shock response in bacteria. *Annual Review of Microbiology* **47**: 321-350.
- Zhang, G., E. A. Campbell, L. Minakhin, C. Richter, K. Severinov & S. A. Darst, (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**: 811-824.
- Zhu, L., Q. Peng, F. Song, Y. Jiang, C. Sun, J. Zhang & D. Huang, (2010) Structure and regulation of the *gab* gene cluster, involved in the γ -aminobutyric acid shunt, are controlled by a sigma54 factor in *Bacillus thuringiensis*. *Journal of Bacteriology* **192**: 346-355.

Réponse bactérienne au stress arsénié

L'arsenic est un métalloïde largement distribué dans l'environnement. Sous l'action des micro-organismes, cet élément peut passer d'une forme oxydée à une forme réduite et inversement chez certaines bactéries. *Herminiimonas arsenicoxydans* est une bactérie hétérotrophe qui est capable de réduire l'arséniate (As(V)) et d'oxyder l'arsénite (As(III)). La réduction de l'As(V) a été largement étudiée ces dernières années, alors que l'oxydation de l'As(III) catalysée par l'arsénite oxydase (Aox) est un processus qui a fait l'objet de peu d'études.

Ce travail de thèse a eu pour objectif principal l'étude de stress arsénié chez les bactéries, en particulier la régulation de l'arsénite oxydase. L'analyse de mutants obtenus par transposition aléatoire d'un mini-Tn5 a conduit à l'identification des gènes dont les produits jouent un rôle dans le contrôle de l'arsénite oxydase : *aoxA*, *aoxB*, *aoxRS*, *rpoN* et *dnaJ*. En parallèle la réponse globale au stress arsénié a été étudiée. Nous avons ainsi montrés que les gènes et protéines régulés par l'arsenic interviennent dans une grande variété de fonctions telles que le transport du phosphate, la motilité, la résistance à l'arsenic *via ars*, le stress oxydatif... Néanmoins, l'arsénite oxydase n'a pas été identifiée lors de ces travaux. Aussi, la cinétique de l'oxydation et les mécanismes de régulation sous-jacents ont été étudiées.

Ces travaux ont ensuite été étendus à des bactéries chimiolithoautotrophes telles que *Thiomonas* sp. et *Rhizobium* sp. NT-26. L'As(III) peut servir de donneur d'électrons chez ces micro-organismes et son oxydation est aussi catalysée par l'arsénite oxydase. L'analyse de la réponse globale au stress arsénié semble indiquer que certains mécanismes de réponses sont conservés tels que la motilité, le transport de phosphate, la résistance à l'arsenic *via ars*. Néanmoins, des spécificités ont pu être observées telles que la biosynthèse d'acide gras, des transporteurs ABC... De la même manière, l'analyse des mutants a permis de valider le rôle de AroR (AoxR) dans la régulation de l'arsénite oxydase chez une bactérie chimiolithoautotrophe et d'autres régulateurs ont également été identifiés : PtxB et MoeB.

Les approches futures devraient permettre de préciser le rôle de ces différents gènes dans la régulation de l'arsénite oxydase chez les bactéries hétérotrophes et chimiolithoautotrophes et ainsi permettre d'appréhender l'implication de cette enzyme dans le métabolisme énergétique.

Mots-clés : stress arsénié, oxydation de l'arsenic, arsénite oxydase, régulation, transposon