



Origine(s) et Fonction(s) de Gènes de Résistance aux Métaux Issus de Métatranscriptomes Eucaryotes de Sols

Antoine Ziller

► To cite this version:

Antoine Ziller. Origine(s) et Fonction(s) de Gènes de Résistance aux Métaux Issus de Métatranscriptomes Eucaryotes de Sols. Ecologie, Environnement. Université de Lyon, 2017. Français. NNT : 2017LYSE1056 . tel-01540853

HAL Id: tel-01540853

<https://theses.hal.science/tel-01540853>

Submitted on 16 Jun 2017

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

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



N°d'ordre NNT : 2017LYSE1056

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 341
Evolution, Ecosystèmes, Microbiologie, Modélisation

Spécialité de doctorat : Biologie Moléculaire, Ecologie Microbienne
et Ecotoxicologie Microbienne
Discipline : Biologie

Soutenue publiquement le 31/03/2017, par :
Antoine, Ysdine, Jean ZILLER

Origine(s) et Fonction(s) de Gènes de Résistance aux Métaux Issus de Métatranscriptomes Eucaryotes de Sols

Devant le jury composé de :

Colpaert, Jan	Professeur des Universités	Hasselt University	Rapporteur
Marques, Laurence	Maître de Conférences	Université de Montpellier	Rapporteuse
Chalot, Michel	Professeur des Universités	Université de Franche-Comté	Examineur
Vogel, Timothy	Professeur des Universités	Université Claude Bernard Lyon 1	Président
Fraissinet-Tachet, Laurence	Maitre de Conférences	Université Claude Bernard Lyon 1	Directrice de thèse



UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

M. le Professeur Frédéric FLEURY

Président du Conseil Académique

M. le Professeur Hamda BEN HADID

Vice-président du Conseil
d'Administration

M. le Professeur Didier REVEL

Vice-président du Conseil Formation et
Vie Universitaire

M. le Professeur Philippe CHEVALIER

Vice-président de la Commission
Recherche

M. Fabrice VALLÉE

Directrice Générale des Services

Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude
Bernard

Directeur : M. le Professeur G.RODE

Faculté de Médecine et de Maïeutique
Lyon Sud – Charles Mérieux

Directeur : Mme la Professeure C. BURILLON

Faculté d'Odontologie

Directeur : M. le Professeur D. BOURGEOIS

Institut des Sciences Pharmaceutiques et
Biologiques

Directeur : Mme la Professeure C.
VINCIGUERRA

Institut des Sciences et Techniques de la
Réadaptation

Directeur : M. X. PERROT

Département de formation et Centre de
Recherche en Biologie Humaine

Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologie	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y. VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E.PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieur de Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE





« Donne un poisson à un homme, et tu le nouriras une journée. Apprends lui à pêcher et tu le nouriras toute sa vie. »

Lao-Tseu

« Un humanisme bien ordonné ne commence pas par soi-même mais place le monde avant la vie, la vie avant l'homme, le respect des autres êtres avant l'amour propre »

Claude Lévy Strauss

« Une théorie, même si elle conditionne le regard, même si elle élabore les données, n'a jamais le poids, le sillage, la traîne d'un paradigme. [...] on peut changer très vite de théorie puisque rien d'autre ne s'attache à elle ; on ne change pas de paradigme. La première est fluide, le second visqueux. »

Bruno Latour





TABLE DES MATIERES

Liste des abréviations	1
Introduction générale.....	4
Chapitre 1. Synthèse bibliographique et contexte	10
I. L'écosystème édaphique	10
A. Les services écosystémiques du sol.....	10
B. Le sol, un biotope complexe abritant une riche biodiversité principalement microscopique	12
C. Importance des microorganismes dans les fonctions écosystémiques des sols	15
II. Les pollutions métalliques des sols	16
A. Les métaux, propriétés chimiques.....	18
B. Origine des métaux dans les sols	20
C. Biodisponibilité des métaux dans les sols	22
III. Réponses cellulaires aux stress métalliques	24
A. Modifications de l'homéostasie métallique des cellules et perturbations des fonctions métaboliques	24
B. Les mécanismes de " résistance " aux métaux.....	34
IV. Les métallothionéines	40
A. Structures des métallothionéines	42
1. Structure primaire	42
2. Structures secondaire et tertiaire	48
B. Diversité et distribution des métallothionéines	52
C. Diversité des formes des métallothionéines.....	60
D. Classifications des métallothionéines	66
V. Discussion et enjeux.....	68
A. La diversité microbienne des sols	68
B. La pollution métallique des sols.....	69
C. Les perturbations métaboliques induites par les métaux	71
D. Les métallothionéines.....	72

VI. Contexte et objectifs de la thèse	76
Chapitre 2. Les CRP sont-elles des métallothionéines ?	80
Introduction.....	80
Article publié dans <i>Journal of Inorganic Biochemistry</i>	84
Données supplémentaires	98
Conclusion.....	98
Chapitre 3. Mise au point de la quantification des gènes emt dans les sols.....	99
Introduction.....	99
Matériels et méthodes.....	105
Sol étudié et préparation des microcosmes.....	105
Extraction et quantification des métaux	106
Co-extraction des ADN et des ARN	107
Mise au point des amorces et sondes	109
Préparation du standard d'ADN plasmidique pour les réactions de PCR	109
Réactions d'amplification des ADN par PCR, qPCR et ddPCR.....	110
Analyses statistiques des données	112
Résultats et discussion.....	114
Réalisation de microcosmes de sol amendés en éléments métalliques	115
Co-extraction d'ADN et d'ARN	119
Mise au point de la PCR	119
Amplifications par PCR quantitative : comparaison de la qPCR en temps réel et de la ddPCR	125
Conclusion et perspectives.....	130
Chapitre 4. Recherche des gènes liés aux séquences emt par une approche de capture ciblée	133
Introduction.....	133
Matériels et méthodes.....	141
Site d'étude et extraction d'ADN à partir d'échantillons de sol.....	141
Conception et synthèse des sondes de capture.....	142
Préparation des extraits d'ADN de sol.....	144
Hybridation en solution	145
Clonage et séquençage des produits de capture	146

Résultats	149
Analyse du pool d'ADN fragmenté	151
Analyse de l'enrichissement en séquences cibles.....	152
Clonage et séquençage des fragments d'ADN issus de la capture B	156
Discussion et perspectives.....	158
Un nombre important de produits de capture aspécifiques ?	159
Des produits de capture de trop petite taille ?	162
Perspectives pour obtenir les régions flanquant les gènes <i>emt</i>	163
Conclusions générales et perspectives de la thèse	165
Les CRP peuvent-elles être considérées comme des métallothionéines ?.....	166
L'expression des gènes <i>emt</i> est-elle induite par des métaux ?	168
De quel(s) groupe(s) d'organisme(s) sont issus les gènes codant les EMT ?	169
Améliorer les connaissances relatives à la diversité des métallothionéines en s'affranchissant des analyses de similarité de séquences.....	171
Pour une synthèse entre la génomique environnementale et la biologie systémique.....	173
Remerciements	176
Liste des figures et tableaux	180
Annexes.....	182
Annexes IV.1 : Séquences et adaptateurs	182
Annexes IV.2 : Amorces et PCR	184
Annexes IV.3 : Solutions préparées pour la capture ciblée.....	186
Références Bibliographiques.....	188
Résumé.....	197





LISTE DES ABREVIATIONS

Signes	
%	Pourcent
°	Degré
°C	Degré Celsius
α	Alpha
β	Béta
γ	Gamma
μ -	Micro
3D	3 dimensions
e.g.	<i>Exempli gratta</i> (Par exemple)
et al.	<i>Et alii</i> (et toutes les autres personnes)
-g	Gramme
k-	Kilo
kb	Kilo base
-L	Litre
LB	Luria-Bertani
-m	Mètre
m-	Milli
-M	Molaire
n-	Nano
sp.	Espèce
spp.	Toutes les espèces du genre
Taq	<i>Thermophilus aquaticus</i>
Tm	<i>Melting temperature</i>
h	Heures
min	Minutes
sec	Secondes
vol	Volume
g	Accélération (9,80665 m.s ⁻²)
Da	Dalton. Unité de masse des atomes et molécules
U	Unité
ppm	Partie par million
p./pp.	Page/pages

Institutions	
ADEME	Agence De l'Environnement et de la Maîtrise de l'Energie
CNRS	Centre National de la Recherche Scientifique
Dr.	Docteur
INRA	Institut National de la Recherche Agronomique
LEM	Laboratoire d'Ecologie Microbienne de Lyon
NCBI	<i>National Center for Biotechnology Information</i>
OCDE	Organisation de Coopération et de Développement Economique
UMR	Unité Mixte de Recherche

Outils techniques	
nr/nt	Base de données NCBI <i>nucleotide collection</i> de séquences de référence
BLASTn	<i>Basic Local Alignment Search Tool nucleotide</i>
ESI-MS	<i>Electrospray Ionization -Mass Spectrometry</i>
ICP-OES	<i>Inductively Coupled Plasma -Optical Emission Spectrometry</i>
PCR	<i>Polymerase Chain Reaction</i>
ddPCR	<i>droplet digital PCR</i> (PCR numérique en gouttelettes)
qPCR	<i>Real time quantitative PCR</i> (PCR en temps réel)
UV-Vis	Ultraviolet-Visible

Biochimie-Biologie moléculaire			
(aq)	Phase aqueuse	HNO₃	Acide nitrique
(org)	Phase organique	HPO₄	Ion hydrogénophosphate
(s)	Phase solide	Me	Métaux
ADN	Acide désoxyribonucléique	MT	Métallothionéines
ADNc	ADN complémentaire	MTF	<i>Metal regulatory Transcription Factor</i>
ARN	Acide Ribonucléique	pH	Potentiel hydrogène
ARNm	ARN messenger	ROS	<i>Reactive Oxygen Species</i>
Ca(NO₃)₂	Nitrate de calcium	SDS	Sodium Dodecyl Sulfate
CdCl₂	Chlorure de cadmium	Ser/S	Sérine
CopA	<i>Copper-exporting P-type ATPase A</i>	SOD	Superoxyde dismutase
CRP	<i>Cystein Rich Protein</i>	SSC	<i>Saline-Sodium Citrate</i>
CuSO₄	Sulfate de cuivre	SSPE	<i>Saline-Sodium Phosphate-EDTA</i>
Cys/C	Cystéine	T	Thionéines
EDTA	Acide éthylène diamine tétraacétique	TO	Thionines
EMT	<i>Environmental Metallothionein</i>	UTR	<i>Untranslated Region</i>
Glu/E	Glutamine	ZnSO₄	Sulfate de Zinc
Gly/G	Glycine	Oligonucléotide F	<i>Forward</i>
GSH	Glutathion (réduit)	Oligonucléotide R	Reverse
GSSH	Glutathion (oxydé)	Oligonucléotide P	<i>Probe</i>
HCl	Acide chlorhydrique	Not I	Enzyme de restriction isolée de <i>Nocardia otitidis</i>
His/H	Histidine	BSA	Albumine sérique bovine

Atomes			
Ac	Actinium	Mn	Manganèse
As	Arsenic	Mo	Molybdène
At	Astate	N	Azote
Ba	Baryum	Na	Sodium
Be	Béryllium	Ni	Nickel
Br	Brome	O	Oxygène
C	Carbone	P	Phosphore
Ca	Calcium	Pb	Plomb
Cd	Cadmium	S	Soufre
Ce	Cérium	Sb	Antimoine
Cl	Chlore	Sr	Strontium
Co	Cobalt	Te	Tellure
Cr	Chrome	Th	Thorium
Cu	Cuivre	Tl	Thallium
F	Fluore	V	Vanadium
Fe	Fer	W	Tungstène
Hg	Mercure	Zn	Zinc
I	Iode	Pb	Plomb
La	Lanthane	Al	Aluminium
Li	Lithium	Se	Sélénium
Mg	Magnésium		



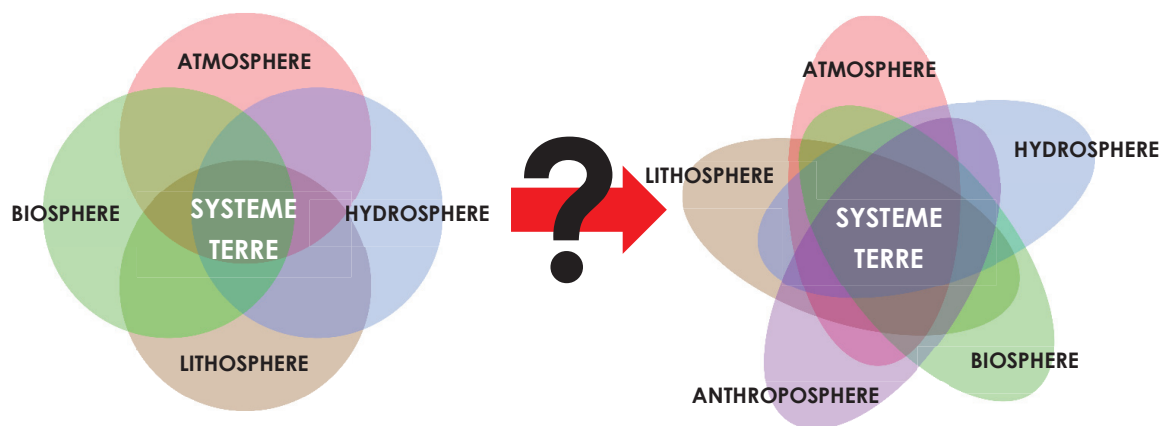


Figure A : Diagrammes conceptuels du système Terre. D'un point de vue fonctionnel, le système Terre peut être découpé en sous-systèmes qui sont l'atmosphère, la lithosphère, l'hydrosphère, la biosphère et récemment l'anthroposphère.



INTRODUCTION GENERALE

Durant l'Holocène (de -10 000 ans au 19^{ème} siècle), le système complexe qu'est la planète Terre a été marqué par une longue période de stabilité environnementale (**Biraben 2003 ; Rockström et al. 2009**). Cette stabilité reposait sur des processus physiques et chimiques imbriqués qu'il est difficile de comprendre sans prendre en compte les dimensions historiques et structurales des différents sous-systèmes terrestres (**Figure A**). Depuis plusieurs années, nous constatons que cette stabilité est menacée par les activités humaines qui exercent une pression croissante et continue sur les milieux. Si les choses restent telles qu'elles le sont actuellement, à l'échelle de seulement quelques générations humaines, le « système Terre » pourrait basculer vers un nouvel état dont les caractéristiques pourraient entraîner certaines sociétés humaines vers un effondrement (**Rockström et al. 2009**). Nous passons ainsi de l'Holocène, ère stable qui a vu naître l'Homme, à l'Anthropocène, ère instable marquant l'avènement de l'Homme en tant que force de changements globaux (« anthroposphère ») (**Biraben 2003 ; Rockström et al. 2009 ; Zalasiewicz et al. 2015**).

D'un point de vue fonctionnel, le « système Terre » peut être décomposé en quatre sous-systèmes ou compartiments principaux (**Figure A**) : l'atmosphère, la lithosphère, l'hydrosphère et la biosphère. Cette dernière, faisant référence au « domaine du vivant », est composée d'une diversité très importante d'organismes qui interagissent entre eux à différents niveaux d'organisation spatiale et temporelle. Cette biodiversité est très importante pour le fonctionnement du système Terre car elle permet d'une part, le transfert de matières et d'énergies entre l'hydrosphère, la lithosphère et l'atmosphère, en étant moteur de flux biogéochimiques ; et d'autre part, elle permet le maintien dans le temps des écosystèmes face aux multiples perturbations par « un effet tampon » (**Rockström et al. 2009 ; Olivier et al. 2015**). Par sa position



centrale au sein du système Terre, sa complexité et son hétérogénéité, la pédosphère (les sols) abrite une riche biodiversité qui est principalement microscopique et inconnue. Cette biodiversité des sols joue des rôles importants pour les sociétés humaines, notamment en assurant la nutrition minérale des plantes ou en recyclant les déchets organiques (**Adhikari & Hartmink 2016**). Grâce à leurs diversités spécifique, métabolique et *a fortiori* fonctionnelle, les microorganismes des sols sont les principaux moteurs de certains flux de matières et d'énergies à l'échelle globale (**Falkowski et al. 2008**) comme par exemple, les flux de soufre (oxydo-réduction du soufre), d'azote (fixation du diazote atmosphérique) ou de métaux (oxydo-réduction du fer ou du manganèse,...) (**Staley & Orrians 2000**). Leurs fonctions au sein des écosystèmes sont assez bien connues mais les connaissances concernant leur inventaire, la dynamique des successions microbiennes dans l'environnement, l'organisation des relations structurant les communautés, ainsi que leur histoire évolutive restent encore très lacunaires.

Cette biodiversité est aussi menacée par les grands changements de l'Anthropocène à un tel point que l'on parle de la 6^{ème} plus grande extinction de masse d'espèces dans l'histoire de la vie (**Barnosky et al. 2011**). Par exemple, les pollutions chimiques, diffuses et chroniques, pourraient être une des causes majeures de la diminution de la biodiversité (**Rockström et al. 2009, Amorim et al. 2016**). Cependant l'importance des pollutions chimiques sur la diminution de la biodiversité reste très compliquée à démontrer étant donnée la complexité des relations entre d'un côté un polluant donné et tous les autres composés présents dans l'environnement qui vont conditionner sa toxicité pour un individu et d'un autre côté, un « individu » interagissant au sein d'une communauté complexe d'organismes qui va rencontrer ce polluant (**Ramade 1992, pp.45-49**). L'extraction et l'utilisation d'éléments métalliques est une des plus anciennes perturbations d'origine humaine ayant débuté dès l'Antiquité. Elle est aussi l'une des plus difficiles à gérer car les éléments métalliques ne sont pas biodégradables et peuvent s'accumuler dans l'environnement pendant



des centaines d'années à des millénaires constituant une pression de sélection à toutes les échelles spatiales et temporelles (**Ramade 1992, pp.45-49**). En ce qui concerne les microorganismes, les phénomènes de stress métalliques vont avoir différents effets en fonction de la dose (létale ou sub-létale), de la fréquence (stress aigu ou chronique) et de « l'état » de l'organisme (stade de développement, assemblage des communautés, capacités physiologiques et comportementales) (**Piggott et al. 2015**). Bien que certains métaux présents à des concentrations physiologiques soient essentiels au bon fonctionnement de la cellule, sous certaines conditions les métaux peuvent avoir un effet délétère au niveau moléculaire en altérant les molécules organiques ce qui va perturber les métabolismes et à l'échelle cellulaire pouvant aller jusqu'à la mort de la cellule. Cette activité va se répercuter sur les autres niveaux d'organisation : à l'échelle individuelle par modification de la valeur sélective, à l'échelle populationnelle par modification des paramètres démographiques, à l'échelle des communautés par modification des réseaux d'interactions, à l'échelle de l'écosystème par modification de la productivité et de la stabilité des processus et à l'échelle globale en perturbant les flux de matières et d'énergies (**Ramade 1992, p.84**). Cependant, le stress métallique ne date pas de l'Anthropocène. Les premières formes de vie, il y a environ 3,8 milliards d'années, étaient déjà confrontées à ce type de stress et ont même utilisé les propriétés chimiques des métaux à leur avantage en développant toute une diversité de métalloprotéines (**Yannone et al. 2012**). Les organismes ont donc depuis très longtemps mis en place des mécanismes moléculaires permettant la gestion des « stocks » intracellulaires de métaux, d'acquisition de ceux-ci en cas de carence et de contrôle du stress métallique en cas d'excès qui leur ont permis de perdurer et de se propager sur toute la planète (**Merchant & Helmann 2012 ; Scott et al. 2013**). Les mécanismes de résistance aux métaux passent par la modification du métabolisme par exemple en augmentant le pool intracellulaire de molécules (métabolites ou protéines) chélatrices de



métaux limitant les effets toxiques des métaux réagissant chimiquement avec les molécules organiques.

Que ce soit en écologie microbienne ou en écotoxicologie microbienne, l'étude et l'observation des microorganismes *in situ* est souvent plus difficile d'un point de vue méthodologique (difficultés d'isolement, non cultivabilité,...) que l'étude des molécules organiques dont ils sont composés (acides nucléiques, protéines, lipides). Aussi, beaucoup d'études en écologie et écotoxicologie microbienne sont depuis plusieurs années basées sur des méthodes de biologie/écologie moléculaire et de biochimie qui permettent indirectement d'étudier les communautés microbiennes des sols, sédiments et milieux aquatiques.

Au sein de mon laboratoire d'accueil, une approche de génomique environnementale (métatranscriptomique fonctionnelle), qui consiste à rechercher de nouveaux gènes de fonction par complémentation fonctionnelle de levures sensibles aux métaux, a été mise en œuvre. Cette approche a notamment permis d'isoler à partir de sols une famille de gènes d'origine eucaryote codant des protéines impliquées dans la résistance cellulaire au stress métallique (**Lehembre et al. 2013**). Ces protéines sont sans homologues dans les bases de données (probablement issues d'organismes d'origine taxonomique inconnue). Comme leurs séquences en acides aminés sont riches en résidus cystéines conservées, elles ont été nommées CRP pour « *Cysteine-Rich Proteins* ». Cette famille semble être apparentée à une famille de chélateurs de métaux intracellulaires bien connus, les métallothionéines. Mon projet de recherche de doctorat intitulé « Origine(s) et fonction(s) de gènes de résistance aux métaux issus de métatranscriptomes eucaryotes de sols » a pour objectif de déterminer si ces séquences inconnues sont des métallothionéines ou appartiennent à une nouvelle classe de protéines. Ce projet a été réalisé dans un triple contexte d'écologie microbienne, d'écotoxicologie microbienne et de biologie moléculaire. Il combine des questionnements sur les mécanismes de résistance aux stress métalliques, la



nature des interactions entre molécules organiques et métaux et les méthodes d'étude des communautés microbiennes au travers de 5 chapitres.

Chapitre 1. Synthèse Bibliographique et Contexte. Cette synthèse bibliographique porte sur l'écosystème édaphique, les pollutions métalliques des sols, les conséquences des pollutions métalliques sur les systèmes biologiques et enfin plus spécifiquement sur les métallothionéines. A la fin de ce chapitre, sont présentés le contexte de ce travail et les objectifs de ma thèse.

Chapitre 2. Les CRP sont-elles des métallothionéines ? Ici sont regroupés les résultats de la comparaison des CRP avec les métallothionéines connues via des analyses bio-informatiques, microbiologiques et biochimiques. Les résultats de ce chapitre ont fait l'objet d'un article publié dans *Journal of Inorganic Biochemistry*. Dans ce chapitre, les protéines CRP ont été renommées EMT pour « *environmental metallothioneins* ».

Chapitre 3. Mise au point de la quantification des gènes *emt* dans les sols. Dans ce chapitre, sont présentées la réalisation de microcosmes de sol pollués par des éléments métalliques et la mise au point d'une quantification des gènes *emt* par deux méthodes de PCR quantitative afin de pouvoir à terme quantifier l'abondance relative de leurs transcrits dans des sols amendés en métaux.

Chapitre 4. Recherche des gènes associés aux séquences *emt* par une approche de capture ciblée. Dans ce chapitre, après une brève introduction sur les méthodes disponibles pour obtenir des régions bordant un gène, sont détaillées les différentes mises au point d'une méthode de capture de gènes réalisées à partir d'extraits d'ADN environnementaux.

Chapitre 5. Conclusions générales et perspectives de la thèse.





CHAPITRE 1

Synthèse bibliographique et contexte

I. L'écosystème édaphique

Dans le langage courant, le « sol » est synonyme de « support », dans le sens où « c'est ce sur quoi on marche ». Le Larousse le définit comme la « couche superficielle de l'écorce d'une planète tellurique » ou encore comme une « surface, en générale plane et horizontale [...] ». Le sens du mot « sol » va dépendre des disciplines scientifiques qui l'étudient. L'écologue y verra un élément du biotope faisant partie de l'écosystème alors que le géologue y verra la partie superficielle de la lithosphère. Dans ce manuscrit, nous considérons " le sol " comme « la couche la plus externe de la croûte terrestre, composée de roches, d'air, d'eau et d'êtres vivants, siège d'échanges intenses de matière et d'énergie entre l'air, l'eau et les roches. Le sol, en tant que partie de l'écosystème, occupe une position-clé dans [les cycles biogéochimiques] » (**Gobat et al. 2010, p.11**).

Les deux paragraphes ci-dessous vont illustrer cette définition du sol en mettant l'accent sur le sol en tant que système écologique complexe ayant une place centrale dans le fonctionnement général des écosystèmes et le sol en tant qu'habitat d'une biodiversité riche et principalement microscopique.

A. Les services écosystémiques du sol

Un sol est un système écologique dont la complexité et le dynamisme reposent en grande partie sur sa position charnière au sein du système Terre, générant et contrôlant les flux de matières et d'énergies entre la géosphère, la biosphère, l'atmosphère et l'hydrosphère. En effet, il est à l'interface entre



l'atmosphère et la biosphère, intervenant par exemple dans la fixation de l'azote (**Galloway et al. 2004**) et du carbone atmosphérique (**Schimel 1995**). Il est aussi à l'interface entre la géosphère et la biosphère, intervenant dans le transfert d'éléments des minéraux vers les organismes vivants (**Falkowski et al. 2008**). Le sol remplit 4 services écosystémiques principaux (**De Groot et al. 2002 ; Adhikan & Hartemink 2016**) :

- **Le sol est un habitat** permettant aux organismes de se développer, de vivre et de se multiplier en tamponnant par exemple les fluctuations de températures.
- **Le sol régule les processus écologiques** tels que le climat via l'humidité générée par le couvert végétal ou encore l'épuration des eaux et des déchets organiques via les activités biochimiques des organismes qu'il contient.
- **Le sol est un producteur** par exemple de nourriture à la base des réseaux trophiques terrestres (minéraux assimilables par les végétaux) ou de matières premières essentielles aux systèmes sociaux et économiques.
- **Le sol est source de bien-être** notamment esthétique, récréatif et/ou culturel.

Ces services vont être en grande partie dépendantes des propriétés biophysicochimiques du sol dont une large part est sous l'influence de la biodiversité présente dans les sols et en particulier la diversité des microorganismes (**Montoya et al. 2006 ; Falkowski et al. 2008**).



B. Le sol, un biotope complexe abritant une riche biodiversité principalement microscopique

Le sol est un biotope de nature triphasique : solide (matière minérale, matière organique vivante ou morte), liquide (solution du sol) et gazeuse (air du sol) (**Gobat et al. 2010, p.5**). La répartition de tous les éléments constitutifs du sol est très hétérogène à la fois dans un volume de sol considéré et dans le temps selon de multiples niveaux d'organisation donnant une grande diversité de types de sol. Nous allons pouvoir trouver par exemple des spodosols dans les régions froides et humides et des mollisols dans les régions tempérées (**Ehrlich 2016**). Les types de sol rencontrés vont être conditionnés par la localisation géographique et la nature des éléments qui les composent (géodiversité et biodiversité) selon 10 paramètres majeurs : structure, texture, capacité de fixation et d'échanges d'ions, nature et abondance des complexes argilo-humiques, porosité, fertilité, régime hydrique, potentiel redox et pH (**Gobat et al. 2010, p.44**). La diversité et la variabilité spatio-temporelle et multiscalaire des sols permettent la présence d'un nombre important de niches écologiques ce qui en fait un grand réservoir de biodiversité.



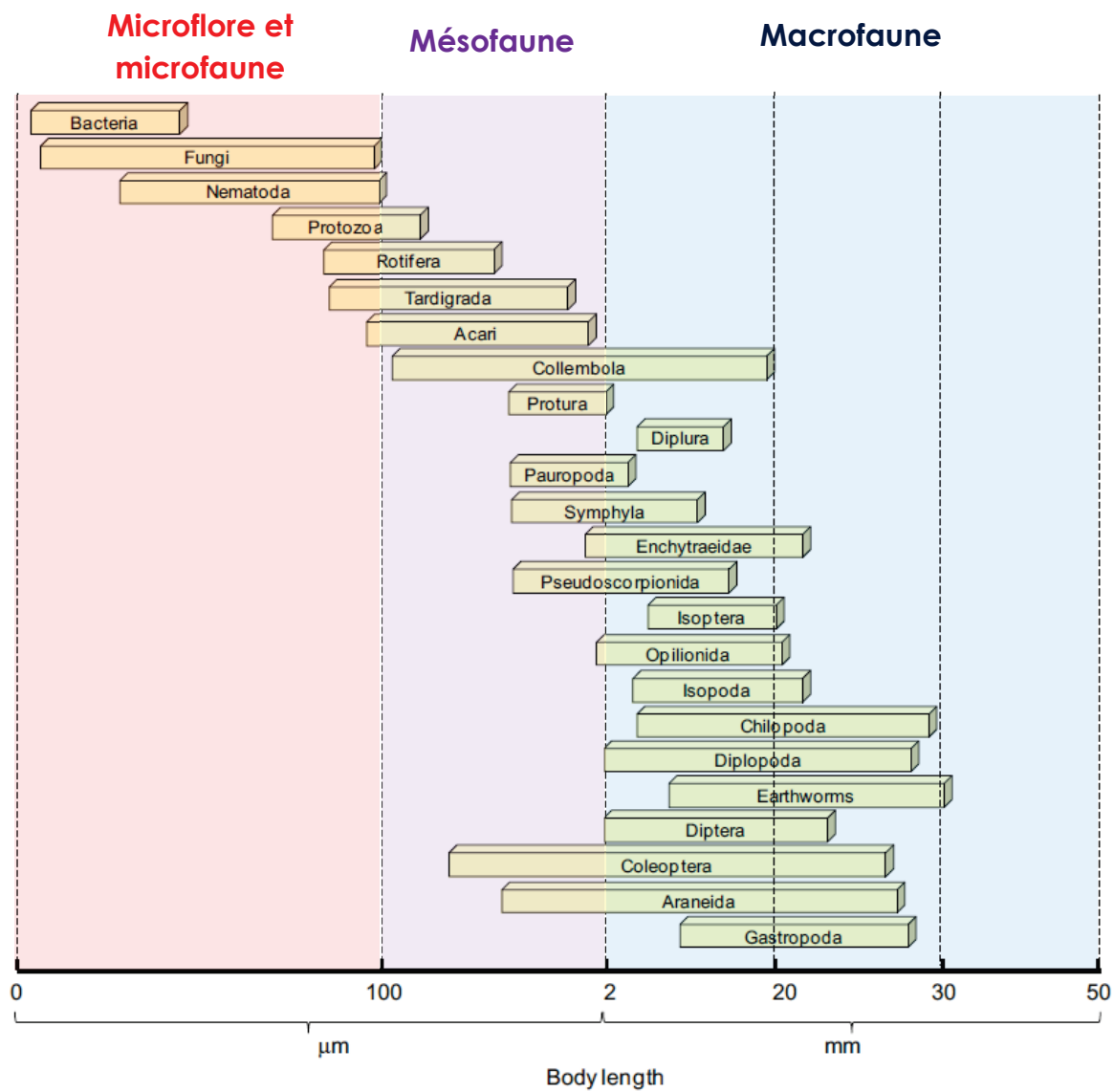


Figure I.1 Biodiversité du sol. Classification de la biodiversité des sols selon la taille des organismes.

Adapté de : Briones 2014



On y observe de très nombreuses espèces appartenant à la plupart des *taxa* existants (**Wall et al. 2001**). La biodiversité du sol peut être appréhendée en fonction de la taille des organismes qui la composent ou en fonction du rôle écologique des organismes.

Les organismes des sols peuvent être séparés en 3 groupes en fonction de leur taille (**Briones 2014**) (**Figure I.1**) :

- une macrofaune comprenant des organismes visibles à l'œil nu ayant une taille supérieure à 2 mm (divers métazoaires comme par exemple les vers de terre et les macro-arthropodes),
- une mésofaune regroupant des organismes eucaryotes d'une taille comprise entre 100 µm et 2 mm (tardigrades, collemboles et acariens),
- et une microfaune/flore d'une taille allant de 1 à 100 µm regroupant des organismes appartenant aux bactéries, archées, eucaryotes unicellulaires et pluricellulaires comme les nématodes. Cette microflore/faune que nous appellerons « les microorganismes » du sol représentent la part la plus importante de la biodiversité du sol en abondance et en richesse (**Kirchman 2012, p.9**). De par leur effectif, ils vont ainsi fortement peser sur les processus écologiques du sol.

Lorsque l'on s'intéresse aux fonctions écologiques des organismes du sol (**Gobat et al., pp.73-148 ; Briones 2014**), les macroorganismes (méso à macrofaune) ont plusieurs rôles clefs en fonction de leur place dans les réseaux écologiques : i) les organismes photosynthétiques (plantes) contribuent fortement à l'apport et à la redistribution dans les réseaux trophiques de la matière organique en fixant le carbone atmosphérique, en excréant une large gamme de composés organiques et après leur mort en tant que matière organique. Ils jouent aussi un rôle important dans les flux d'eau et de gaz par leur exploration du sol via les systèmes racinaires ; ii) les prédateurs (tardigrades, acariens,...) peuvent réguler les populations d'organismes qu'ils consomment ; iii) les organismes saprotrophes (collemboles, vers de terre,...) favorisent la



minéralisation de la matière organique morte en la fragmentant ainsi que l'aération du sol et la redistribution horizontale et verticale des éléments au sein des sols lorsqu'ils se déplacent. Les microorganismes, qu'ils soient procaryotes ou eucaryotes, ont énormément de rôles fonctionnels différents grâce à leur grande diversité spécifique et métabolique. Ils peuvent mobiliser les éléments insolubles constitutifs des minéraux (bactéries sulfato-réductrices,...), fixer l'azote (bactéries fixatrices d'azote) et le carbone atmosphérique (cyanobactéries et micro-algues), recycler la matière organique (champignons saprotrophes) et influencer les populations d'autres organismes par prédation, parasitisme ou symbiose.

Les microorganismes des sols constituent la composante la plus importante de sa biodiversité d'un point de vue populationnel (abondance, richesse) et fonctionnel (espèces " clés de voûte ", à la base des réseaux trophiques). Cependant, l'inventaire de la diversité microbienne a été principalement réalisé sur des populations de procaryotes, en négligeant les eucaryotes (**Rinke et al. 2013**). Pourtant ceux-ci jouent des rôles essentiels dans les réseaux écologiques telluriques et le fonctionnement des sols. Par exemple en consommant certaines populations de procaryotes ou de champignons de manière sélective, les microorganismes eucaryotes modifient indirectement la minéralisation du soufre, du phosphore et de l'azote ainsi que la transformation de la matière organique (**Coûteaux & Darbishire 1997**). Grâce à des stratégies trophiques variées (photosynthèse, prédation, saprotrophie, mixotrophie), certains sont potentiellement impliqués dans beaucoup de processus écologiques différents (**Coûteaux & Darbishire 1997 ; Müller et al. 2013**).

C. Importance des microorganismes dans les fonctions écosystémiques des sols

La biodiversité microbienne du sol joue un rôle central dans toutes les fonctions écosystémiques des sols en agissant directement sur ses propriétés.



Les microorganismes interviennent dans la fragmentation et l'altération de la matière solide minérale (bactéries géotrophes) ou organique morte (bactéries et champignons saprotrophes) (**Barrios 2007 ; Briones 2014**). Ces activités vont conditionner la formation des complexes argilo-humiques qui sont des éléments hybrides issus de l'association entre des argiles (particules minérales très fines), de l'humus (particules organiques) et des éléments métalliques (Al, Fe,...). Ils conditionnent l'aération du sol, la rétention de l'eau, la rétention des éléments et les échanges d'ions et vont ainsi influencer la fertilité des sols (**Gobat et al. 2010, pp.64-69**). Par leur richesse (plusieurs centaines de milliers d'espèces), leur redondance fonctionnelle et leur abondance (plus d'un milliard de cellules par gramme de sol), les microorganismes telluriques agissent aussi sur la stabilité des fonctions écosystémiques des sols face à divers types et régimes de perturbations (**Bender et al. 2016**).

II. Les pollutions métalliques des sols

Les éléments métalliques sont utilisés par les sociétés humaines depuis la fin de la préhistoire (minerais) et ont dès lors pris une place essentielle et croissante dans tous les secteurs d'activités modernes. Cette matière première est utilisée par exemples dans l'orfèvrerie (bijoux en or), l'armement (plastrons en cuivre, épées en acier, avions en alliages de métaux,...), la construction (béton armé,...), les technologies de l'information et de la communication (terres rares des circuits imprimés) (**Arndt et al. 2012, pp.9-10 ; Nuss & Eckelman 2014**). Mais bien avant cela, les métaux étaient déjà utilisés par les toutes premières formes de vie, à l'instar de l'eau, et du carbone, dans les processus cellulaires de base par exemple en tant qu'échangeur d'électrons lors des réactions de catalyse enzymatique (**Dupont et al. 2010**). Ainsi, leur concentration à l'intérieur des cellules est un caractère régulé et sélectionné depuis des milliards d'années et peut être considérée comme un trait fonctionnel fondamental des systèmes cellulaires (**Williams 2001, 2002**).



<

Figure I.2 : Classification périodique des éléments de Mendeleïv. Ce tableau de classification est composé de : 7 Lignes et 18 colonnes. Les éléments d'une même colonne ont des propriétés chimiques similaires (même type de couches de valence et même nombre d'électrons).

Adapté de : <http://www.elementschimiques.fr/>



Le développement des activités humaines associées aux deux dernières révolutions industrielles s'est accompagné d'une utilisation croissante d'éléments métalliques entraînant des pollutions dont le scandale des « boues rouges » près de Marseille (**Landrevie 2015**) en est un des exemples les plus récents. En 2000 (**Directives 91/689/CEE**), la commission européenne a placé certains métaux dans la catégorie des « substances dangereuses » pour la santé : Cd, Pb, Hg, Ni, Se, Te, Tl, Et, Cr⁶⁺, Sb, As.

Dans cette partie, après avoir défini ce que nous entendons par « métaux », nous verrons l'origine des pollutions métalliques dans les sols et leur biodisponibilité.

A. Les métaux, propriétés chimiques

Les métaux constituent un groupe d'éléments chimiques partageant des propriétés communes: 1) ils sont solides et malléables à température ambiante (sauf le mercure), 2) ils sont de bons conducteurs d'électricité avec des colorations particulières sous forme solide, 3) et ils ont une tendance à se transformer en cations. Lorsque l'on regarde la classification périodique des éléments (**Figure I.2**), les éléments métalliques représentent 66% des éléments existants et regroupent les familles des métaux alcalins (Li, Na,...), des métaux alcalino-terreux (Be, Mg,...), des lanthanides (La, Ce,...), des actinides (Ac, Th,...), des métaux de transition (Fe, Cu,...) et autres métaux (Al, Pb,...). Ils ont tendance à interagir avec les familles d'éléments plus électronégatifs : les non-métaux (C, N, O, P, S, Se) et les halogènes (F, Cl, Br, I, At) qui eux ont tendance à former des anions. Sachant que les molécules organiques (protéines, lipides et acides nucléiques) sont riches en éléments non-métaux (C, N, O et P), la plupart des métaux pourront se lier aux biomolécules.



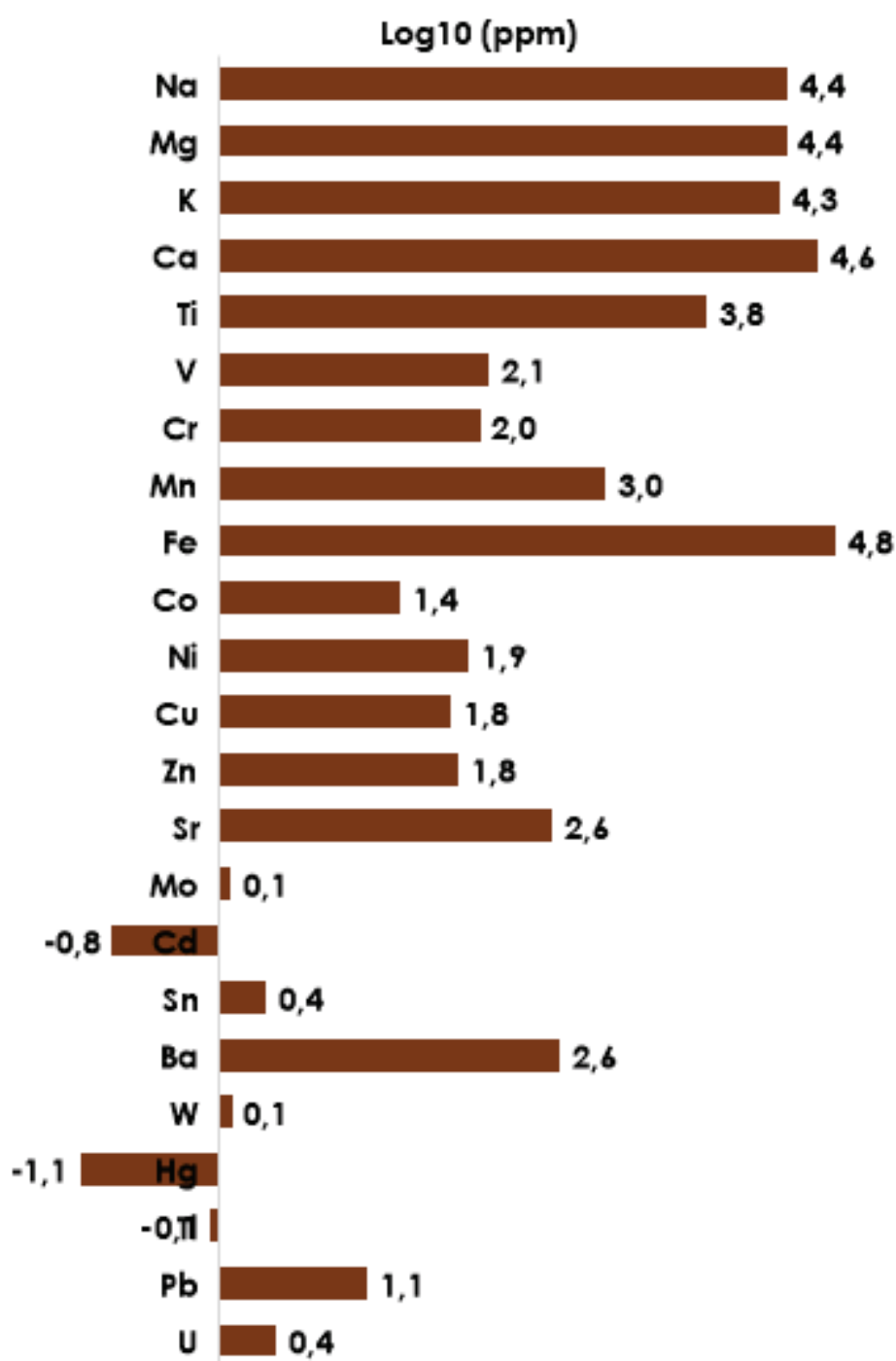


Figure I.3 : Diagramme représentant l'abondance des éléments métalliques essentiels dans la croûte terrestre en $\log_{10}$ (ppm).

Source données : <http://www.elementschimiques.fr/>



B. Origine des métaux dans les sols

Les apports dans les sols des éléments métalliques proviennent du sous-sol (apports endogènes) ou de la surface des sols (apports exogènes).

Les métaux sont des composants des différents minéraux qui constituent les roches du sous-sol (" roche mère "). Certains de ces minéraux sont particulièrement riches en métaux comme par exemple la sphalérite (ZnS) contenant du Zn ou la bauxite (Al_2O_3) contenant de l'Al. Aussi un sous-sol riche en ces minéraux pourra constituer une source de pollution en apportant des métaux dans les sols situés au-dessus. Ces apports endogènes sont possibles lors de l'altération des roches des sous-sols sous l'action de processus abiotiques (dissolution par l'eau, influence du pH et du potentiel redox) mais aussi biologiques (fragmentation des roches lors de l'exploration racinaire, bio-altération des minéraux par excrétion d'acides organiques) (**Tack 2010**). A l'échelle du globe, l'abondance des minéraux métallifères est très hétérogène (**Figure I.3**). On parle de " gisement ", lorsque des roches sont riches en minéraux métallifères et ces gisements peuvent être exploités par des industries minières permettant la remobilisation de ces éléments de la lithosphère vers la surface du sol, ce qui peut se révéler problématique pour la santé des populations et de l'environnement avoisinants (**Nriagu et al. 1990 ; Arndt et al. 2012, pp.57-58**). Cette remobilisation représente une forte source d'apports exogènes.



Les apports exogènes proviennent principalement d'aérosols (particules en suspension dans l'air) et des eaux (irrigations, pluies) dont la majorité est générée par des activités humaines. Les sources d'apports exogènes d'origine « naturelle » peuvent être des volcans en éruption ou des roches en surface (altération). Les autres apports sont issus des activités agricoles intensives (engrais minéraux, produits phytosanitaires, déchets organiques), industrielles (production d'énergie par combustion, exploitations minières, fonderies) et urbaines (incinération des déchets, épandage des boues d'épuration). Pour illustrer cela, en France les sources principales de contamination des sols au Cd proviennent pour 54% des engrais minéraux, 25% des déjections animales, 14% des retombées atmosphériques, 5% des boues et composts et pour 2% des amendements calciques et magnésiens (**Données du ministère du développement durable, 2012**).

Suite à ces pollutions, les métaux des sols vont soit être transportés *via* les flux hydriques lorsqu'ils sont dissous dans l'eau (lixiviation), soit se lier aux éléments solides du sol et perdurer sur plusieurs années. En effet, les métaux ne sont pas biodégradables et peuvent constituer un risque de stress chroniques sur les microorganismes des sols. La quantité de métaux fixés dans les sols va dépendre des propriétés du sol et seulement certaines espèces métalliques mobiles pourront être assimilées par les microorganismes des sols. On parle de " biodisponibilité " des métaux.



C. Biodisponibilité des métaux dans les sols

La biodisponibilité d'un élément métallique correspond à sa capacité à interagir avec les systèmes biologiques, à être assimiler/internaliser par les cellules. Cette interaction n'est possible que si les métaux sont sous une forme soluble dans l'eau en tant que i) cations hydratés (e.g. Me^{2+}), ii) oxy-anions (MeO^- ,...) ou iii) complexes organiques (Me-ligands organiques). La majeure partie des métaux du sol sera sous forme non biodisponible associée à la phase solide par i) liaison électrostatique réversible, ii) complexation réversible avec la matière organique et iii) minéralisation (carbonates, phosphates, sulfides) difficilement accessible aux êtres vivants (**Tack 2010**).

Les paramètres qui jouent sur la biodisponibilité des métaux sont : i) les propriétés du métal (composition électronique, état d'oxydation, état de complexification et structure moléculaire) et ii) les caractéristiques du sol (potentiel redox, pH, composition chimique, transferts d'éléments entre les différents compartiments du sol) (**Tack 2010**). Par exemple, la diminution du pH déstabilise les liaisons chimiques entre un métal et son ligand organique ayant pour conséquence sa libération dans la solution du sol sous une forme de cations hydratés (si les conditions physicochimiques le permettent). Autre exemple, un sol avec un potentiel redox oxydant favorise la minéralisation de la matière organique entraînant la libération de certains éléments comme le Zn et le Cd. Aussi, il n'y aura pollution que lorsqu'un apport en éléments métalliques dans un sol sera associé à des facteurs favorisant la biodisponibilité des métaux.

Si toutes ces conditions sont réunies les conséquences des pollutions aux métaux sur les microorganismes vont dépendre : i) du métal (nature, quantité), ii) de l'organisme considéré (dose assimilée, stade de développement, capacité métabolique), iii) du régime de perturbation (fréquence, durée), iv) et de la résultante de l'ensemble complexe d'interactions métal-environnement-cellule sur la durée de l'exposition (**Piggott et al. 2015**).



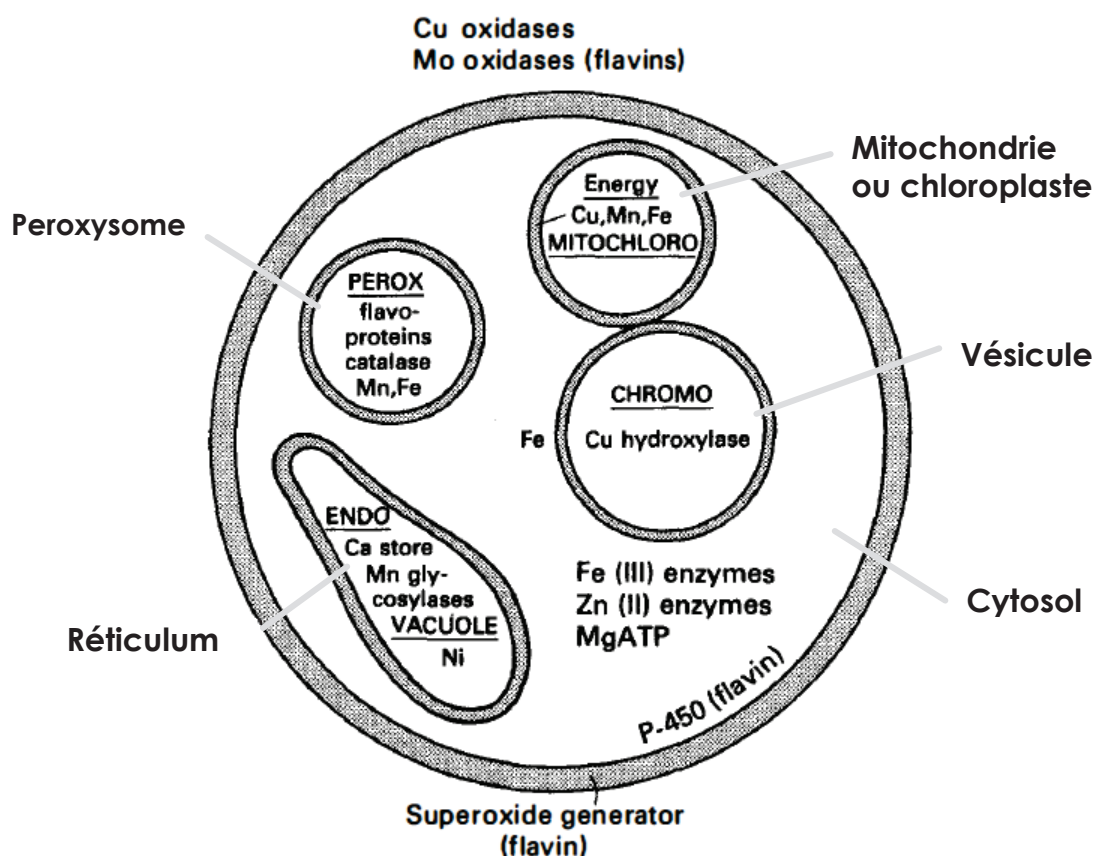


Figure I.4 : Compartimentation cellulaire de certaines espèces métalliques.

Ce schéma représente une cellule eucaryote avec les principales espèces de métaux retrouvées dans les différents compartiments cellulaires.

Source image : Fraústo da Silva & Williams 2001.



III. Réponses cellulaires aux stress métalliques

Toute cellule est composée d'éléments métalliques libres ou associés à des molécules organiques dont la concentration optimale a été optimisée en fonction de l'histoire évolutive des organismes et des conditions environnementales présentes (**Dupont et al. 2010 ; Merchant & Helmann 2012**).

Lors de pollutions métalliques, les espèces de métaux bio-disponibles assimilées vont perturber le fonctionnement cellulaire. Dans cette partie, nous verrons dans un premier temps comment les métaux perturbent le physiologie métallique normale puis quelles réponses sont mises en place par la cellule pour s'affranchir de ce stress.

A. Modification de l'homéostasie métallique des cellules et perturbation des fonctions métaboliques

• Métaux physiologiques et métallomes

Un " métal physiologique " est un métal qui est impliqué dans le métabolisme cellulaire. On trouve 13 métaux physiologiques essentiels au fonctionnement cellulaire qui sont majoritairement des métaux alcalins et alcalino-terreux (Na, Mg, K, Ca) puis des métaux de transition (V, Mo, W, Mn, Fe, Co, Ni, Cu, Zn). Beaucoup d'autres métaux sont utilisés par de rares organismes (Sr, Ba, Cd,...).

Ces métaux vont être sous des formes bio-disponibles qui sont soit libres soit liées à des composés organiques. A noter que la répartition dans la cellule est très hétérogène et nous allons retrouver plusieurs espèces d'un même métal (Cu^+ , Cu^{2+} ,...) dans les cellules mais les proportions vont changer en fonction des compartiments considérés (**Figure I.4**). Par exemple, chez les eucaryotes, nous allons pouvoir observer à la fois des espèces de Cu sous forme libre ($\text{Cu}^+_{(\text{aq})}$ et $\text{Cu}^{2+}_{(\text{aq})}$) et sous forme liée à des protéines (Cu-CopA, Cu-métallothionéine, métallo-enzymes).





Mais la majeure partie du Cu cellulaire sera localisée dans des vésicules sous forme libre et très peu dans le cytoplasme sous forme chélatée (**Palumaa 2013**). Les espèces métalliques physiologiques ainsi retrouvées vont dépendre de la réactivité d'un métal, des conditions physicochimiques locales et des processus métaboliques dans lesquels elles sont impliquées. Les métaux sous forme mobile interviennent dans le transfert d'énergie et la signalisation cellulaire (Na^+ , K^+ , Mg^{2+} , Ca^{2+} , Fe^{2+} , Mn^{2+} , Zn^{2+}). Les métaux liés à des molécules organiques interviennent dans la chimie redox (Fe, Cu, Mn, Mo, Ni, Co) ou la chimie acido-basique (Zn^{2+} , Ni^{2+} , Fe^{3+} , Mn^{2+} , Mg^{2+}). Certains métaux vont donc être multifonctionnels (Mg, Zn et Fe) (**Merchant & Helmann 2012 ; Fraústo da Silva & Williams 2001**). Ainsi chaque cellule aura une quantité et une répartition de métaux qui lui sont propres et qui varient au cours de sa vie.

On définit le « métallome » comme étant l'ensemble des métaux présents dans les cellules à un moment donné (**Williams 2001 ; Yannoni et al. 2012**). A titre d'exemple, le métallome de *Synechococcus* sp. (Bactérie phytoplanctonique) est composé majoritairement de métaux alcalins et alcalino-terreux (K, Ca et Mg) et de très peu de métaux de transition (Fe, Mn, Zn, Cu, Co, Cd) (**Merchant & Helmann 2012**). Cette composition ne semble pas être le fruit du hasard car les métaux alcalins et alcalino-terreux sont les métaux les plus abondants des océans et de la croûte terrestre ; ils sont facilement accessibles et faiblement réactifs en présence d'éléments non métalliques (**Dupont et al. 2010**). A l'échelle des compartiments cellulaires (cytoplasme, organites), on va retrouver un métallome différent par compartiment qui reflète les fonctions métaboliques dans lesquels ces éléments sont impliqués. Ainsi, par exemple chez les eucaryotes, le K qui est impliqué dans les potentiels de membrane et le pouvoir tampon du cytosol sera surtout localisé dans le



A. Série des métaux monovalents :

Cu > Na, K

Série des métaux divalents/Série de Irving-Williams :

(Mg, Ca) < Mn < Fe < Co < Ni < Cu > Zn < Cd

Série des métaux trivalents :

Co > Mn > Fe > Cr > Al

B.

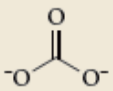
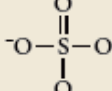
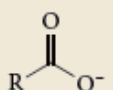
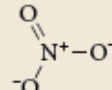
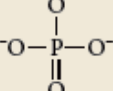
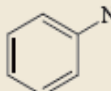
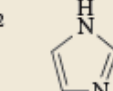
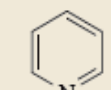
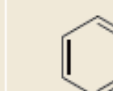
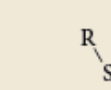
	Hard	Borderline	Soft
Acids	Na ⁺ , K ⁺ , Mg ²⁺ , Ca ²⁺ , Cr ³⁺ , Al ³⁺ , Ga ³⁺ , Co ³⁺ , Fe ³⁺	Cu ²⁺ , Zn ²⁺ , Pb ²⁺ , Bi ³⁺ , Ni ²⁺ , Co ²⁺ , Fe ²⁺	Cu ⁺ , Au ⁺ , Ag ⁺ , Hg ²⁺ , Hg ⁺ , Cd ²⁺
Bases	 Carbonate  Sulphate  Carboxylates  Nitrate R-OH Alcohols H₂N-R Amines  Phosphate R-O-R Ethers Also: H₂O, OH⁻, NH₃, hydrazine	 Aniline  Imidazole  Pyridine  Nitrite R-N=N⁺=N⁻ Azides Also: N₂	 Phenyl groups R-SH Thiols  Ethylene  Thioethers ≡N⁻ Cyanide Also: H₂S, H₂⁻

Figure I.5. Sélectivité des métaux.

- A. **Séries des métaux mono-, di- et trivalents.** Ces séries ont été obtenues expérimentalement et permettent de classer les métaux du moins au plus réactif.
- B. **Tableau présentant différents acides et bases qualifiés de dur ou mou en fonction de leur réactivité chimique.** La théorie « HSAB : *Hard-Soft Acid Base* » prédit que les acides durs vont avoir une forte affinité pour les bases dures, les acides intermédiaires pour les bases intermédiaires, les acides mous pour les bases molles.

Source image : Lemire et al. 2013 © 2013, Rights Managed by Nature Publishing Group.



cytosol sous forme de cation hydraté (K^+) alors que le cuivre très réactif sera concentré dans des vésicules sous forme libres (Cu^{2+}) permettant de digérer des composés internalisés par endocytose (**Palumaa 2013 ; Fraústo da Silva & Williams 2001**).

• Les interactions métaux-ligands et fonctions métaboliques

En chimie inorganique, on appelle les molécules capables de se lier aux métaux des « ligands ». La denticité ou le « mordant » correspond au nombre de métaux liés par ligand. On parle de ligand monodenté s'il est lié à un atome métallique, bidenté s'il est lié à deux atomes et polydenté s'il est lié à plus de deux. La chélation est un cas particulier de la liaison entre un ligand polydenté à un métal. Le métal est lié au ligand en plusieurs points entraînant un repliement en cercle du ligand autour du métal. Le ligand est alors qualifié de chélateur (EDTA, métallothionéines,...). En fonction de son état de valence (nombre de charges), un métal sera capable de former un nombre de liaisons fini avec un ligand donné selon une géométrie particulière. Par exemple, le Cu^{2+} peut réaliser des liaisons de géométrie tétraédriques au sein de métalloprotéines ou de cristaux alors que le Cd^{2+} peut réaliser soit des liaisons tétraédriques soit octaédriques (**Rulíšek & Vondrášek 1998**). Lorsqu'un métal va se retrouver lié à un réseau de ligands et d'autres métaux on va parler de "complexes" (complexes argilo-humiques,...).

L'affinité d'un métal pour un ligand semble dépendre : des propriétés physico-chimiques du métal, des propriétés du ligand, des propriétés du solvant dans lequel se passe l'interaction et de l'environnement moléculaire (abondance d'autres métaux et autres ligands) (**Dudev & Lim 2008**). Cette affinité a été bien étudiée dans des conditions physiologiques et a donné des « séries » lorsque les métaux sont complexés et une table empirique de sélectivité (**Figure I.5**). Ces données permettent de prédire les affinités entre métaux physiologiques et ligands (**Lemire et al. 2013 ; Fraústo da Silva &**



Williams 2001). Par exemple, dans la théorie HSAB (*Hard and Soft Acids and Bases*), le Cd^{2+} réagit préférentiellement avec des groupements thiol (RSH) ou de l'éthylène (C_2H_4) et remplace le Cu^+ lorsque celui-ci est lié à des molécules organiques.

Parmi les ligands organiques, les protéines occupent une place de choix. Sous le terme " métalloprotéome " sont regroupées toutes les protéines actives lorsqu'elles sont liées à un métal. Le métalloprotéome représente entre 30 et 40% des protéines dans une cellule et diffère en fonction de l'histoire évolutive des organismes comme pour le métallome (**Merchant & Helmann 2012 ; Dudev & Lim 2008**). Par exemple, les bactéries ont de l'ordre de 10% de protéines à Fe et 5 à 6% de protéines à Zn dans leur métalloprotéome avec des métalloenzymes fonctionnant surtout avec du Fe, Mn et Mo mais qui restent capables de correctement fonctionner avec plusieurs métaux différents. Ces enzymes ont en quelque sorte un large spectre de co-facteurs (enzymes cambialistiques). Contrairement aux bactéries, les eucaryotes ont environ 10% de protéines à Zn et entre 5 à 6% de protéines à Fe. Les métalloenzymes seront surtout à Ca, Cu et Zn et ne peuvent fonctionner convenablement qu'avec un seul métal contrairement à celles des bactéries (**Merchant & Helmann 2012**). Cette différence de composition met en avant l'importance du Zn et du Cu chez les eucaryotes et pourrait jouer un rôle très important dans les interactions entre eucaryotes et procaryotes notamment dans la lutte contre les bactéries pathogènes avec la séquestration d'éléments comme le Fe pour limiter la croissance des bactéries ou le relargage de Cu par des macrophages humains (**Lemire et al. 2013**).

Ainsi le métallome et le métalloprotéome sont liés l'un à l'autre et sont des traits fonctionnels essentiels pour les cellules. Ils sont régulés respectivement par des mécanismes d'acquisition, de transport, et de confinement des métaux et par des mécanismes de synthèse, lyse et transformation des ligands. En cas de déséquilibre de l'un, la composition de l'autre est modifiée pour limiter les effets délétères sur les processus cellulaires.



• La cytotoxicité des métaux

Un stress métallique peut avoir des effets sub-létaux ou létaux sur les cellules. Dans le cas d'un stress métallique sub-létal, la cytotoxicité va se manifester à plusieurs échelles. Tout d'abord à l'échelle moléculaire les métaux les plus réactifs vont s'associer de manière aspécifique aux molécules organiques et engendrer la production d'espèces réactives d'oxygène (**Lemire et al. 2013**). En effet, comme susmentionné, les métaux tels que le Cu, le Fe ou le Zn ont une forte affinité pour les éléments non-métalliques (C, N, P,...) très abondants dans les molécules organiques. Ils vont se lier à ces molécules de manière aspécifiques et entraîner des catalyses enzymatiques de biomolécules telles que l'ADN ou le repliement non fonctionnel d'enzymes en remplaçant le métal structural par celui en excès comme par exemple le remplacement irréversible du Zn par du Cd. La formation d'espèces réactives (réactions d'Haber-Weiss, de Fenton,...), notamment celles de l'oxygène, induit la modification du potentiel redox et du pH des cellules perturbant les liaisons chimiques des molécules organiques ainsi que leur repliement. Par exemple, en présence d'un excès de Fe^{2+} , de Cr^{5+} , d' As^{3+} , de Te^{4+} , de Cd^{2+} et de Cu^{2+} sous forme libre, et en conditions acides, il y a formation du radical hydroxyle ($\cdot\text{OH}$) à partir de peroxyde d'hydrogène (H_2O_2) (**Rico et al. 2009**). Ces molécules très réactives vont être prises en charge par des antioxydants présents dans la cellule (superoxyde dimutase, thiol,...) mais rendront la cellule vulnérable à des stress oxydatifs supplémentaires.





A l'échelle cellulaire, ces réactions vont avoir pour conséquences à court terme : i) de perturber le fonctionnement métabolique (génétoxicité, réactions catalytiques « parasites »), ii) de modifier ce fonctionnement avec la synthèse de molécules en réponse à la nouvelle stœchiométrie métallique et donc iii) de rendre la cellule potentiellement plus vulnérable face à de multistress mais aussi moins vulnérable face aux stress de même nature. A long terme, ces cellules auront : i) des modifications génétiques, ii) un métallome et un métallopeptome modifiés et iii) une valeur adaptative (*fitness*) réduite ou augmentée en fonction des conditions environnementales.

Dans le cas de stress létaux, les fonctions cellulaires sont perturbées sévèrement conduisant à l'inactivation du métabolisme et à la mort des cellules. La sévérité va dépendre du métal considéré et de la capacité de la cellule à " tamponner " cet excès de métal par ce qu'on pourrait appeler son " ligome " (métallopeptomes et autres molécules organiques pouvant réguler la concentration des métaux de manière directe ou indirecte). Dans ce cas, les cellules n'ont pas la capacité de changer suffisamment leur métabolisme pour qu'il fonctionne convenablement dans ces nouvelles conditions et vont donc perdre leur intégrité. Par exemple, l'excès léthal de métaux de transition comme le Cu, entraîne une augmentation des interactions aspécifiques entre le Cu et des ligands cellulaires conduisant à des réactions « parasites » en chaîne perturbant la stabilité des fonctions métaboliques (inactivation de métallopeptases, création d'espèces réactives de l'oxygène, modification des flux ioniques, déstabilisation des potentiels membranaires, perte de l'intégrité cellulaire et désagrégation de la cellule) (**Lemire et al. 2013**). Pour limiter l'effet délétère de ces mécanismes et assurer la stabilité des fonctions métaboliques, différentes stratégies sont possibles. Ces stratégies passent notamment par l'induction transcriptionnelle de gènes spécifiquement induits en réponse à l'excès de métaux dans le milieu intracellulaire. Cette modification de l'ensemble des ligands cellulaires permet de rééquilibrer le métallome et d'assurer un fonctionnement métabolique viable.



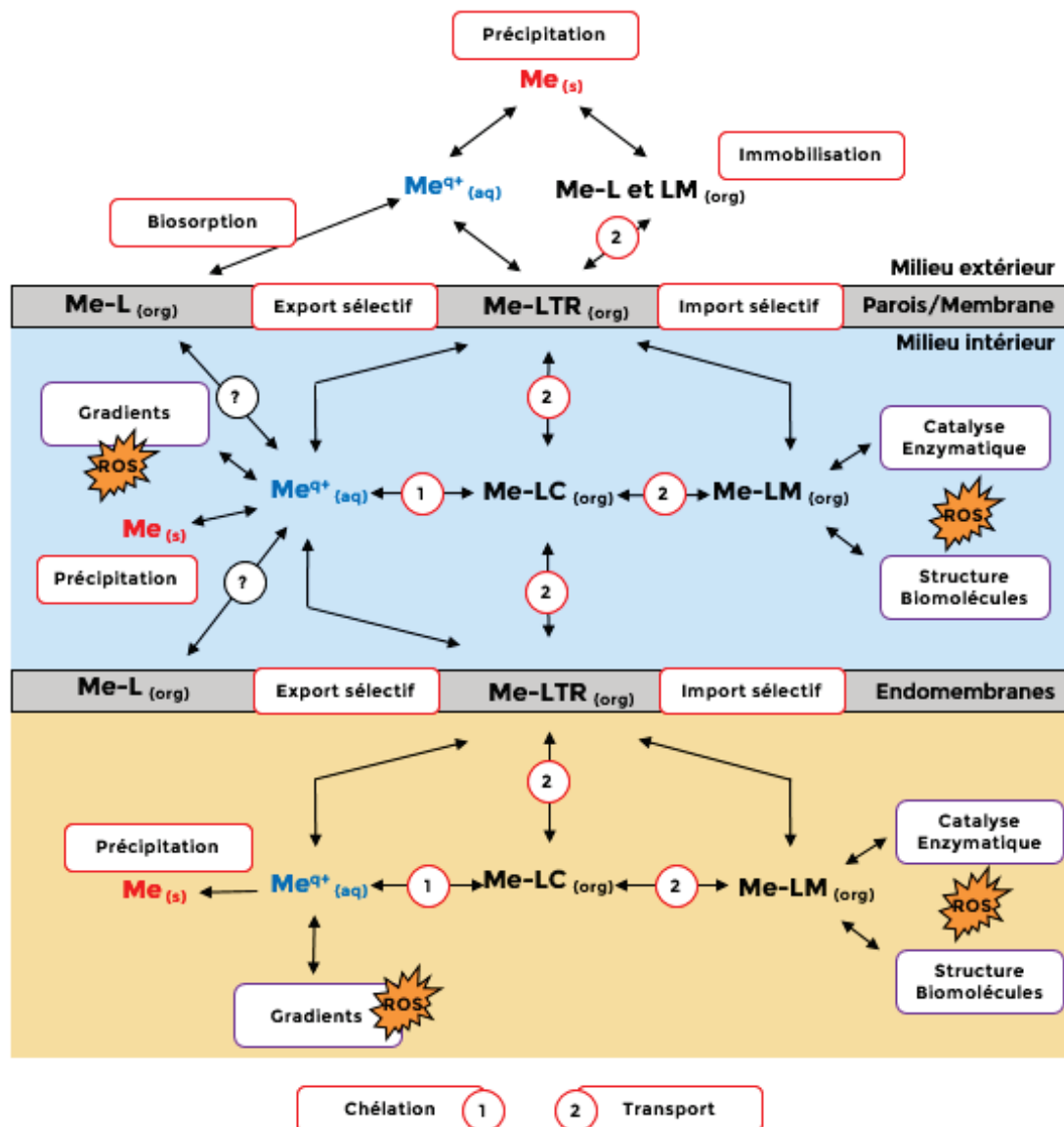


Figure I.6. Schéma simplifié des mécanismes de résistance aux métaux dans le milieu extracellulaire, dans les cellules et dans les endomembranes. Tous ces mécanismes sont interconnectés et leur importance va dépendre du type de métal, du type d'organismes, des concentrations en métaux et en ligands, du potentiel redox, du pH et de la cinétique des transferts de molécules. Les ligands peuvent être des chélateurs (**LC**), des transporteurs ou des récepteurs membranaires (**LTR**) et des molécules intervenant dans le métabolisme comme les protéines, lipides et saccharides (**LM**).

$Me_{(s)}$: Métaux en phase solide.

$Me^{q+}_{(aq)}$: Cations métalliques en phase aqueuse où q+ représente la charge

$Me-L_{(org)}$: Métaux liés à un ligand en phase organique.

Les mécanismes de régulation des métaux **sont entourés en rouge** alors que les fonctions métaboliques, dans lesquels les métaux sont impliqués **sont entourées en violet**.

Adapté de : Fraústo da Silva & Williams 2001



B. Les mécanismes de "résistance" aux métaux

En réponse à la présence de métaux, les processus cellulaires sont réarrangés de manière à répondre le plus précisément possible au stress induit par l'élément métallique. Par exemple lorsqu'on applique un stress métallique aigu en utilisant des concentrations sub-létales, on peut observer une réponse en deux temps. Dans un premier temps, une réponse précoce permet l'induction de gènes impliqués dans le remplacement des protéines endommagées et la synthèse de molécules nécessaires à l'acclimatation à la nouvelle stœchiométrie métallique (régulateurs de transcription, protéases, transporteurs intracellulaires, synthèse de protéines, systèmes antioxydants, chélateurs et protéines membranaires d'export). Dans un deuxième temps, seule subsiste l'induction de gènes spécifiques au stress métallique (systèmes antioxydants, chélateurs, transporteurs membranaires) (**Kaur et al. 2006 ; Yepiskoposyan et al. 2006**).

Les stratégies d'acclimatation sont multiples et vont dépendre à la fois des propriétés du métal et des capacités métaboliques des organismes. On peut distinguer différents modes de résistance aux métaux selon le compartiment cellulaire considéré (**Figure 1.6**) :

- **Dans le milieu extracellulaire, on observe des mécanismes d'immobilisation, de biosorption et de biominéralisation des métaux.**

Les cellules excrètent et libèrent des molécules organiques qui vont se lier aux métaux, les immobiliser dans le milieu et limiter leur réactivité. Ces molécules peuvent être non spécialisées (acides nucléiques, exopolysaccharides, acides organiques) ou spécialisées (chélateurs spécifiques). Par exemple les sidérophores sont des molécules chélatrices de métaux retrouvées chez les bactéries, les champignons et les plantes avec une forte affinité pour le Fe^{3+} . Elles sont induites et excrétées en cas de carence en Fe mais aussi en présence d'autres métaux comme le Mo chez *Azotobacter*



vinelandii. Certains sidérophores peuvent, comme la Pyoverdine de *Pseudomonas aeruginosa*, être capables de chélater beaucoup d'autres métaux dans des conditions physiologiques mais toujours avec des constantes d'affinité inférieures à celles du Fe (**Schalk et al. 2011**).

Certains ligands organiques retrouvés dans le milieu extracellulaire conduisent à la biominéralisation des métaux sous certaines conditions. Par exemple, la liaison du fer à pH neutre sur les fibres extracellulaires de *Rhodobacter sp.*, riches en C organique, conduit à la formation de nanocristaux de goethite (α -FeOOH) (**Miot et al. 2009**). Les métaux s'agglutinent naturellement sur certaines parois cellulaires. Chez les bactéries à Gram positif, les métaux se lient aux groupes carboxyles des acides glutamiques du peptidoglycane. Chez les eucaryotes, les métaux peuvent se lier sur les parois via les groupements amines de la chitine par exemple (champignons).

La biosorption des métaux sur les couches S (*Surface layer*) de bactéries et d'archées se fait de manière passive et aspécifique (**Merroun et al. 2005**). Avec le temps et en présence de beaucoup de métaux, ces parois se retrouvent complètement minéralisées. A l'échelle de la population, cela favorise la minéralisation des éléments métalliques dans le milieu (**A. Kich, échanges personnels**).

La présence de métaux peut aussi induire des changements dans la composition des parois comme par exemple chez les champignons (**Bellion et al. 2006**). Elle peut induire une forme de résistance (**Macfie & Welbourn 2000**). Par exemple, chez 3 espèces de ciliés, la présence de Cd et de Zn induit une perte de la motilité et une différenciation membranaire similaire à ce qui peut être observé peu de temps avant un enkystement (**Martín-González et al. 2005**).



- **Dans la membrane cellulaire, on observe une grande diversité de transporteurs.**

Les transporteurs membranaires permettent des échanges constants d'éléments entre les milieux intra- et extracellulaire. En fonction des familles de transporteurs (Nies 1999 ; Srivastava & Kowshik 2013), ils permettent l'import, l'export ou les deux en même temps, en plus d'avoir des spectres de substrats plus ou moins larges. La modification de la combinaison de transporteurs membranaires va permettre la régulation de la quantité de métaux intracellulaires, par exemple via la répression d'un transporteur (permettant l'import) de type ABC chez l'Archées *Hallobacterium NRC-1* et l'augmentation de l'efflux via les P₁ATPases en présence de fer (Kaur et al. 2006). Toutefois, il n'a jamais été observé un import de métaux totalement bloqué dans une cellule physiologiquement active. Ainsi, les métaux peuvent entrer par le biais de pompes à influx peu spécifiques en cas de pollutions métalliques (Horner et al. 2013). Les transporteurs constituent un goulot d'étranglement pour l'entrée et la sortie des éléments métalliques à la fois dans la cellule et dans les organites dédiés. Il semble qu'aucun mécanisme de résistance aux métaux ne puisse s'en affranchir.

- **Dans le milieu intracellulaire, on observe une immobilisation par chélation, un transport et une bio-atténuation des métaux.**

Une fois dans le cytosol, les métaux libres réactifs sont pris en charge par plusieurs types de molécule organique. Ces molécules ou ligands (chélateurs intracellulaires, acides organiques) peuvent se lier à plusieurs métaux différents comme par exemple les métallothionéines qui peuvent se lier au Zn²⁺, Cu⁺, Cd²⁺ ou les phytochélatines se liant au Cd²⁺.

Les chélateurs intracellulaires sont des protéines de petite taille et contiennent des régions riches en acides aminés à forte affinité pour les métaux comme les Cys, Glu, Gly, His,.... Ils permettent le transport des éléments



métalliques vers des organites et le transfert des métaux à d'autres molécules organiques. Afin d'assurer les processus cellulaires de base, un pool de chélateurs doit être maintenu dans le cytosol. La production de certains d'entre eux est par ailleurs souvent induite par différents stress et notamment la présence de métaux. (**Sies 1999 ; O'Halloran & Culotta 2000 ; Palumaa 2013**). Ces chélateurs sont des molécules centrales dans le métabolisme et la résistance métallique car ils régulent les flux de métaux cellulaires.

Beaucoup d'enzymes d'organismes chimio-lithotrophes sont capables de catalyser des réactions changeant l'état de valence des métaux vers d'autres formes et de s'en servir comme source d'énergie. Dans certains cas, ces réactions sont utiles si elles font passer un métal d'une forme toxique à une forme non toxique comme par exemple, la mercurique réductase qui catalyse la réduction de l'ion Hg^{2+} en Hg^0 (**Mukhopadhyay et al. 2002**). Dans ces cas, on parle de bio-atténuation.

• La compartimentation intracellulaire et la biominéralisation.

La biominéralisation est un mécanisme commun chez les organismes lithotrophes (**Edwards & Bazylinski 2008**). En présence d'une forte concentration de métaux, il se produit une accumulation et une précipitation de ceux-ci sous forme de granules généralement par liaison avec des sulfides de type As_2S_3 par exemple chez *Desulfotomaculum auripigmentum*. Chez des bactéries et champignons, on observe aussi des granules de polyphosphates (MeHPO_4). Par ailleurs, les bactéries magnétotactiques stockent des minéraux de magnétite (Fe_3O_4) et de greigite (Fe_3S_4) par l'intermédiaire de molécules organiques dans des structures appelées magnétosomes (**Edwards & Bazylinski 2008**). Chez les archées, les métallochaperonnes CutA (Copper-tolerance) se lient aux métaux entraînant leur multimérisation puis leur précipitation (**Srivastava & Kowshik 2013**). Cette protéine bien que présente aussi chez les



bactéries et les eucaryotes ne provoque la précipitation des métaux que chez les archées (**O'Halloran & Culotta 2000**).

La précipitation se déroule souvent dans des compartiments cellulaires spécialisés tels que les vacuoles chez les champignons sans doute car ils permettent d'obtenir les conditions nécessaires à cette minéralisation qui ne sont pas forcément celles régnant dans le cytosol. Cette compartimentation repose sur les gradients et les systèmes de transport d'éléments naturellement en place au sein des cellules. La sélectivité des endo-membranes est permise par ses transporteurs membranaires (**Clemens 2001**). Par exemple, la levure stocke du Co^{2+} , Mn^{2+} , Zn^{2+} , Mg^{2+} et K^{+} , sous forme de granules de polyphosphate dans ses vacuoles (**Gadd 1990**).

Il existe ainsi plusieurs manières pour une cellule de résister aux effets cytotoxiques des métaux. Il semble néanmoins que deux mécanismes prépondérants soient indispensables à la résistance aux métaux des systèmes cellulaires : le transport membranaire de métaux et la chélation intracellulaire.

Parmi les chélateurs intracellulaires décrits, certains semblent impliqués dans une fonction comme les phytochélatines produites en réponse à un stress métallique alors que d'autres sont multifonctionnelles comme les métallothionéines semblant indiqué que leur propriété vis-à-vis des métaux sont essentielles pour l'adaptation des organismes aux changements environnementaux.





IV. Les métallothionéines

Les métallothionéines (MT) sont définies comme étant un groupe de petites protéines (24-162 acides aminés) riches en cystéines (15-35%) organisées en motifs conservés. Elles sont pauvres en acides aminés aromatiques et sont capables de chélater les métaux. Leur très large répartition, notamment chez les eucaryotes (mais elles sont aussi présentes dans certains groupes bactériens tel que celui des Cyanobactéries), témoigne de leur importance dans les processus cellulaires visant à maintenir l'homéostasie métallique et à tolérer différents types de stress. Ainsi, ces protéines sont transcrites de manière basale et/ou en réponse à plusieurs types de stress à la fois généraux (oxydatifs) ou spécifiques (métalliques) ou à des signaux moléculaires (hormones, médiateurs cellulaires). Leur liaison aux métaux est permise grâce essentiellement aux groupements thiol des cystéines, mais parfois aussi à la présence d'histidines et leur chaîne latérale azotée, conduisant à une interaction de type métal-ligand. (**Blindauer & Leszczyszyn 2010 ; Capdevila et al. 2012**). Les MT sont classées soit selon leur similarité de séquences, soit selon leur préférence pour un type de métal donné.

Ci-dessous nous allons développer les caractéristiques structurales de ces protéines avec leur structure primaire, secondaire et tertiaire puis leur distribution dans le monde vivant aux échelles tissulaire et cellulaire, puis leur diversité de formes et d'interaction avant de terminer par les deux classifications de MT en vigueur.



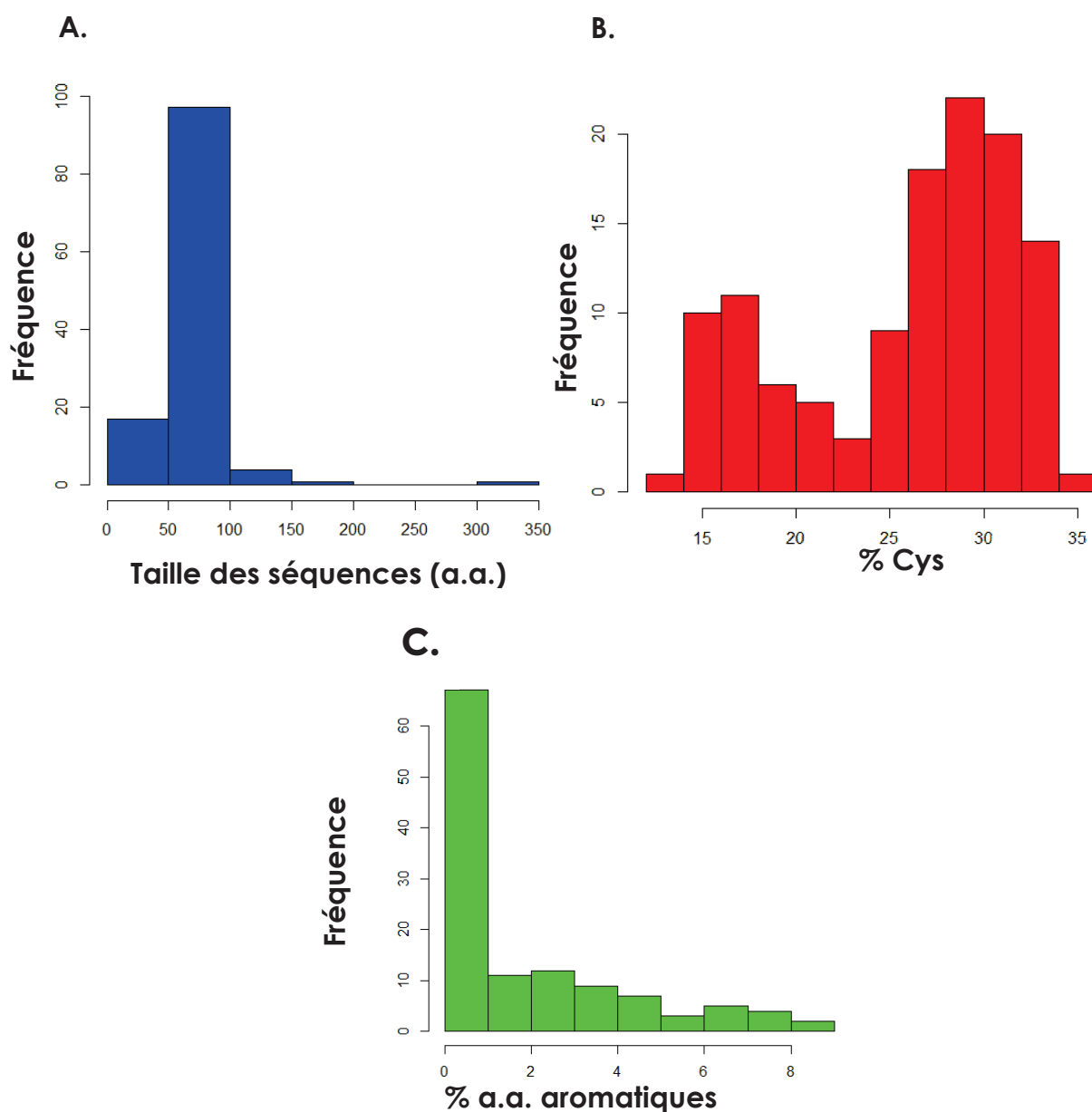


Figure 1.7 : Caractéristiques de la structure primaire des MT à partir de séquences protéiques présentes dans Swiss-Prot. A. Fréquence de tailles de leurs séquences en acides aminés. **B.** Fréquence en Cystéines (en % d'acides aminés). **C.** Fréquence en acides aminés aromatiques (en % de séquences).



A. Structures des métallothionéines

Les métallothionéines sont des protéines d'une grande variabilité dans l'organisation de leurs séquences primaires. De plus, elles n'ont pas nativement de structure secondaire mais acquièrent une structure tertiaire au moment de leur liaison avec les métaux. Cette structure semble dépendre du nombre et de la nature des métaux liés.

1. Structure primaire

• Caractéristiques et composition en acides aminés

La majeure partie des MT possède une taille comprise entre 24-85 acides aminés (**Figure 1.7A**). Quelques-unes atteignent des tailles plus conséquentes comme les MT des Ciliés *Tetrahymena spp.* (99 - 162 acides aminés), de *Trichomonas vaginalis* (308 acides aminés) (*Excavata*) et des champignons basidiomycètes *Cryptococcus neoformans* (122 -183 acides aminés) et *Tremella mesenterica* (257 acides aminés). Cette taille plus importante semble être due à des phénomènes de duplication de gènes engendrant des protéines chimériques avec une organisation des cystéines en modules répétés correspondant à plusieurs petites MT ancestrales accolées (**Gutiérrez et al. 2011 ; Palacios et al. 2014 ; Iturbe-Espinoza et al. 2016**).

La liaison Me-MT se fait principalement par l'intermédiaire des atomes de soufre présents dans les cystéines (Me-SCys). Le nombre de Cys liées va dépendre du type de métal. La composition importante en Cys est la principale caractéristique des MT avec entre 15 et 33% de résidus par protéine (**Figure 1.7B**) (le pourcentage moyen de Cys des protéines est en général d'environ 2%). Ce fort pourcentage en Cys permet aux MT de potentiellement se lier à beaucoup de métaux à la fois par rapport aux autres métalloprotéines.





Figure I.8 : Séquences primaires de métallothionéines illustrant l'organisation ou non des motifs de Cys en modules. **A.** Métallothionéine de souris (P02802) (Métazoaire) n'ayant pas de structure particulière. **B.** Métallothionéine de pois chiche (CAA65008) (*Chloroplastidia*) ayant une structure bimodulaire. **C.** Métallothionéine de Blé (P30569) (*Chloroplastidia*) présentant une structure trimodulaire. **D.** Métallothionéine de *Tetrahymena* (*Alveolata*) avec une structure en modules répétés.

Les Cys sont représentées en rouge, les His en vert.

Les modules conservés au sein d'une même sous-famille de métallothionéines sont encadrés en bleu et les régions intermodulaires sans Cys sont soulignées.

Les modules répétés au sein d'une même métallothionéine sont soulignés par une flèche.



Plus de 60% des MT ne possèdent pas d'acides aminés aromatiques mais certaines peuvent en posséder jusqu'à 8% (MT de plantes ou de champignons) (**Figure I.7C**). Pour celles qui en possèdent, il n'est pas rare de retrouver des histidines conservées qui interviennent aussi dans la coordination des métaux (Me-His) notamment parmi les MT de champignons, métazoaires, plantes, cyanobactéries et protéobactéries. Les His pourraient avoir un rôle de modulation de la liaison avec les métaux divalents et permettre la réduction de la sensibilité des MT à l'oxydation (**Blindauer 2008**).

Trinchella et ses collaborateurs en 2008 (**Trinchella et al. 2008**) ont montré qu'au sein des MT de vertébrés, les MT de mammifères ont des index d'hydropathie plus élevés, comparés aux autres MT de vertébrés. Cet index se calcule en fonction de la charge électrique d'un acide aminé par rapport aux acides aminés voisins et est lié à la plasticité et à la flexibilité globale de la protéine. Avec leur index plus élevé, les MT de mammifères seraient donc moins flexibles que celles des autres vertébrés. Cette moindre flexibilité pourrait être liée à un plus grand polymorphisme des MT de mammifères (17 gènes codant des MT chez l'homme) (**Capasso et al. 2003 ; Trinchella et al. 2008 ; Zalewska et al. 2014**)

• Organisation des cystéines en motifs

Les métallothionéines (MT) sont riches en Cys conservées et ces Cys sont agencées d'une manière bien particulière appelée « motifs » (**Figure I.8**). Au total, neuf types de motifs différents ont été recensés, certains sont retrouvés uniquement dans quelques clades et/ou quelques isoformes de MT comme le motif CCxCC (où x est n'importe quel autre acide aminé) que l'on ne retrouve que chez les vertébrés, CxCxC spécifique des MT du nématode *Caenorhabditis elegans* et de la levure *Saccharomyces cerevisiae* ou CCxC présent uniquement chez l'algue brune *Fucus vesiculosus*. D'autres motifs sont retrouvés dans toutes les MT comme le motif CxC. Ces motifs sont parfois



accompagnés d'acides aminés conservés qui peuvent être importants pour la stabilité de la liaison avec le métal comme les sérines conservées du motif SCCSCCx des vertébrés (**Emoto 1996**). La composition en motifs peut être utilisée pour comparer les métallothionéines entre elles notamment chez les MT de *Tetrahymena* spp. (**Gutiérrez et al. 2009 ; 2011 ; Espart et al. 2015**). Actuellement, ni ces motifs, ni les autres acides aminés ne permettent de conclure quant à l'affinité d'une MT pour un métal en particulier. Des MT avec des motifs différents peuvent avoir des affinités pour des métaux comparables lors d'études *in vitro* comme par exemple la sous-famille CuMT (contenant les motifs CxC, xxCxx) de *Tetrahymena* spp. et Cup1 (ayant les motifs CxCC, CC, CxC et xxCxx) de *Saccharomyces cerevisiae* qui ont toutes une forte affinité pour le cuivre. A l'inverse des protéines avec les mêmes motifs peuvent avoir une affinité différente comme pour les MT d'escargot CdMT (CxC, xxCxx), CuMT (CxC, xxCxx) et CdCuMT (CxC, xxCxx) (**Palacios et al. 2011a**). Il semble que les acides aminés environnant ces motifs Cys soient importants pour l'affinité d'une MT pour un métal donné.

• Organisation des motifs en modules

A l'échelle de toute la séquence protéique, certaines MT présentent des regroupements de motifs en Cys que l'on appelle modules (**Figure I.8**).

Par exemple, les MT de plantes, de *Fucus vesiculosus* (Straménopile), de *Yarrowia lipolytica* (Champignon, Ascomycète) et de *Rhizophagus irregularis* (Champignon, Gloméromycète) ont une organisation plurimodulaire. Leurs Cys sont regroupées en modules soit dans deux régions C- et N-terminale (MT-1, MT-2, MT-3 de plantes, de *Fucus vesiculosus* et de *Rhizophagus irregularis*), soit dans trois régions C-, N-terminale et centrale (plantes, sous-famille MT-4), soit dans deux régions N-terminale et centrale (*Yarrowia lipolytica*). Dans le cas des MT de plantes qui ont été bien étudiées, ces régions sont séparées par une région intermodulaire appelée *linker* qui est hydrophobe et possède des acides



aminés aromatiques et une taille variable en fonction des MT. Par exemple, la taille du *linker* varie entre 30 et 50 acides aminés pour MT-1, MT-2 et MT-3 et entre 12 et 14 acides aminés pour MT-4 (**Figure I.8B et I.8C**) (**Domènech et al. 2006 ; Freisinger 2008**).

Chez les métallothionéines de plus grandes tailles retrouvées chez les organismes *Tetrahymena thermophila* (Alvéolé, Ciliés), *Trichomonas vaginalis* (Excavata, Métamonade), *Cryptococcus neoformans* (Champignon, basidiomycète) et *Tremella mesenterica* (Champignon, basidiomycète), les acides aminés sont organisés en modules répétés qui peuvent être alignés entre eux (**Figure I.8D**). Cette organisation en modules répétés en plus de la taille plus grande que la moyenne semble aller dans le sens de l'hypothèse de l'origine chimérique de ces MT due à des duplications géniques (**Gutiérrez et al. 2011**).



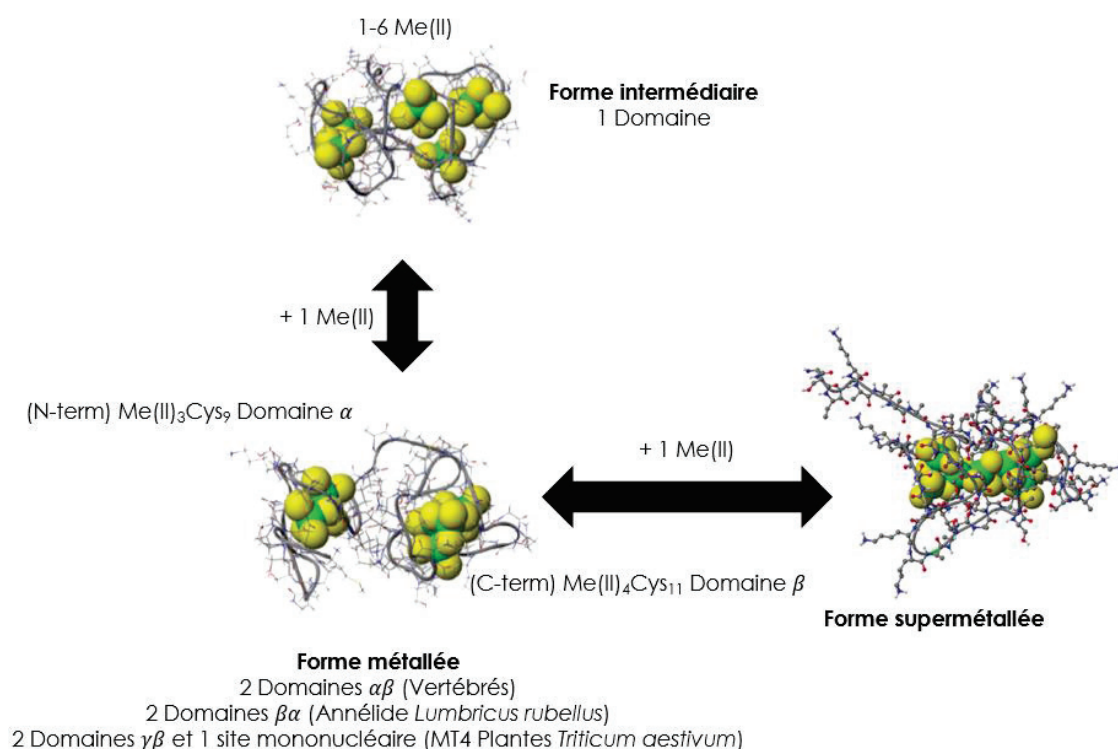


Figure I.9 : Diversité des structures tertiaires de métallothionéines liées aux métaux. Les MT ont plusieurs types de structure tertiaire en fonction du degré de métallation. La forme dominante des MT est la forme intermédiaire partiellement métallée. Les formes intermédiaires et supermétallées ne se composent que d'un seul domaine, alors que la forme métallée se compose en général de 2 domaines. Comme indiqué, plusieurs types de formes métallées peuvent être rencontrés.

Source images : rhMT1 (humains) (Sutherland & Stillman 2014 © 2014, Royal Society of Chemistry) .



2. Structures secondaire et tertiaire

En l'absence de métal, l'apoprotéine (protéine sans métal) ne forme pas de véritable structure secondaire ou tertiaire. Cependant, des études ont montré qu'en solution aqueuse, elle adopte une conformation compacte exposant les Cys à l'extérieur, les rendant plus facilement accessibles aux ions métalliques. A pH acide, l'apoprotéine perd cette pseudo-organisation et est beaucoup plus relâchée (**Sutherland & Stillman 2014**).

Les MT de plantes qui ont une organisation plurimodulaire ont localement des structures secondaires de type feuillets β au niveau des *linkers*. Les fonctions hypothétiques de ces régions intermodulaires pourraient être : 1) "l'étiquetage" pour un adressage vers d'autres organelles ou pour l'ancrage à des membranes, 2) la protéolyse ou 3) favoriser le repliement de la protéine vers une conformation particulière (**Freisinger 2008**).

C'est lorsque les MT se lient aux métaux, qu'elles acquièrent une structure tertiaire organisée qui va dépendre de la nature et du nombre d'atomes de métaux liés. Selon le nombre d'ions métalliques liés, on peut observer deux formes de protéines : soit en deux domaines, soit en un seul domaine (**Figure I.9**). Les métaux divalents (Me^{2+}) comme le Zn^{2+} et le Cd^{2+} vont pouvoir se lier à 4 Cys (on parle dans ce cas de coordination tétragonale) alors que les métaux monovalents (Me^+) comme le Cu^+ ou l' Ag^+ vont se lier à 2-3 Cys (coordination di- ou trigonale). Cela a deux conséquences: 1) les MT sont susceptibles de se lier à plus de Me^+ que de Me^{2+} et 2) la structure tridimensionnelle va dépendre du type et du nombre de métaux liés.

Les MT liées aux métaux mono- ou divalents peuvent adopter une structure soit en deux domaines globulaires, soit en un seul domaine (**Figure I.9**). La structure secondaire paradigmatique des MT est celle des mammifères. Par exemple, la protéine MT-2A du lapin possède deux domaines : un domaine bêta (β) et un domaine alpha (α) séparés par un très court *linker* de 3 acides aminés qui semble très important pour la liaison aux métaux (**Chan et al. 2002**).





Mais on peut rencontrer d'autres structures en deux domaines avec par exemple, la MT de l'annélide *Lumbricus rubellus* qui possède deux domaines inversés par rapport aux MT de Mammifères ($\alpha\beta \blacktriangleright \beta\alpha$) et la MT-4 de *Triticum aestivum* (Plante) qui possède deux domaines $\gamma\beta$ avec en plus un site de liaison à un métal.

D'autres MT sont retrouvées en un seul domaine comme Cup1 de *Saccharomyces cerevisiae* (**Atrian & Capdevila 2013**) ou encore SmtA de la bactérie *Synechococcus* sp. avec un domaine en doigt de zinc faisant intervenir 2 His et 2 Cys dans le cluster métallique. En présence de beaucoup de métaux, il a été observé que des MT en deux domaines adoptent une structure protéique en un seul domaine, contractée, leurs permettant de lier un métal de plus et se retrouvent en un seul domaine, sous une forme « supermétallée » (Remarque : métallation = processus de liaison entre une protéine et des métaux) (**Sutherland & Stillman 2014**). Néanmoins, étant donné que très peu de structures en 3 dimensions de métallothionéines sont disponibles et qu'elles ont été obtenues à partir d'organismes phylogénétiquement proches, il est possible que la structure paradigmatique en deux domaines ($\alpha\beta$) ne soit pas la plus courante.

L'intérêt fonctionnel de cette structure en deux domaines n'est pas clair. Il est possible que chaque domaine ait une affinité différente pour un métal et/ou qu'un domaine permette la liaison spécifique à certaines métalloprotéines dont l'activation nécessite le transfert de métaux (**Liu et al. 2000**).



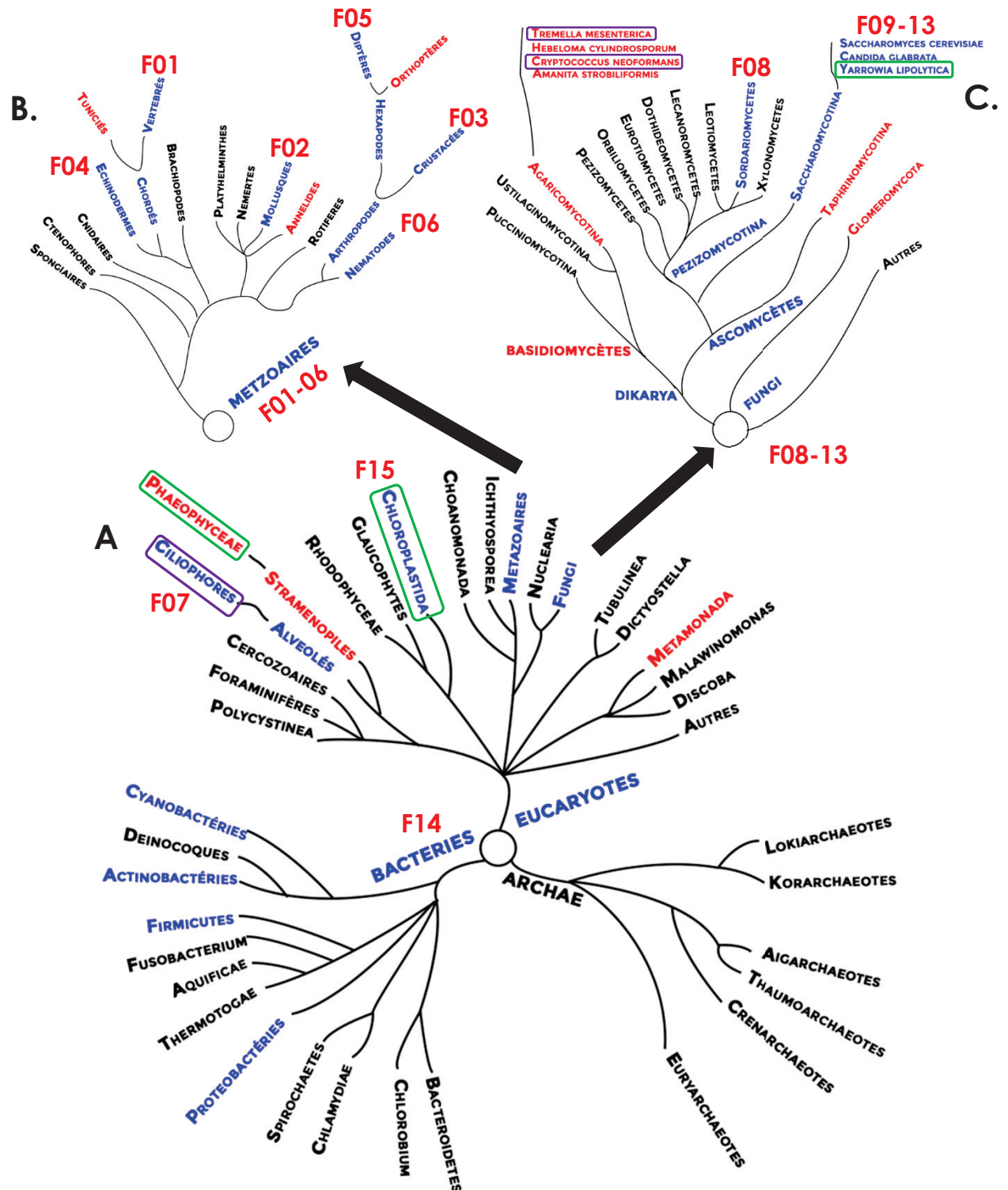


Figure I.10 : Distribution des métallothionéines au sein de l'arbre de vie. En **A.** sont représentés les principaux clades d'organismes vivants, en **B.** les principaux clades de métazoaires et en **C.** les principaux clades de champignons. En bleu, les clades dans lesquels les MT ont été affiliées à une famille. Les numéros de familles de MT sont indiqués en rouge (Binz & Kagi 1999). En rouge, les clades dans lesquels des MT ont été identifiées, mais ne peuvent être affiliées à une famille spécifique. Entourées en vert, les protéines avec une organisation en Cys pluri-modulaire et en violet une organisation en modules répétés.



B. Diversité et distribution des métallothionéines

Les structures primaires (séquences en acides aminés) des MT sont très diversifiées en termes de tailles, de motifs en Cys et d'organisation de ces motifs. Ces structures sont néanmoins à l'origine de la capacité des MT à se lier aux métaux. Si on considère que chaque cellule et chaque compartiment cellulaire peuvent être considérés comme un environnement métallique particulier, ou " nano-environnement ", cette grande diversité de séquences de MT, pourrait être liée à l'adaptation de chacune de ces séquences aux concentrations cellulaires en métaux qui diffèrent d'un organisme à l'autre et d'un compartiment cellulaire à l'autre. Il semble donc que l'histoire évolutive des organismes qui possèdent les MT ainsi que les variations locales de la concentration en métaux pourraient être des facteurs déterminants associés à la diversité et à la distribution spatiale et temporelle des séquences de MT (Capdevila & Atrian 2011).

• Origine et distribution taxonomique des MT

Les métallothionéines sont principalement retrouvées parmi les organismes eucaryotes, notamment parmi les vertébrés qui regroupent près de 82% des 722 séquences de MT déposées dans Swiss-Prot, ou parmi les plantes (9% des séquences) (Figure I.10). Les MT ne sont pas toutes homologues entre elles et les niveaux taxonomiques à partir desquels elles sont homologues entre elles sont très hétérogènes et dépendent des MT considérées. Elles peuvent être homologues au niveau du domaine comme par exemple BmtA chez les bactéries, au niveau d'embranchements comme les MT de vertébrés ou d'espèces comme certaines MT de Champignons. Mais même si deux séquences de MT sont homologues, elles possèdent de fortes dissimilarités entre elles comme par exemple pour les 4 MT de *Yarrowia lipolytica* où MT-1 et MT-3 partagent 93% d'identité entre elles, MT-2 et MT-4 90% d'identité, mais entre ces deux groupes, sont observés seulement 42% d'identité (García et al. 2002).



Parfois cette différence est tellement importante qu'il est impossible de savoir si deux sous-familles de MT présentes au sein d'une même cellule sont issues d'un ancêtre commun ou si elles sont issues de phénomènes de convergence évolutive comme c'est le cas pour les séquences des sous-familles CuMT et CdMT de *Tetrahymena thermophila* avec 29% d'identité entre elles (séquences ABF68755 et AAP42952, alignement global, matrice de substitution Blossum90, algorithme de Needleman-Wunch).

Compte tenu de l'extrême hétérogénéité des séquences, il est impossible de déterminer si toutes les MT décrites ont une origine commune. L'hypothèse communément admise est qu'elles sont issues de multiples événements de convergence évolutive (**Capdevila & Atrian 2011**). Le besoin de chélateurs de métaux dans le métabolisme du Zn et du Cu pourrait avoir pour origine la grande oxygénation de la Terre, il y a 2,4 milliards d'années, suite à l'apparition et la prolifération des organismes photosynthétiques. Cette toute première pollution globale d'origine biologique a radicalement changé la surface de la planète, la faisant passer d'un milieu globalement réducteur à un milieu oxydant. Les conséquences ont été progressives et globales avec la précipitation du fer océanique et la solubilisation du Cu et du Zn sous formes ioniques (**Scott et al. 2013**). S'en sont suivis de multiples événements et processus évolutifs s'étalant sur des milliards d'années conduisant à des changements métaboliques importants en vue de la gestion et de l'utilisation de ces nouvelles sources métalliques biodisponibles ainsi qu'une grande diversification d'espèces d'organismes. Tout cela aurait conduit à l'apparition de protéines comme les MT permettant de maintenir l'homéostasie métallique cellulaire (**Capdevila & Atrian 2011**).



• Compartimentation des métallothionéines

Les différentes sous-familles et isoformes de MT présentent des répartitions différentes en fonction des tissus à l'échelle de l'organisme et en fonction des organelles au sein de la cellule qui ont une répartition en éléments qui leur est propre (stœchiométrie).

A l'échelle de l'organisme, s'observe une compartimentation à la fois spatiale et temporelle des différentes sous-familles et isoformes de MT. Par exemple, les plantes possèdent 4 sous-familles (MT-1, 2, 3 et 4) ayant chacune plusieurs isoformes. Chez *Arabidopsis thaliana*, les MT-1 sont plus exprimées dans les racines que dans les feuilles, MT-2 dans les feuilles et à un moindre degré dans les racines, MT-3 dans les fruits matures et MT-4 dans les graines. Chacune de ces MT présente donc des profils d'expression différents qui nous informent de leurs fonctions au sein des organes. Toutefois, les isoformes MT-1a et MT-2b sont exprimées dans les mêmes organes et se lient tous deux au Cu mais MT-1a répond fortement à l'augmentation de concentration de Cu alors que la transcription du gène MT-2b n'est pas induite en réponse à ce métal. Ces deux MT sont aussi induites dans des organes en senescence en réponse à la remobilisation des métaux. Ainsi, dans des tissus jeunes et adultes, MT-1a est impliquée dans la régulation de la concentration cellulaire en Cu et MT-2b dans le fonctionnement cellulaire basal, alors que dans les organes sénescents, toutes les deux seront impliquées dans le recyclage des métaux (**Guo et al. 2003 ; Leszczyszyn et al. 2013**). Un autre exemple de différenciation fonctionnelle au niveau de l'organisme est celui de l'insecte orthoptère *Oxya chinensis* (criquet), où les deux types de MT identifiées sont exprimés à tous les stades de développement et dans tous les tissus mais leur expression est nettement plus forte dans les œufs et dans le système nerveux central de l'adulte (**Liu et al. 2014**).

A l'échelle de la cellule, les métallothionéines (Vertébrés) peuvent être retrouvées dans la plupart des compartiments intracellulaires : cytosol (**Filipović & Raspor 2003**), noyau (**Tsujikawa et al. 1991**), mitochondries (**Costello et al.**



2004 ; Maret et al. 2002) ou lysosomes (**Lerch et al. 1985**). Dans ces compartiments, la quantité de MT est souvent proportionnelle à la quantité de métaux (**Filipović & Raspor 2003**). De plus, les MT ne sont pas nécessairement cantonnées à un tissu ou un compartiment cellulaire. Elles sont mobiles en fonction des besoins en métaux des différents compartiments. Chez les mammifères, les MT peuvent être excrétées et se retrouver dans plusieurs fluides extracellulaires (sang, fluides cérébraux spinaux) (**Lynes et al. 2006**). Dans les cellules, les MT peuvent migrer du cytosol vers le noyau en réponse à des stress oxydatifs et métalliques (**Cherian & Apostolova 2000**).

• Régulation des MT

La compartimentation des différentes isoformes de MT permet la mise en place de mécanismes de régulation de l'expression des gènes permettant de répondre aux variations intracellulaires locales de la stœchiométrie en métaux. Les MT peuvent être exprimées de manière constitutive et/ou en réponse à plusieurs types de stress et de molécules signal. La régulation de l'expression des MT peut intervenir au niveau transcriptionnel, avec l'implication éventuelle de mécanismes épigénétiques (eucaryotes), post-transcriptionnel, traductionnel et post-traductionnel.

La plupart des informations disponibles concernent la régulation transcriptionnelle des MT avec des études relativement anciennes. Dans la région 5' promotrice, les MT peuvent présenter un grand nombre de séquences régulatrices sur lesquelles se fixent plusieurs facteurs de transcription. Certaines de ces séquences sont retrouvées en plusieurs exemplaires (**Stuart et al. 1984**). On retrouve par exemple les motifs MRE (*Metal Responsive Elements*), ARE (*Antioxidant Response Element*), GRE (*Glycocorticoids Response Elements*), CRE (*cAMP Response Element*), TRE (*TPA Response Element*) et IRE (*Interferon Response Element*) dont la présence dépend des organismes, des compartiments et des isoformes (**Hager et al. 1981 ; Nebes et al. 1988**). Cette



diversité de séquences régulatrices va permettre aux MT de répondre à un grand nombre de stimuli et permettre à chaque isoforme d'avoir une activité spécifique. Par exemple, les MT-1 et MT-2 humaines sont présentes à un niveau basal dans presque tous les tissus. En présence de Cd et de Cu, la transcription de ces deux séquences augmente de façon différentielle : MT-1 est induite 10 fois en présence de Cd et de Cu et MT-2 est induite 3 fois en présence de Cd et 30 fois en présence de Cu (**Vasconcelos et al. 1996**). Ces motifs régulateurs dans la région promotrice fonctionnent en association avec différents facteurs de transcription. Par exemple, les motifs MRE fonctionnent en association avec les facteurs de transcription de type MTF (*Metal regulatory Transcription Factor*) qui acquièrent une structure active en présence de métaux (**Heuchel et al. 1994**).

La transcription des MT peut aussi être modulée par des modifications épigénétiques de la région promotrice. Ainsi, chez les mammifères, la région promotrice est riche en îlots CpG pouvant être méthylés et en sites de liaison d'histones qui peuvent être acétylés entraînant un relâchement de la chromatine. Cette régulation est directement liée à l'activité de deux enzymes : l'ADN méthyltransférase (Dnmt) et l'Histone Désacétylase (HDAC). Plus ces enzymes sont actives et moins la transcription de MT-1 humaine est importante (**Ghoshal et al. 2002 ; Imoto et al. 2010**).

Une fois la transcription induite, des mécanismes d'épissage alternatif en fonction des tissus peuvent parfois produire différentes isoformes comme pour MT-1G humaine (**Mehus et al. 2014**). Les ARNm peuvent aussi présenter des demi-vies variables et être sélectionnés et traduits préférentiellement par les ribosomes, en fonction des isoformes et même du type de métal inducteur. Mais le fonctionnement précis de ces mécanismes est encore inconnu et reste peu étudié (**Hui et al. 1987 ; Sadhu & Gedamu 1989 ; Vasconcelos et al. 1996**).





Une fois les ARNm traduits en thionéines (métallothionéines sans métal), ces protéines vont pouvoir subir des modifications post-traductionnelles liées à leur fonctionnement cellulaire. Elles vont pouvoir se lier à des nombres et des types de métaux variables en fonction des conditions cellulaires, mais aussi être oxydées par la présence d'espèces réactives de l'oxygène. Si les protéines natives sont plus sensibles à la protéolyse, les protéines oxydées sont au contraire plus récalcitrantes. La protéolyse des protéines liées aux métaux va dépendre du type et du nombre de métaux liés (**McKim et al. 1992 ; Klaassen et al. 1994**). Les degrés de métallation et d'oxydation des protéines doivent donc être pris en compte dans le renouvellement des MT.

La concentration en MT fonctionnelles est donc finement régulée. Si la régulation transcriptionnelle des MT a été très bien étudiée avec la caractérisation de nombreux stimuli à la fois généraux ou spécifiques pouvant induire leur transcription, leur régulation aux autres niveaux post-transcriptionnels est moins bien documentée à l'exception de quelques études réalisées chez l'Homme. Ces mécanismes de régulation semblent dépendre des organismes, des isoformes et de l'état de la cellule (concentration en métaux, potentiel redox, etc...).



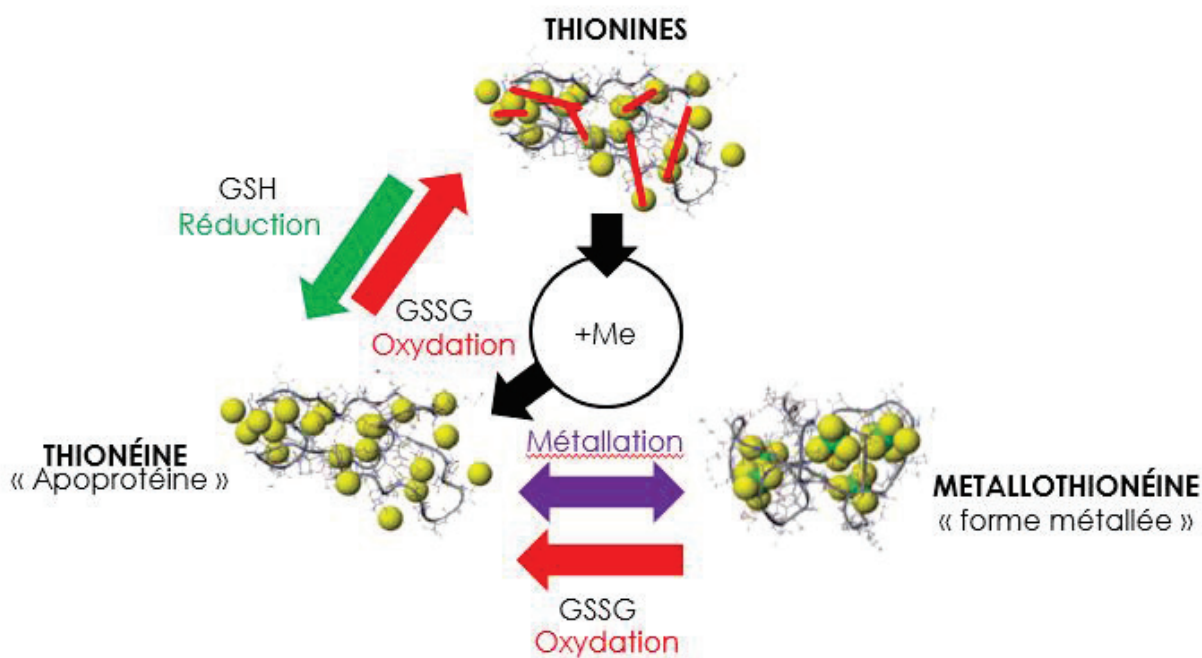


Figure I.11 : Trois formes de protéines sont trouvées dans les cellules : Les thionéines, formes réduites avec une forte affinité pour les métaux ; Les thionines, formes oxydées, avec des ponts disulfures qui se lient peu aux métaux ; Les métallothionéines, thionéines liées aux métaux. Les MT ont plusieurs types de structures tertiaires en fonction du degré de métallation. L'association au glutathion permet de réduire les thionines. Le système est particulièrement sensible à l'oxydation.

Source images : rhMT1 (humains) (Sutherland & Stillman 2014 © 2014, Royal Society of Chemistry)

Adapté de : Bell & Vallee 2009



C. Diversité des formes de métallothionéines

Si les métallothionéines ont des distributions et des modalités de régulation variées à l'échelle d'une cellule ou d'un organe, leur variabilité individuelle (à l'échelle d'une seule protéine) est très peu abordée. Ci-dessous nous allons voir que les MT sont des protéines présentent sous plusieurs formes dans les compartiments cellulaires ce qui leur permet d'interagir avec beaucoup d'acteurs métaboliques différents ce qui semble expliquer leur implication dans beaucoup de processus cellulaires.

• Trois principales formes de MT dans les cellules en fonction de l'équilibre redox et métallique

Dans la cellule, les métallothionéines (MT) vont être retrouvées sous trois formes moléculaires (**Bell & Vallee 2009**) (**Figure 1.11**) : 1) les thionéines (T), formes réduites pouvant se lier à des espèces chimiques réactives comme des métaux oxydés ou des ROS ; 2) les thionines (TO), formes oxydées qui réagissent peu ; 3) et les métallothionéines (MT), formes métallées pouvant se lier à des espèces chimiques plus réactives que celles déjà liées à des éléments. Les équilibres entre ces trois formes dépendent des ratios de synthèse/dégradation des protéines, de la quantité de métaux disponibles et des conditions redox de la cellule (**Krezel et al. 2007**). Les thionéines correspondent à l'apoprotéine alors que les métallothionéines correspondent à l'holoprotéine (protéine ayant adoptée une structure 3D et qui est fonctionnelle) lorsqu'elles sont totalement métallées.

La quantité cellulaire de chacune de ces trois formes est peu connue mais il semblerait que les thionéines (T) peuvent être en quantité non négligeable comme par exemple dans le foie, les reins et le cerveau de rat où les T sont en même quantité que les métallothionéines (MT). Ce ratio peut varier en fonction des tissus et des cellules (**Maret et al. 2002**) permettant de tamponner le milieu intracellulaire en métaux, en particulier en Zn^{2+} chez les



eucaryotes, et en espèces chimiques réactives telles que des radicaux libres. Les quantités relatives de chacune de ces trois espèces de protéines pourraient jouer un rôle essentiel dans les processus cellulaires et donner des informations sur l'état métallique et redox du nano-environnement cellulaire.

• Interactions avec les métaux

Les thionéines (T) vont se lier aux métaux pour devenir des MT. Elles se lient aux métaux *via* leurs atomes de soufre (Cys) et éventuellement leurs atomes d'azote (His) par une liaison covalente de coordination (dative) en plusieurs points entraînant un repliement autour du métal (chélation). Aussi, les T n'adoptent une structure 3D qu'une fois liées à un métal (**Duncan & Stillman 2006**). Il a été démontré *in vitro* que des MT sont susceptibles de chélater 18 espèces métalliques différentes mais la stabilité de la liaison est généralement maximale pour le Zn et le Cu selon la série des métaux divalents de Irving-Williams : $Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} < Zn^{2+}$ (**Irving & Williams 1948**). Concernant le cuivre, on va surtout retrouver du Cu^+ lié aux MT. En effet, la forme Cu^{2+} présente une plus forte affinité pour les molécules possédant des groupements amines (N) alors que le Cu^+ a plus d'affinité pour les thiols (S). De plus le Cu^+ est plus abondant dans un environnement réducteur comme c'est aussi le cas pour les T et MT. Plusieurs espèces métalliques peuvent être retrouvées liées à une même MT (**Bofill et al. 2009**) entraînant à l'échelle cellulaire une grande diversité de MT métallées car chacune d'entre elles se lie potentiellement à des nombres et des types de métaux différents (**Sutherland et al. 2012**).

Cette liaison avec les métaux est sensible à l'oxydation, qui entraîne la séparation des métaux et des atomes de soufre.



• Interactions avec les espèces réactives de l'oxygène et de l'azote

En présence d'espèces réactives : de l'oxygène (ROS), tels que le peroxyde d'hydrogène (H_2O_2) ou l'anion superoxyde ($\cdot O_2^-$), et de l'azote (NOS), tels que le monoxyde d'azote ($\cdot NO$), les métallothionéines (MT) vont s'oxyder pour perdre progressivement leurs métaux et devenir des thionéines (T), puis les cystéines vont former des ponts disulfures entre elles et donner des thionines (TO). Bien entendu, certaines MT peuvent avoir quelques soufres oxydés et les autres libres ce qui ne l'empêchera pas de se lier à des métaux mais réduira sa capacité maximale de métallation. L'oxydation des MT a ainsi pour conséquence au niveau cellulaire l'augmentation de la concentration en ions métalliques libres.

Mais cette oxydation n'est pas irréversible (**Figure I.11**), elle est en effet régulée par le glutathion. La forme réduite du glutathion (GSH) va permettre la réduction des TO en T tandis que la forme oxydée du glutathion (GSSG) va permettre la libération et le transfert de métaux par oxydation des métallothionéines (**Jiang et al. 1998 ; Bell & Vallee 2009**). Les MT oxydées vont ainsi pouvoir être régénérées et de nouveau disponibles pour se lier à des métaux ou des ROS. Il apparaît ainsi que les systèmes [GSH:GSSG] et [T:MT:TO] sont interconnectés et jouent des rôles importants dans la régulation des équilibres métalliques et rédox au sein de la cellule.

• Interactions avec les métalloprotéines

Dans la cellule, les éléments du système [T:MT:TO], que nous appellerons le « système thiol », peuvent former des dimères ou se lier à d'autres protéines pour échanger des ions métalliques.

Les MT peuvent se lier entre elles selon deux modalités : des ponts disulfures ou des ponts métalliques. Dans un environnement oxydant, les MT forment des ponts disulfures intramoléculaires (thionines) mais aussi intermoléculaires, ce qui a pour conséquence la formation de dimères



(**Zangger et al. 2001**). Par ailleurs, un métal peut établir plusieurs liaisons avec des atomes de soufre et lorsque ce métal se lie à des cystéines de deux MT différentes, on a la formation d'un pont métallique et d'une protéine dimérique (**Palumaa et al. 1992**).

Le système thiol interagit potentiellement avec toutes les métalloprotéines qui ont besoin de métaux pour leur repliement ou leur activité comme les facteurs de transcription ou les métallo-enzymes. Ci-dessous, trois cas sont détaillés : l'interaction Zn-MT et ferritines qui est cytotoxique, l'activation de la Cu/Zn-superoxyde dismutase par l'interaction avec une Cu-MT et l'échange de Zn entre une Zn-MT et les facteurs de transcription en doigt de zinc MTF permettant leur activation/inactivation.

Dans le milieu intracellulaire, l'homéostasie du Fe est essentielle pour le fonctionnement des cellules. Cependant, le Fe^{2+} est très toxique en présence d'oxygène car il entraîne la formation de ROS. La ferritine est une protéine assurant le stockage du fer sous forme Fe^{3+} non toxique. La liaison entre la ferritine et les MT entraîne la réduction du Fe^{3+} en Fe^{2+} qui est libéré et conduit à la formation de ROS. Les MT s'oxydent devenant des thionines (TO) et libèrent du Zn sans pouvoir se lier aux métaux libres à cause des ponts disulfures. Cela provoque la formation de ROS en chaîne ainsi que la libération de Zn^{2+} perturbant le métabolisme. Il semble que ce phénomène se produise lors de dysfonctionnements métaboliques (**Orihuela et al. 2011**).

La Cu/Zn-SOD est une enzyme impliquée dans la détoxification des ROS en transformant les anions superoxyde en peroxyde d'hydrogène et oxygène. Pour fonctionner, cet enzyme a besoin de Cu et de Zn. Les Cu-MT peuvent donner du cuivre aux apo/Zn SOD seulement en condition oxydante ou en présence de NO. Or la Cu/Zn SOD est requise pour lutter contre le stress oxydatif. Ainsi les CuMT interviennent dans la régulation du stress oxydatif en activant les Cu/Zn SOD (**Liu et al. 2000**).



Les MTF sont des facteurs de transcription en doigt de zinc régulant la transcription de la métallothionéine MT-I des mammifères. La liaison au Zn leur permet d'adopter un repliement en doigt de zinc sans lequel leur liaison à l'ADN est impossible. En présence de Zn libre, les MTF vont activer la transcription de thionines (T) qui vont pouvoir se lier au pool de Zn libre. Une fois ce pool réduit, les T et les MT qui restent vont se lier aux MTF pour prendre le Zn et les inactiver. On a donc en fonction des conditions (oxydantes, concentration en métaux, ratio MT:T:TO) une régulation de la transcription des MT et autres protéines dépendantes du Zn via l'activation des facteurs de transcription MTF (**Heuchel et al. 1994 ; Andrews 2000**).

• **Conséquences des interactions sur les processus cellulaires.**

Les propriétés de liaison du système thiol avec les métaux, les ROS, et les différentes métalloprotéines en font un système central au niveau des processus cellulaires de base chez les eucaryotes. Ce système intervient : (i) dans la régulation des équilibres redox en tant qu'antioxydant, (ii) dans la régulation des équilibres métalliques en tant que métallochaperonnes (iii) dans la signalisation des déséquilibres redox et métalliques en tant que transducteurs de signaux via la libération d'ions métalliques, (iv) dans la détoxification cellulaire face au surplus de métaux et de radicaux libres en tant que chélateurs intracellulaires, (v) dans la régulation de la transcription de gènes et de l'activité de métalloprotéines via l'échange de métaux. Toutes ces fonctions sont interdépendantes et si l'une est modifiée, les autres le seront également (**Giles et al. 2003 ; Maret 2005 ; Sutherland et al. 2012**).



A.

Family 1: vertebrate MTs
 m1: mammalian MT-1
 m2: mammalian MT-2
 m3: mammalian MT-3
 m4: mammalian MT-4
 m: n.d. mammalian MT
 a1: avian MT-1
 a2: avian MT-2
 a: n.d. avian MT
 b: batracian MT
 t: teleost MT

Family 2: mollusc MTs
 mo1: mussel MT-1
 mo2: mussel MT-2
 mog: gastropod MT
 mo: n.d. mollusc MT

Family 3: crustacean MTs
 c1: crustacean MT-1
 c2: crustacean MT-2
 c: n.d. crustacean MT

Family 4: echinodermata MTs
 e1: echinodermata MT type 1
 e2: echinodermata MT type 2

Family 5: diptera MTs
 d1: diptera MT type 1
 d2: diptera MT type 2

Family 6: nematoda MTs
 n1: nematoda MT type 1
 n2: nematoda MT type 2

Family 7: ciliata MTs
 ci: ciliata MT

Family 8: fungi-I MTs
 f1: fungi-I MT

Family 9: fungi-II MTs
 f2: fungi-II MT

Family 10: fungi-III MTs
 f3: fungi-III MT

Family 11: fungi-IV MTs
 f4: fungi-IV MT

Family 12: fungi-V MTs
 f5: fungi-V MT

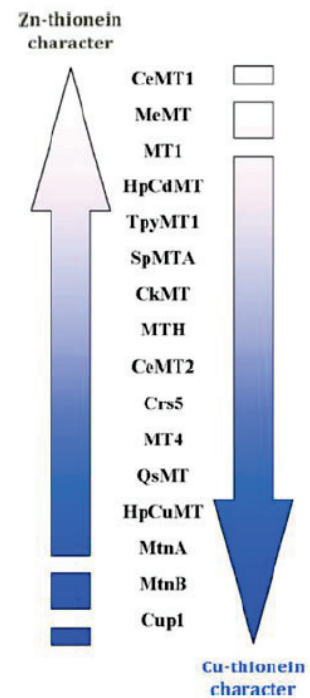
Family 13: fungi-VI MTs
 f6: fungi-VI MT

Family 14: prokaryota MTs
 pr: prokaryota MT

Family 15: planta MTs
 p1: plant MT type 1
 p2: plant MT type 2
 p2v: plant MT type 2 variant, described as a clan of p2
 p3: plant MT type 3
 p21: plant MT type 2x1
 pec: plant EC MT-like protein

Family 99: phytochelatins and other non-proteinaceous MTs

B.



C.

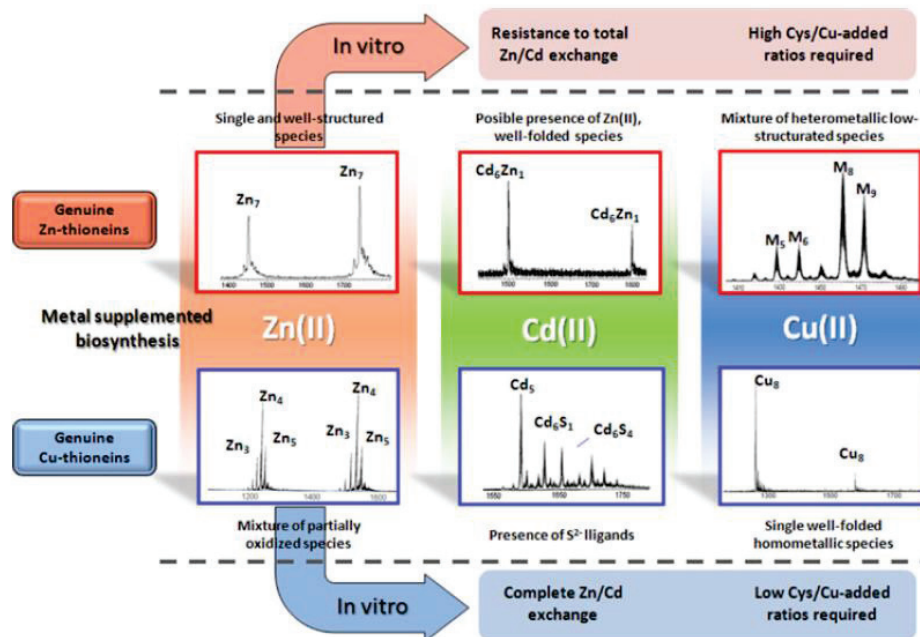


Figure I.12 : Classifications des Métaallothionéines **A.** Classification phylogénétique (Binz & Kagi 1999). Cette classification est construite suite à des analyses phylogénétiques à partir des séquences d'acides aminés et d'acides nucléiques. Elle comporte 15 familles différentes à l'intérieur desquelles on trouve des subdivisions en sous-familles. **B.** Gradient de MT selon leur « caractère » de Zn-thionéine ou Cu-thionéine. **C.** Classification fonctionnelle (Palacios *et al.* 2011b). Cette classification est obtenue après analyse par spectrométrie de masse des produits de la synthèse de MT recombinantes en présence de Zn, Cd et Cu et des remplacements de métaux *in vitro*.

Source images : © 2011, SBIC



D. Classifications des métallothionéines

Les métallothionéines peuvent être classées selon leur similarité de séquences (acides aminés et acides nucléiques) traduisant des relations de parenté entre les protéines et selon leurs propriétés de liaison aux métaux traduisant des fonctionnalités communes.

• Classification phylogénétique basées sur la similarité de séquence

La classification des métallothionéines basée sur la similarité des séquences est la plus ancienne des deux et traduit des relations de parenté entre les différentes séquences de MT qui sont homologues entre elles (**Binz & Kagi 1999**). Cette classification sépare les MT en 15 familles différentes. Au sein d'une même famille, les séquences de MT sont homologues entre elles avec des subdivisions en différentes sous-familles (**Figure I.12A**). L'analyse phylogénétique de chacune des familles a été construite à partir d'alignements multiples de séquences protéiques et nucléiques en incluant, lorsque cela était possible, les régions 5'-UTR contenant des séquences régulatrices de la transcription des gènes de MT. Ces différentes familles ont été placées sur l'arbre phylogénétique du vivant présenté **figure I.10**. Cette répartition est très hétérogène avec 6 familles différentes (F01-06) rien que pour l'embranchement des métazoaires, 1 pour les ciliés, 6 pour les champignons ascomycètes, 1 pour les plantes et 1 pour les procaryotes, sans compter les séquences qui ne sont affiliées à aucune famille comme celles des champignons basidiomycètes. A noter qu'aucune analyse phylogénétique entre toutes les MT ne peut être réalisée étant donné que les familles de MT ne sont pas homologues entre elles.



• **Classification fonctionnelle basée sur les propriétés de liaison aux métaux *in vitro*.**

La classification phylogénétique a été complétée par une classification fonctionnelle basée sur les propriétés de liaison des MT au Zn, Cu et Cd. Après purification des complexes [Me-MT], des analyses spectrométriques (ESI-MS) et spectroscopiques (dichroïsme circulaire, ICP-OES, spectrométrie UV-Vis) sont menées pour quantifier les espèces métalliques liées (**Palacios et al. 2011b**). De cette classification émerge le concept de « comportement » ou « caractère » où les MT peuvent être placées sur un gradient avec d'un côté les protéines ayant une forte affinité pour les métaux monovalent (Cu-thionéines) et de l'autre les MT ayant une forte affinité pour les métaux divalents (Zn-thionéines) alors que certaines MT présentent une affinité intermédiaire (**Figure I.12B et C**).

- Les Zn-thionéines : en présence de Zn ou de Cd, les Zn-thionéines forment un complexe [Me²⁺-MT] majoritaire avec un seul type de métal lié. La protéine se replie bien d'après les données de dichroïsme circulaire. En présence de Cu, les Zn-thionéines forment des complexes hétérométalliques [Cu,Zn-MT] avec moins de métaux liés dû à l'oxydation et la formation de ponts disulfures ayant pour conséquences un repliement incomplet de la protéine.
- Les Cu-thionéines : lorsqu'on analyse le nombre de Zn et de Cd liés aux Cu-thionéines, on observe la formation de plusieurs complexes [Me²⁺-MT] sans complexe dominant, la présence de S²⁻ et Cl⁻ liés au complexes [Cd-MT] et un repliement incomplet de la protéine. En présence de Cu, les Cu-thionéines forment des complexes homométalliques [Cu-MT] avec très peu de Zn et un repliement complet.



V. Discussion et enjeux.

A. La diversité microbienne des sols

Si le rôle des communautés de champignons des sols et des facteurs influençant leur diversité et leurs activités sont bien documentés (**Anderson & Cairney 2004 ; Barbi et al. 2016**), ce n'est pas le cas des eucaryotes unicellulaires (protistes) bien que ceux-ci représentent la composante la plus riche en espèces des organismes eucaryotes (**Adl et al. 2012 ; Bates et al. 2013**). Beaucoup de protistes participent activement à la régulation des populations de bactéries par prédation (**Matz & Kjelleberg 2005**) mais la structure globale des réseaux auxquels ils participent est encore obscure et reste à être étudiée à grande échelle avec l'aide des nouvelles technologies comme ce qui a été réalisé dans le milieu marin au travers du projet Tarra Océan (**Lima-Mendez et al. 2015 ; Guidi et al. 2016**).

Même si les progrès techniques et conceptuels liés à l'informatisation de la biologie, au *Big Data* et aux approches « -omiques » (**Nesme et al. 2016**) nous ont permis de connaître les communautés microbiennes des sols comme jamais auparavant, l'inventaire de la diversité des organismes présents est encore très loin d'être complet et tend à être remplacé par des pools d'amplicons ne permettant pas d'apprécier la diversité réelle de ces organismes sur plusieurs niveaux d'organisation. Ces révolutions techniques sont confrontées à un ensemble de biais technologiques et conceptuels entraînant une incertitude finale difficilement quantifiable (du prélèvement de sol aux assemblages de séquences).



B. La pollution métallique des sols

La pollution des sols par des éléments métalliques, est souvent problématique au niveau des sites anciennement industrialisés et des terres arables. En effet, l'existence de ces sites affecte la santé humaine tout en constituant des ressources économiques. Selon un rapport de l'ADEME (**La recherche / Sites et sols pollués / novembre 2014 / N°493 / 111**), la pollution métallique concerne 3 millions de sites en Europe et environ 350 000 en France. Cependant elle ne prend pas en compte la pollution diffuse, caractérisée par la présence de faibles concentrations en métaux lourds dans les sols à l'échelle globale depuis l'Antiquité (**Nriagu 1990**) et dont les proportions se sont considérablement accrues pendant les révolutions industrielles successives depuis le 18^{ème} siècle. Ajoutée à cela l'agriculture intensive qui augmente la concentration et la biodisponibilité des métaux dans les sols et les contamine en divers produits phytosanitaires, l'augmentation des transports mondiaux et l'accroissement de la diversité d'éléments chimiques rejetés dans les écosystèmes (nouvelles molécules, faible biodégradabilité) et nous avons les ingrédients d'une pression sur les systèmes vivants encore jamais rencontrée sur Terre. Par exemple, très peu d'informations sont disponibles sur les conséquences écologiques des pollutions aux « terres rares » de la famille des lanthanides, qui rentrent dans la composition d'une multitude de produits critiques pour les économies tournées vers le tertiaire. Ils sont en effet largement utilisés dans les technologies de l'information et de la communication et la mécanique de pointe (ordinateurs, smartphones, superalliages de l'aéronautique,...). La diversité microbienne des sols est la première affectée par les expositions métalliques. Les réponses allant de l'échelle cellulaire à celles des communautés sont très loin d'être toutes comprises. Par exemple, à ce jour aucune étude ne fait état des conséquences des pollutions aux « terres rares » sur la diversité microbienne des sols et les interactions de ces polluants émergents avec les autres polluants (hydrocarbures, micro-plastiques, nano-particules,...).



Les métaux sont aussi associés à des problèmes médicaux dépassant leur simple toxicité. Les bactéries ont la capacité de s'échanger des plasmides portant de multiples résistances (MDR), non seulement aux antibiotiques, mais aussi aux métaux (**Baker-Austin et al. 2006 ; Lemire et al. 2013**). Cela pourrait favoriser le maintien de gènes de résistance aux antibiotiques au sein de populations bactériennes présentes sur des sites contaminés aux métaux et leur diffusion vers des espèces pathogènes.

Enfin, dans le cas des études sur la réaction des communautés microbiennes face aux pollutions, nous sommes souvent amenés à comparer des études très différentes en tous points : sols différents, méthodes de mesure de la diversité différentes, espèces métalliques considérées différentes, quantités, durées d'exposition, volumes prélevés différents,... ce qui rend les résultats difficilement comparables et généralisables et pourrait expliquer parfois des effets contradictoires de certains métaux.

Comprendre quels réseaux écologiques et métaboliques sont touchés par les apports en métaux et l'histoire de leurs relations avec les cellules, c'est potentiellement pouvoir éviter des problèmes à la fois de santé publique, écologiques, sociaux et économiques à venir avec par exemple la dispersion d'organismes pathogènes résistants aux antibiotiques, la déstabilisation des fonctions écologiques des sols à long terme et l'intoxication par ingestion de denrées alimentaires contaminées. Au final, les pollutions métalliques sont des perturbations caractérisant l'anthropocène qui associées aux autres stress pourraient faire basculer les systèmes sols vers des états dont les conséquences sont inconnus et probablement catastrophiques pour les sociétés humaines actuelles (**Röckstrom et al. 2009**). En effet, en prenant une perspective globale propre aux sciences du système Terre, tous les cycles biogéochimiques (C, N, P, S,...) sont connectés aux cycles biogéochimiques des métaux car tous les organismes utilisent des enzymes à cofacteurs enzymatiques pour leur métabolisme ce qui limite leur écologie. Ainsi, en modifiant les cycles



biogéochimiques globaux des métaux, on affecte tous ces autres processus (**Benjamin & Honeyman 2000 ; Kim et al. 2013**).

C. Les perturbations métaboliques induites par les métaux

La stoechiométrie des métaux bio-disponibles dans l'environnement est une propriété évoluant à l'échelle des temps géologiques. A l'échelle cellulaire, cette stoechiométrie et la composition en éléments métalliques et en ligands sont finement régulées et peuvent être considérées comme des traits fonctionnels des organismes. En changeant les métaux disponibles, les activités humaines favorisent des organismes adaptés à ces nouvelles conditions et entraînent des modifications à long terme des génomes aux conséquences imprévisibles.

Il semble que beaucoup de molécules et de stratégies soient communes aux mécanismes de maintien de l'homéostasie de base, de réponse aux phénomènes de carence et de résistance aux métaux comme par exemple l'utilisation de chélateurs de métaux. Certaines de ces molécules sont excrétées en cas de carence (par exemple les sidérophores), présentes dans les cellules pour gérer les métaux réactifs et induites en cas d'excès en métaux (**Merchant & Helmann 2012**). Aussi il est possible que la gestion des métaux fonctionne sur des combinaisons de « réseaux protéiques de réponses multiples » permettant de répondre très finement à chaque stress métallique.

Comprendre les réseaux métaboliques dans les cellules et les réseaux écologiques dans les sols est un enjeu majeur de la biologie des systèmes. Le recours à une combinaison de technologies de pointe comme les techniques de séquençage massif, de microfluidique et les outils de modélisation des systèmes semble inéluctable pour étudier ces réseaux afin de récupérer un nombre élevé de données et de les traiter. Cependant, ces techniques ne permettent pas de se passer de méthodes plus « anciennes ou traditionnelles »



et parfois plus fastidieuses permettant de préciser les caractéristiques de chaque élément de ces réseaux. Ainsi, il serait souhaitable d'identifier toutes les molécules et organismes annotés comme « inconnus » dans les bases de données.

L'environnement intracellulaire conditionne la réactivité et les fonctions des métaux dans les processus cellulaires. La diversité des ligands potentiels en est un point clef (**Maret 2005**). Cette diversité est difficilement quantifiable et largement inconnue et les données de la littérature ont été obtenues sur un nombre très limité d'espèces non représentatives de la diversité du vivant. Est-il raisonnable : (i) de généraliser les phénomènes étudiés sur un nombre limité d'espèces ? ; (ii) de les utiliser pour développer des tests simples de laboratoires, standardisés, faciles à mettre en œuvre et recommandés par l'OCDE (**Chapman et al. 1998 ; Calow & Forbes 2003**) qui *in fine* seront prises en compte par des « décideurs politiques » dont les décisions se basent sur des études d'analyses de risques et coût/bénéfices ?

D. Les métallothionéines

Les métallothionéines sont de petites protéines riches en cystéines très diversifiées. Les récentes études ont toutefois mis en lumière la présence de métallothionéines de grandes tailles potentiellement issues de la fusion de plusieurs métallothionéines de plus petites tailles. Ces MT sont très éloignées des MT décrites jusqu'alors en termes de taille et d'organisation augmentant ainsi l'hétérogénéité des séquences d'acides aminés connues au sein de ce groupe de protéines. Cette hétérogénéité de séquences n'est pas adaptée aux algorithmes d'annotation de nouvelles séquences. Ces protéines sont bien souvent retrouvées en tant que « protéines inconnues » ou « protéines riches en cystéines » dans les bases de données. Il serait intéressant de cribler les bases de données sans *a priori* en sélectionnant toutes les protéines riches en cystéines pour vérifier si elles représentent des MT potentielles.



Un des enjeux de la recherche sur les MT actuellement est de relier la composition et l'organisation des séquences primaires d'acides aminés à l'affinité pour un métal donné. Un des autres enjeux est de rééquilibrer les études expérimentales sur ces protéines dont la plupart portent sur des MT de mammifères et de plantes. En effet, en dehors des séquences d'organismes dits « supérieurs » (vertébrés, plantes), très peu d'autres séquences sont disponibles, et nous formulons l'hypothèse que la diversité de ces protéines est encore loin d'être décrite. L'étude de leurs structures tertiaires après liaison aux métaux pourrait donner plus d'informations sur leurs relations évolutives et les liens entre structure primaire et affinité pour un type de métal mais très peu de protéines ont été analysées de façon expérimentale. La diversité des MT au sein d'un organisme peut être grande et l'abondance relative entre ses différentes isoformes ainsi que leur compartimentation et leur expression semblent être soumises à une régulation fine à tous les niveaux. Dans les cellules, les MT se retrouvent sous trois formes dont le ratio pourrait avoir une grande importance dans les capacités d'interaction des MT avec des métaux, des espèces réactives de l'oxygène et de l'azote et des métalloprotéines. Il apparaît que les MT jouent un rôle essentiel dans la stabilité des processus des systèmes cellulaires en intervenant dans la résistance et la résilience des équilibres métalliques et oxydatifs.

Les thionéines (T), les métallothionéines (MT) et les thionines (TO) forment un système thiol fonctionnant de manière indissociable. Cependant, la majeure partie des études se concentre sur les MT et peu sur les T. De plus, ce système semble être connecté avec d'autres systèmes cellulaires impliqués dans la gestion des nano-environnements métalliques et redox notamment : 1) par l'interconnexion avec le système [GSH:GSSH] permettant le maintien de l'équilibre du système thiol, 2) la régulation de l'activité des facteurs de transcription métalliques (MTF) et 3) la régulation de l'activité des SOD. Une hypothèse serait que toutes ces molécules fonctionnent ensemble dans le contrôle de la résistance et la résilience des équilibres métalliques et redox des



systèmes cellulaires. Dans ces processus, le système thiol constitue un élément clef pour deux raisons : 1) il a été montré chez l'homme que tout dysfonctionnement de ce système peut entraîner de sévères maladies comme des cancers, du diabète, des problèmes de réactions inflammatoires et des maladies neurodégénératives (**Raudenska et al. 2014 ; Shamsi & Fatima 2014**) ; 2) chez les plantes et les mammifères, les MT sont induites en réponse à un grand nombre de molécules et de stress différents démontrant leur rôle central dans tous les processus cellulaires généraux : stress métallique, oxydatif, hydrique, thermique,...

Un des derniers enjeux concerne l'origine des MT. Elles semblent provenir de phénomènes d'adaptation des physiologies des organismes à des nano-environnements cellulaires riches en Zn et en Cu comme c'est le cas pour les eucaryotes qui possèdent beaucoup plus de métallo-enzymes à Zn et à Cu que les autres organismes (**Capdevila & Atrian 2011 ; Dupont et al. 2010 ; Decaria et al. 2011**). Il est possible que certains organismes, interagissant avec des eucaryotes au niveau cellulaire, aient acquis un système thiol leur conférant un moyen de survie dans ces microenvironnements cellulaires (concentration ionique, potentiel redox, pool de biomolécules). C'est le cas de certains champignons et bactéries pathogènes ou symbiotiques exprimant des MT en tant que facteurs de virulence pour conditionner leur survie dans leurs hôtes (**Gold et al. 2008 ; Ding et al. 2014**).

Nous sommes donc encore très loin de comprendre précisément le fonctionnement du système thiol ainsi que de connaître la diversité réelle des MT.





VI. Contexte et objectifs de la thèse

Au cours de la thèse de F. Lehembre au laboratoire d'Ecologie Microbienne de Lyon (**Thèse F.Lehembre 2009**), un ensemble de gènes capables de conférer à la levure *Saccharomyces cerevisiae* une résistance au cadmium (Cd) a été isolé en utilisant une approche de métatranscriptomique fonctionnelle. Cela a été réalisé à partir de sols pollués ou non provenant de divers sites de Belgique. Pour chacun de ces sols, des ARNm polyadénylés ont été directement extraits puis convertis en ADNc ; ces derniers ont été clonés pour constituer des banques d'ADNc de sol représentatives de la diversité des gènes exprimés par les différents organismes eucaryotes des sols considérés (**Lehembre et al. 2013**). Certains des gènes isolés présentent des particularités intéressantes qui justifient des études plus approfondies. Parmi ces gènes, certains se regroupent en une famille génique et présentent en commun plusieurs caractéristiques. 1) Ils n'ont aucun homologue dans les bases de données actuelles ; ils codent donc des protéines totalement nouvelles et proviennent de microorganismes potentiellement inconnus ou peu étudiés. 2) Ils présentent des séquences différentes mais les protéines codées présentent toutes certains motifs riches en cystéines, très conservés au sein de cette famille. Par ailleurs, ces protéines semblent être localisées dans le cytoplasme. En se basant sur ces caractéristiques, ces protéines pourraient correspondre à des métallothionéines. Ces gènes ont dans un premier temps été nommés *crp* pour « *Cysteine-Rich Protein* » (**Lehembre et al. 2013**). 3) Ces gènes ont pu être détectés par PCR à partir d'ADN métagénomiques extraits de divers sols et d'eaux douces d'origines climatiques et géographiques très variées. Les organismes porteurs de ces gènes seraient donc très largement distribués et pourraient jouer un rôle dans les mécanismes de résistance aux métaux des microorganismes eucaryotes qui les portent. 4) D'après leur analyse phylogénétique, leurs séquences protéiques se regroupent en 9 sous-familles inégales en nombre de séquences obtenues ainsi qu'en répartition en fonction des sites étudiés.



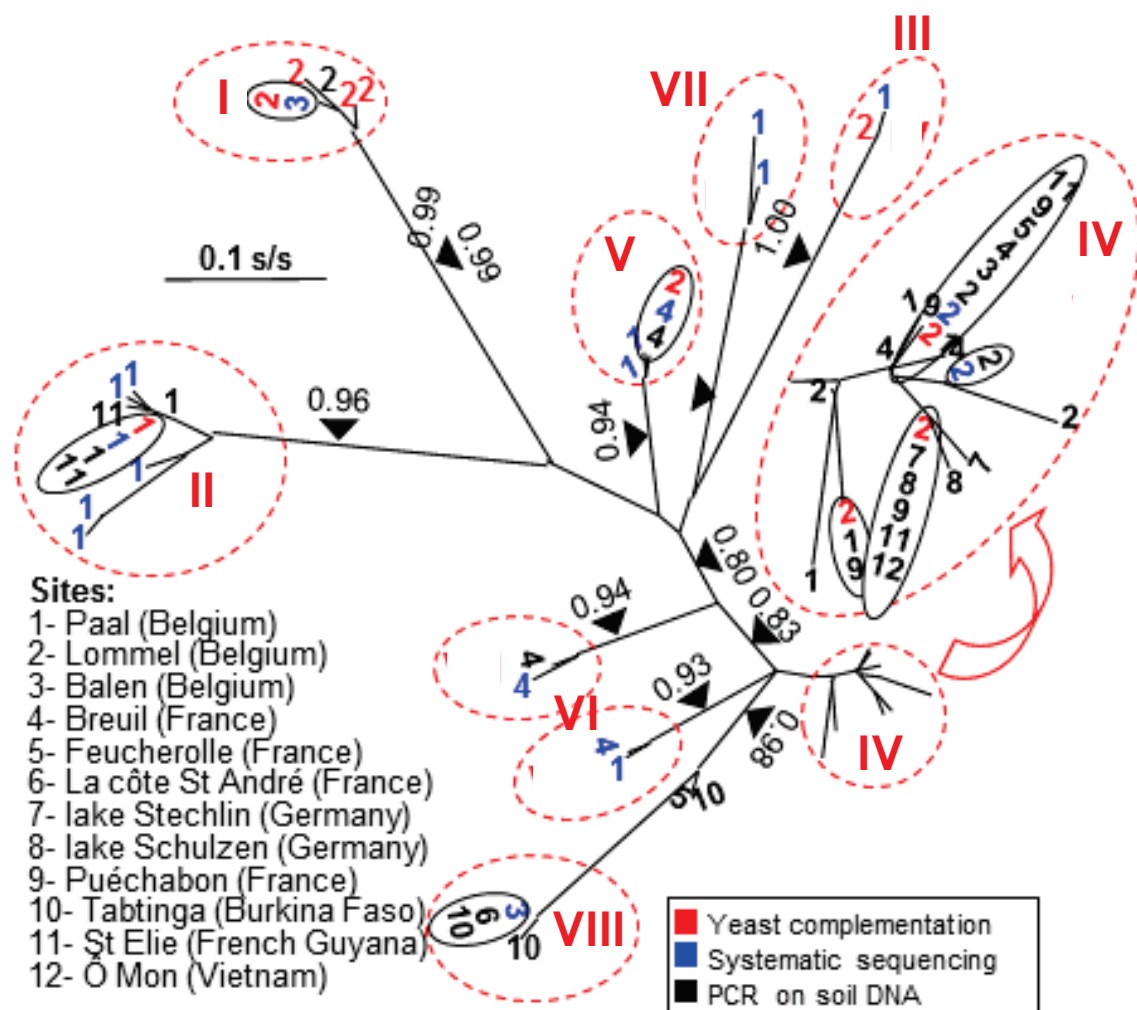


Figure I.13. Arbre phylogénétique (PhyML) des séquences de la famille des CRP. Les séquences ont été obtenues par : complémentation fonctionnelle (en rouge), séquençage de banques d'ADNc (en bleu), PCR sur des ADN métagénomiques extraits de sols et d'environnements aquatiques (en noir).

Source image : adapté de Lehembre et al. 2013



Par exemple, les sous-familles 2, 3, 4, 5, 6 et 7 sont retrouvées dans le sol de Paal qui est non pollué alors qu'on ne retrouve que les sous-familles 1, 3, 4 et 5 dans le sol de Lommel qui est situé près d'une ancienne fonderie de zinc (**Figure I.13**).

Ma thèse a eu pour objectif la caractérisation de cette famille de gènes pour laquelle trois questions ont été posées:

- **Ces protéines peuvent-elles être considérées comme des métallothionéines (Chapitre 2) ?**

En faisant l'hypothèse que ces protéines le sont et en appliquant un raisonnement déductif, nous avons appliqué les mêmes méthodes de caractérisation des métallothionéines putatives aux CRP. Plusieurs protéines CRP recombinantes ont été surproduites dans *Escherichia coli* puis purifiées. L'interaction de ces protéines avec le Cd, le Cu et le Zn a été analysée par des approches spectrométriques ce qui a permis de révéler que cette famille de gènes code de toutes nouvelles métallothionéines.

- **L'expression de ces gènes est-elle induite par des métaux (Chapitre 3) ?**

En nous basant sur les propriétés des métallothionéines, il est probable que ces gènes soient inductibles par la présence de métaux. Les organismes portant les CRP n'étant pas cultivables, nous avons dû mettre au point : 1) des microcosmes de sol amendés ou non de manière exogène par du Cd, du Cu ou du Zn et 2) une méthode de quantification par PCR quantitative des gènes et des transcrits présents au sein de ces microcosmes. Ces mises au point permettront par la suite de quantifier l'expression de ces gènes en présence de métal exogène.

- **De quel(s) groupe(s) d'organisme(s) sont issus les gènes codant les CRP (Chapitre 4) ?**

Ces gènes n'étant connus que sous la forme de séquences d'ADNc isolées, nous avons tenté de rechercher les gènes proximaux qui leurs sont



associés par une approche de capture ciblée sur de longs fragments d'ADN métagénomiques extraits de sol afin d'utiliser ces gènes proximaux pour faire des analyses phylogénétiques et identifier le ou les groupes taxonomiques des microorganismes qui portent ces gènes *crp*. La mise au point de cette technique n'a pas abouti malgré les nombreux essais.





CHAPITRE 2

Les CRP sont-elles des métallothionéines ?

Introduction

Un nombre important de métallothionéines a été découvert ces dernières années remettant en question certaines caractéristiques de ces protéines comme par exemple la " petite " taille de leurs séquences ou l'absence d'acides aminés aromatiques. Ces nouvelles métallothionéines ont été découvertes à partir d'organismes modèles. Les CRP isolées par Lehembre et ses collaborateurs (**Lehembre et al. 2013**) sont de bonnes candidates pour être les premières métallothionéines découvertes par une approche « -omique » à partir d'échantillons environnementaux.

Pour savoir si les CRP sont des métallothionéines, nous avons réalisé les mêmes analyses utilisées pour caractériser les métallothionéines putatives à savoir : 1) des analyses *in silico* (description de l'organisation en acides aminés, comparaison de séquences par calcul de similarités), 2) des études *in vitro* (quantité de métaux liés par protéine) et 3) *in vivo* (complémentation fonctionnelle et test en gouttes). La majeure partie des résultats de ces analyses a été publiée dans *Journal of Inorganic Biochemistry* (JIB) (**Ziller et al. 2017**) en collaboration avec deux équipes de recherche espagnoles à l'origine de nombreuses caractérisations de MT et de leur classification fonctionnelle (**Palacios et al. 2011b**).



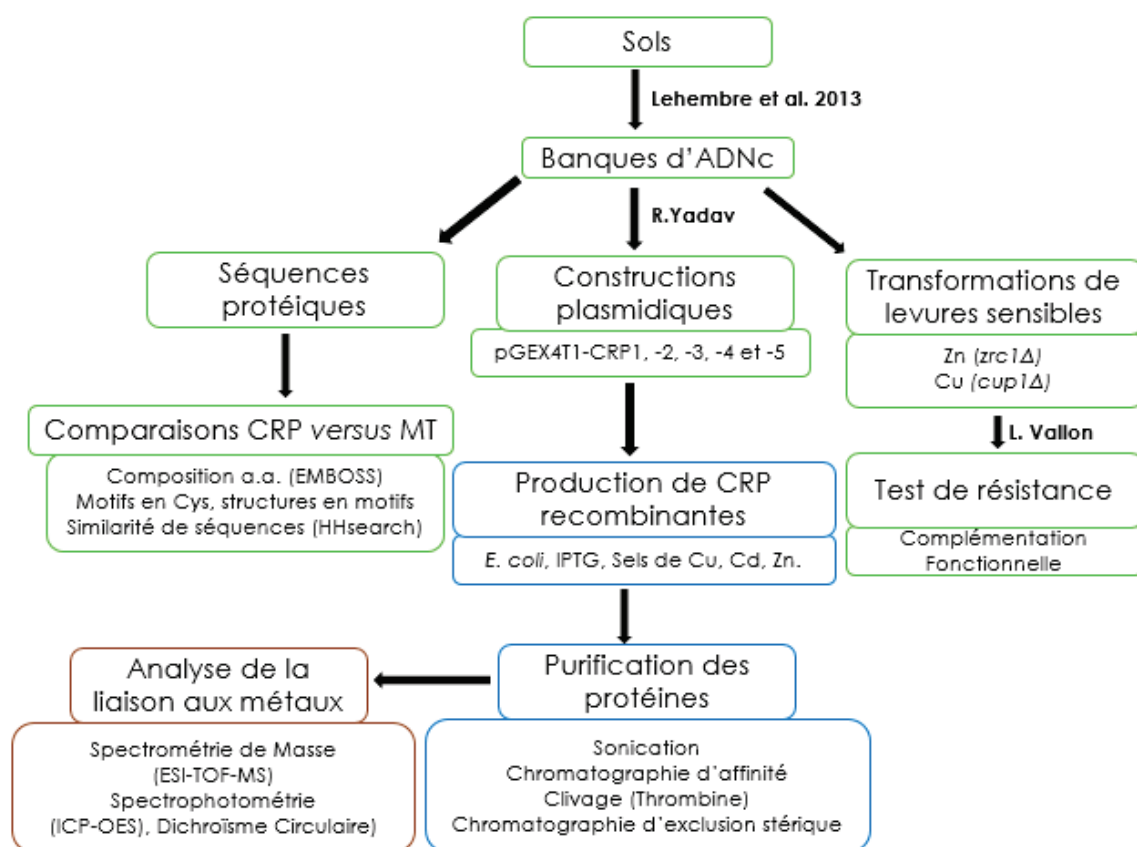


Figure II.1 : Schéma de la stratégie expérimentale. Les CRP et les métallothionéines sont comparées sous 3 points de vue : 1) Les séquences primaires en acides aminés, 2) la capacité à chélater *in vitro* des métaux 3) et la capacité à conférer *in vivo* une tolérance aux métaux lorsqu'elles sont exprimées dans des levures sensibles.

Encadrées en vert, les étapes réalisées au Laboratoire d'Ecologie Microbienne de Lyon, en Bleu, celles réalisées à l'Université de Barcelone et en brun, celles réalisées à l'Université Autonome de Barcelone.



La stratégie expérimentale est représentée **figure II.1**. J'ai réalisé les comparaisons de séquences et les transformations de levure au Laboratoire d'Ecologie Microbienne de Lyon (LEM). Les complémentations fonctionnelles et les constructions plasmidiques surexprimant les protéines recombinantes ont été réalisées au LEM respectivement par Laurent Vallon, le technicien de l'équipe, et le Dr Rajiv Yadav, un post-doctorant de notre équipe financé par le CEFIPRA (Centre Franco-Indien pour la Promotion de la Recherche Avancée), projet 4709-1 « *Gene resources from polluted soils* ». En collaboration avec 2 équipes espagnoles, j'ai effectué les productions et les purifications des protéines recombinantes à l'Université de Barcelone dans l'équipe du professeur Silvia Atrian et les analyses biochimiques à l'Université Autonome de Barcelone dans l'équipe du professeur Mercè Capdevila sous les conseils et l'aide du Dr. Oscar Palacios, lors d'un séjour à Barcelone financé par la région Rhône-Alpes (Bourse CMIRA) et l'école doctorale (E₂M₂).

Dans une première partie, l'article publié dans JIB est présenté. Il fait état de la caractérisation de 5 séquences protéiques de CRP qui se révèlent être toutes des métallothionéines appartenant ainsi à une nouvelle famille de métallothionéines. Les 5 séquences caractérisées ont des propriétés communes avec les MT connues. Elles possèdent des séquences d'acides aminés comprenant 20-24% de résidus cystéines, n'ont pas d'acides aminés aromatiques autres que 2 à 4 His, ont des motifs en cystéines variés (CCC, CxC, CC, CxxC), conservés parmi les CRP, dont la répartition se fait en deux domaines séparés par une séquence relativement longue comme chez les MT de plantes. Mais elles ne sont homologues à aucune MT décrite jusqu'à maintenant. Toutes les CRP chélatent le Cd²⁺, le Zn²⁺ et le Cu⁺ et confèrent, lorsqu'elles sont exprimées dans des levures sensibles, un phénotype de tolérance au Cd et 4 d'entre elles au Zn.

Dans une deuxième partie, des données supplémentaires concernant la purification des protéines sont présentées.





Metagenomics analysis reveals a new metallothionein family: Sequence and metal-binding features of new environmental cysteine-rich proteins

Antoine Ziller^{a,1}, Rajiv Kumar Yadav^{a,1,2}, Mercè Capdevila^b, Mondem Sudhakara Reddy^c, Laurent Vallon^a, Roland Marmesse^a, Silvia Atrian^{d,3}, Òscar Palacios^{b,3}, Laurence Fraissinet-Tachet^{a,*,3}

^a Microbial Ecology, CNRS UMR 5557, INRA UMR 1418, Université Lyon 1, Université de Lyon, F-69622 Villeurbanne, France

^b Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Cerdanyola del Vallès, Barcelona, Spain

^c Thapar University, Department of Biotechnology, Bhadsan Road, Patiala 147 004, India

^d Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, E-08028 Barcelona, Spain

ARTICLE INFO

Article history:

Received 26 July 2016

Received in revised form 28 September 2016

Accepted 11 November 2016

Available online 12 November 2016

Keywords:

Metallothionein

Cysteine-rich protein

Metal-binding protein

Environmental genomics

Metatranscriptomics

ABSTRACT

Metallothioneins are cysteine-rich proteins, which function as (i) metal carriers in basal cell metabolism and (ii) protective metal chelators in conditions of metal excess. Metallothioneins have been characterized from different eukaryotic model and cultivable species. Presently, they are categorized in 15 families but evolutionary relationships between these metallothionein families remain unresolved. Several cysteine-rich protein encoding genes that conferred Cd-tolerance in Cd-sensitive yeast mutants have previously been isolated from soil eukaryotic metatranscriptomes. They were called CRPs for “cysteine-rich proteins”. These proteins, of unknown taxonomic origins, share conserved cysteine motifs and could be considered as metallothioneins. In the present work, we analyzed these CRPs with respect to their amino acid sequence features and their metal-binding abilities towards Cd, Zn and Cu metal ions. Sequence analysis revealed that they share common features with different known metallothionein families, but also exhibit unique specific features. Noticeably, CRPs display two separate cysteine-rich domains which, when expressed separately in yeast, confer Cd-tolerance. The N-terminal domain contains some conserved atypical Cys motifs, such as one CCC and two CXCC ones. Five CRPs were expressed and purified as recombinant proteins and their metal-binding characteristics were studied. All these CRPs chelated Cd(II), Zn(II) and Cu(I), although displaying a better capacity for Zn(II) coordination. All CRPs are able to confer Cd-tolerance, and four of them confer Zn-tolerance in the Zn-sensitive *zrc1Δ* yeast mutant. We designated these CRPs as environmental metallothioneins belonging to a new formerly undescribed metallothionein family.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Metallothioneins (MTs) encompass an array of ubiquitous, small (24–257 amino acids long), cysteine-rich proteins (15–35% of cysteines) with no or a low content of aromatic amino acids. MTs coordinate metal ions through the formation of metal-thiolate complexes via sulfur-metallic ion interactions [1,2]. Although several roles have been proposed for MTs, it is generally accepted that, (i) they are basal metal carriers in metal homeostasis processes, and, (ii) constitute an important line of defense against metallic stress resulting from high

concentrations of either essential (e.g. Zn, Cu, ...) or non-essential (e.g. Cd, ...) metal ions.

Known MTs come essentially from vertebrates and plants (respectively, 82 and 9% of the total number of MT sequences in Swiss-Prot). Based on sequence similarity analysis, MTs were classified into 15 different families [3]. However, besides the shared structural characteristics listed above, MTs share little overall sequence conservation between families suggesting that different MT families may have emerged independently in several occasions during the course of evolution and that their shared structural features are the result of multiple convergent evolutionary processes [4]. Another MT classification scheme, based on more functional criteria, such as the metal-binding preferences of the MT polypeptides, has also been proposed [5]. This classification distinguishes Zn-thioneins, displaying a preference for divalent metal ions, from Cu-thioneins showing preferences for monovalent ones. However, beyond this simplistic binary model, a more comprehensive analysis of metal-binding features of a large number of MTs led to the proposal of a continuous gradation between both extreme behaviors [6].

* Corresponding author at: Ecologie Microbienne, Université Lyon 1, 10, rue Dubois, Bât Lwoff, F-69622, Villeurbanne, France.

E-mail address: fraissin@univ-lyon1.fr (L. Fraissinet-Tachet).

URL: <http://www.ecologiemiocriennelyon.fr/> (L. Fraissinet-Tachet).

¹ These two first authors contributed equally to the work.

² Present address: Department of Botany, University of Allahabad, Allahabad 211 002, India.

³ These three last authors contributed equally to the work.

Nomenclature

CRP	cysteine-rich protein
ESI-MS	electrospray ionization mass spectrometry
FPLC	fast protein liquid chromatography
GST	glutathione-S-transferase
ICP-AES	inductively coupled plasma atomic emission spectroscopy
IPTG	isopropyl β -D-thiogalactopyranoside
MT	metallothionein
PCR	polymerase chain reaction
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Despite the wealth of knowledge regarding MTs, all of them come from known, trackable species not representative of the known phylogenetic diversity of the living world. Indeed, most eukaryotic MTs come from a few phyla not representative of the entire eukaryotic domain [7,8]. MTs have been frequently described and studied in vertebrates, plants and fungi, but rarely in the Alveolata (e.g. MTs from the ciliate *Tetrahymena*), the Stramenopiles (MTs from the brown alga *Fucus vesiculosus*, the oomycete *Phytophthora cinnamomi* and the diatom *Thalassiosira pseudonana*) and the Excavata (one MT from the pathogen *Trichomonas vaginalis*) [9] and never, to the best of our knowledge, in the phyla Rhizaria and Amoebozoa. In this respect, the screening of “environmental nucleic acids” (DNA or RNA), directly extracted from environmental matrices (soil, water, sediments...), represents an attractive strategy to isolate novel genes from a large diversity of organisms, cultivable or not [10]. Because of its tremendous heterogeneity, the soil environment hosts a very large array of taxa, many of them undescribed and representative of most of the phylogenetic diversity of the microscopic Eukarya [11,12].

Hence, based on a functional metatranscriptomic approach, using RNA directly extracted from metal-contaminated or uncontaminated soil samples, new families of eukaryotic genes involved in resistance to Cd or Zn have been identified [13]. Several of these genes, able to restore Cd resistance to Cd-sensitive yeast mutants, encoded cysteine-rich proteins (CRPs). Of these soil CRPs, a group of 11 genes encoded proteins of similar lengths (110–133 amino acids), sharing conserved amino acid residues including several cysteine motifs but had no homologs deposited in public databases, which prevented taxonomic affiliation and protein family identification. However, all of their features suggested that these polypeptides could represent new metallothioneins potentially originating from yet unexplored eukaryotic taxa.

The objectives of this study were to determine if these cysteine-rich proteins (CRPs) can be classified as metallothioneins, by comparing their features to those of well-known MTs. Our study includes (i) a detailed analysis of CRP sequence features, (ii) the heterologous expression of full-length and truncated sequences in Cd-sensitive yeast mutants and, (iii) the characterization of the metal-binding ability of the purified recombinant CRPs in presence of Cu, Cd and Zn metal ions.

2. Materials and methods

2.1. Amino acid sequence features and similarity comparison

CRP cDNAs were previously isolated from eukaryotic metatranscriptomes and their sequences included a 3' poly-A tail suggesting that they derived from an eukaryotic mRNA [13]. CRP amino acid sequences [International Nucleotide Sequence Database (DDBJ, ENA, GenBank) accession Nos. CCG34099 (CRP1), CCG34101 (CRP2), CCG34096 (CRP3), CCG34094 (CRP4) and CCG34097 (CRP5)] were aligned using ClustalO available at the EBI website (<http://www.ebi.ac.uk/Tools/>

[msa/clustalo/](http://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/)). To build a MT dataset, we retrieved annotated and reviewed MT sequences from the Swiss-Prot database (<http://www.uniprot.org/>) using the PSI-BLAST tool. Amino acid composition for each protein was calculated using the EMBOSS pepstat tool (http://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/) and Cys motifs were counted manually.

To ascertain homology between CRPs and MT sequences from families 3, 7, 11 and 15 from crustaceans, ciliates, Ascomycota fungi (MT family 11) and plants, respectively, multiple sequence alignments were built for each family using MUSCLE [14] and the Jalview software [15] and too divergent sequences were discarded manually. One HMM profile was then built for each of the MT families 3, 7, 11 and 15 and compared to CRP sequences using HMMER 3 package (hmmerr.org/) (details are shown in Supplementary Table S1).

2.2. Cloning of truncated CRP sequences

To express truncated CRPs in yeast, expression vectors were constructed using PCR (polymerase chain reaction)-based strategies with full-length CRP5 cDNA sequence (accession No. HE716995) as a template by using high-fidelity *Platinum* DNA polymerase (Life Technologies, Germany). Gene fragments encoding the N- (CRP5N) and C- (CRP5C) terminal domains of the protein were amplified using the primer pairs CRP5NF/CRP5NR and CRP5CF/CRP5CR, respectively (Table S2). Each of these primers contained *NotI* restriction sites and in-frame start or stop codons in forward (F) or reverse (R) primers, respectively. After PCR amplification from the CRP5 cDNA, the PCR products were purified by agarose gel electrophoresis, digested by *NotI* and ligated in the *NotI*-digested and dephosphorylated *Sfi*1-pFL61 yeast-*E. coli* shuttle vector [13]. Integrity and identity of the cloned gene fragments were confirmed by Sanger sequencing. The two resulting partial sequences were deposited in the GenBank/ENA/DDJB database under accession No. LT629274 and LT629275 for CRP5N and CRP5C respectively.

2.3. Yeast metal tolerance assays

Zn and Cd tolerance assays were performed using the Zn-sensitive *zrc1Δ* and Cd-sensitive *ycf1Δ* *Saccharomyces cerevisiae* mutant strains, respectively. These two strains derive from the wild-type strain BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) that was used as a positive control for metal tolerance. All three strains were purchased from Euroscarf. The *YCF1* gene encodes an ABC transporter that confers Cd tolerance through the transport of Cd conjugates into vacuoles [16, 17]; the *ZRC1* gene encodes a transporter that sequesters Zn into vacuoles [18,19]. All these strains were transformed with either the *Sfi*1-pFL61 empty vector or the same plasmid into which either the complete CRP1, -2, -3, -4 or -5 genes [13] or the truncated CRP5N or CRP5C sequences [this study Section 2.2] had been cloned. Transformations were performed using the standard lithium acetate method [20]. For metal tolerance assays, yeast cells were grown overnight in liquid YNB (0.67% yeast nitrogen base, 2% glucose supplemented with a complete amino acid solution without uracil, pH 5.9). Cell suspensions were diluted to OD_{600nm} = 1.0. Five microliters of culture 10-fold dilution series were then spotted on YNB plates containing or not toxic concentrations of ZnSO₄ (15 mM) or CdCl₂ (40 μ M). Growth was monitored every day for a maximum of 7 days at 30 °C.

2.4. Recombinant plasmid construction for GST-CRP fusion protein synthesis

The five CRP cDNA sequences [accession numbers for CRP1 (HE716997), CRP2 (HE716999), CRP3 (HE716994), CRP4 (HE716992) and CRP5 (HE716995)] previously isolated from soil metatranscriptomes [13] were sub-cloned into the expression vector pGEX-4T-1 (GE Healthcare, United Kingdom) in frame with the glutathione-S-transferase (GST) coding region to synthesize the corresponding fusion proteins.

For this purpose, the five cDNAs were first amplified from the cDNA-Sfi1pFL61 vectors, using specific upstream and downstream PCR primers (Table S3), which introduce two restriction sites, one before the start codon (*Bam*HI) and one after the stop codon (*Sal*I). PCR was performed using the high fidelity *Platinum* DNA polymerase (Life Technologies, Germany) with the following conditions: initial denaturation at 95 °C for 2 min; 30 cycles consisting of 95 °C for 15 s, 58 °C for 30 s, 72 °C for 30 s and a final extension at 72 °C for 5 min. After a *Bam*HI-*Sal*I digestion and a gel extraction, PCR products corresponding to the full-length CRP coding regions were ligated into the corresponding sites of the pGEX-4T-1 expression vector. Correctness of the fusion between the CRP and GST reading frames was controlled by Sanger DNA sequencing, and the constructs were transformed into BL21 *Escherichia coli* cells for protein synthesis. Length of the five different GST-CRP fusion proteins was estimated following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

2.5. Recombinant protein production and purification.

Each CRP protein was synthesized in Zn-, Cd- or Cu-supplemented cultures. For each culture, one BL21 *E. coli* transformed clone was grown at 37 °C in 5 L of Luria Bertani (LB) ampicillin-supplemented medium inoculated with an overnight 500 mL pre-culture. GST-CRP gene expression was induced after 1 h 30 min of culture by adding 100 µM of isopropyl β-D-thiogalactopyranoside (final concentration) 30 min before metal supplementation (300 µM ZnCl₂, 300 µM CdCl₂ or 500 µM CuSO₄, final concentrations). After 3 h of incubation at 30 °C, the cultures were centrifuged for 5 min at 7700g and the pellets were resuspended in a mixture of ice cold PBS (1.4 M NaCl; 27 mM KCl; 100 mM Na₂HPO₄; 18 mM KH₂PO₄; pH 7.3) and 8.9 mM β-mercaptoethanol just before cell sonication (Branson Sonifier 250; 0.6 Hz) at 4 °C with pulses of 0.6 s for 8 min. The GST-CRP proteins were purified as described in Cols et al. [21]. Briefly, they were recovered from the total

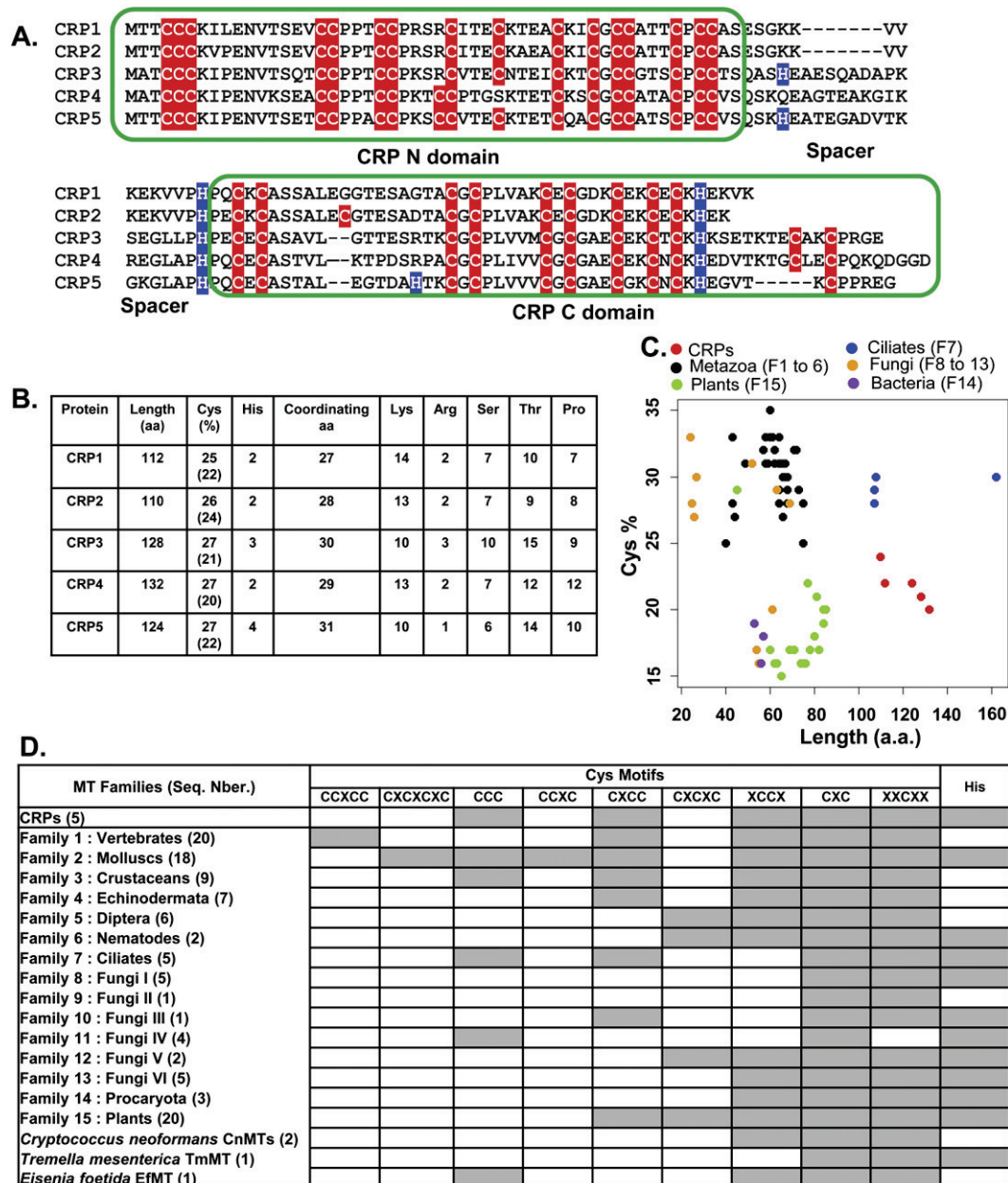


Fig. 1. Main sequence features of the five CRP1, 2, 3, 4 and 5 proteins and of representative of the different known MT families. A. Sequences alignment of the 5 studied CRPs. Cys residues are highlighted in red and His in blue. Both N- and C-terminal cysteine-rich domains (framed by green rectangles) are separated by a cysteine-free spacer. B. Length, cysteine content and MT typical amino acid composition of CRPs. C. Scatter plot giving the percentage of Cys versus protein length for different MTs and CRPs. D. Table giving the presence (grey boxes) or absence (white boxes) of His and Cys motifs (where X is any amino acid and C a cysteine residue) in the different MT protein families and in the CRPs.

bacterial cell extracts by batch affinity chromatography, using a Glutathione Sepharose matrix (GE Healthcare). After three washes with ice cold PBS, GST-CRP fusion proteins were cleaved by addition of thrombin and digestion overnight at 17 °C. The mixtures were then centrifuged to remove the beads of Glutathione Sepharose, and the supernatant was concentrated using Centrprep 3 kDa cut-off Microcons (Amicon). Protein samples were then fractionated by size-exclusion fast protein liquid chromatography (FPLC) using a Superdex-75 column (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 7.0. Fractions containing proteins were identified by measuring absorbance at 254 and 280 nm and stored at –80 °C. During all the steps of the purification procedure, argon was bubbled in samples and buffers in order to avoid protein oxidation.

2.6. Characterization of the metal-CRP complexes by optical spectroscopy

The S, Zn, Cd and Cu contents of all the CRP preparations were analyzed by Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) in a Polyscan 61E (Thermo Jarrel Ash) spectrometer, measuring S at 182.04 nm, Zn at 213.85 nm, Cd at 228.80 nm and Cu at 324.80 nm. Samples were treated as described in Bongers et al. [22]. Protein concentrations were calculated from the ICP-AES sulfur measurements, assuming that all the S atoms were contributed by the Cys residues of the CRP peptides.

2.7. Characterization of the metal-CRP complexes by mass spectrometry

The molecular weights of the recovered complexes were determined by electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) on a Micro TOF-Q instrument (Bruker Biospin, France) interfaced with a Series 1200 HPLC Agilent pump, equipped with an autosampler, all of which were controlled by the Compass Software. Calibration was performed with the ESI-L Low Concentration Tuning Mix (Agilent Technologies, Germany). Samples containing Zn-CRP or Cd-CRP complexes were analyzed under the following conditions: 20 µL of protein solution injected through a polyether heteroketone tubing (1.5 m × 0.18 mm internal diameter) at 40 µL min^{–1}; capillary counter-electrode voltage, 5 kV; desolvation temperature, 90–110 °C; dry gas, 6 L min^{–1}; spectra collection range, 800–2500 m/z. The carrier buffer was a 5:95 mixture of acetonitrile:ammonium acetate (15 mM, pH 7.0). Alternatively, the corresponding Cu-complexes were analyzed

as follows: 20 µL of protein solution injected at 40 µL min^{–1}; capillary counter-electrode voltage, 3.5 kV; lens counter-electrode voltage, 4 kV; dry temperature, 80 °C; dry gas, 6 L min^{–1}. In this latter case, the carrier was a 10:90 mixture of acetonitrile:ammonium acetate, 15 mM, pH 7.0. For the analysis of the apo-peptides and Cu-CRP complexes at acidic pH, 20 µL of the corresponding sample were injected under the same conditions described previously, but using a 5:95 mixture of acetonitrile:formic acid (pH 2.4), as liquid carrier, which causes the complete demetalation of the peptides loaded with Zn, but keeps the Cu ions bound to the peptides. Under all the conditions assayed, the error associated with the mass measurements was always lower than 0.1%. Masses for the holo-species were calculated as previously described [23].

3. Results and discussion

3.1. CRP sequence and metal resistance features

The protein sequences of the five CRPs analyzed in this study are between 110 and 132 amino acids long, contain 20–24% of cysteines and 2–4 His but no other aromatic residues (Fig. 1A and B). The ratio Lys/Arg is clearly in favor to Lys and the ratio Ser/Thr is in favor to Thr. These CRPs were initially selected for their ability to restore Cd tolerance when expressed in *ycf1Δ* yeast mutant [13]. We show here that CRP2 is the most capable form to restore Zn tolerance when expressed in *zrc1Δ*, zinc-sensitive yeast mutant, followed in decreasing capacity by CRP4, CRP5, CRP1 and CRP3 (Fig. 2A). None of the CRPs, however, complemented for Cu tolerance when expressed in a *cup1Δ* copper-sensitive yeast mutant (data not shown). It is worth noting that, compared to Zn, the ability to restore Cd tolerance followed a different order for the 5 CRPs. While CRP2 conferred also the highest level of Cd resistance, the order for the four others was CRP3 > CRP4 > CRP5 [13]. Despite the fact that CRP sequences could not be globally aligned to the sequences of other known MTs, all of these features make them potential members of the MT superfamily according to Binz and Kagi [3].

When we compared CRPs between them, the five protein sequences present little dissimilarity (Fig. 1A). Differences essentially concern non-cysteine residues. Length differences result mainly from differences in the central region and in the cysteine-poor terminal tail. Nevertheless, if we count the number of coordinating residues (Cys and His)

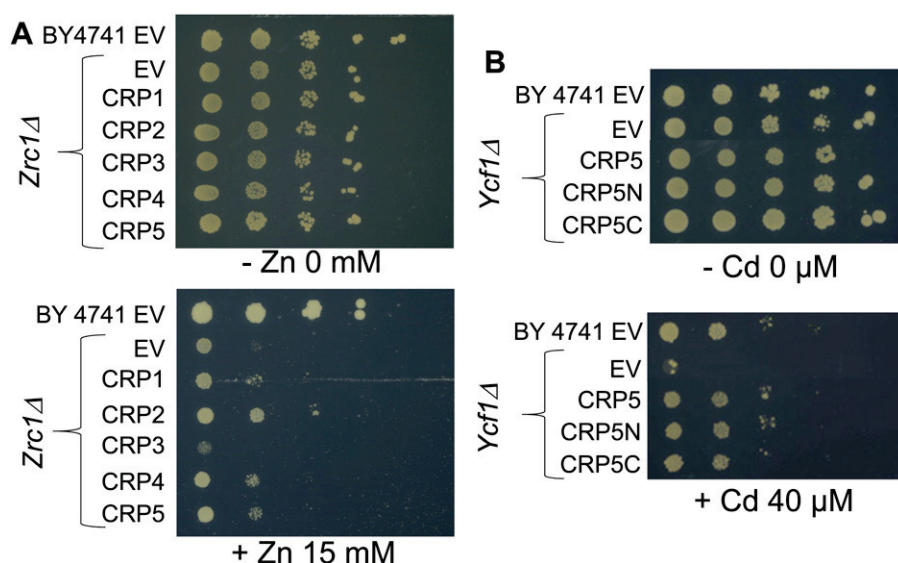


Fig. 2. Metal tolerance assay of yeast cells expressing full length or truncated CRPs (drop test assays). A. Zn-tolerance assay using the Zn-sensitive *S. cerevisiae* *Zrc1Δ* strain expressing or not (EV = empty vector) the full-length CRP1, -2, -3, -4 or -5 cDNAs. B. Cd-tolerance assay using the Cd-sensitive *S. cerevisiae* *Ycf1Δ* strain expressing or not the full-length or truncated (N or C-terminal domains) CRP5 sequences. In both tests, the wild-type strain BY4741 was used as a positive metal-tolerant control.

potentially participating in metal coordination, CRPs differ from each other in this order, CRP5 (31 residues) > CRP3 (30) > CRP4 (29) > CRP2 (28) > CRP1 (27) (Fig. 1B). Such differences could potentially account for the distinct metal-binding abilities that are presented in the next section.

Closer analysis of CRP sequences revealed an uneven distribution of cysteines, with the presence of two distinct N- and C-terminal cysteine-rich domains of *ca* 45–52 amino acids in length, separated by a 18 (CRP1 and 2) or 25 (CRP3, 4 and 5) amino acid-long spacer devoid of cysteines (Fig. 1A). Such a bipartite organization is typical of plant MTs, although in this case the linker is longer in proportion of the shorter Cys-rich moieties, which contain fewer cysteines [24–26]. Also similarly to plant

MTs, the composition of the CRP spacer region is rich in hydrophobic amino acids (Val, Leu, Ile, Ala), and among the charged residues, Lys clearly predominates. Several hypotheses have been proposed for the potential functions of this linker in plants. It might be involved in plant apo-protein folding when binding metal ions or could play a role in the coordination between the two protein domains which could be important for metal binding affinities and specially when MTs are used as metal chaperones in cell basal physiological processes [27–29].

Concerning their length, CRPs are in the upper limit of the known MT size range that does not follow a normal distribution (Fig. 1C). Indeed, most characterized MTs have a length comprised between 24 and 85 amino acids and MTs longer than 110 amino acids have only been

Table 1

Experimental molecular masses (from ESI-MS analyses) and calculated molecular masses for the distinct metal-CRP species identified in the corresponding metal-CRP syntheses. Major species are highlighted in bold. M = Zn + Cu. ns, not studied as no recombinant proteins could be purified from the corresponding cultures.

Metal-supplemented production	Identified species	Experimental mass (Da)	Theoretical mass (Da)	Metal-supplemented production	Identified species	Experimental mass (Da)	Theoretical mass (Da)
Zn-CRP1	Zn ₇	12,340	12,343.7	Cu-CRP1	pH 7.0:		
	Zn ₈	12,406	12,407.1		M ₁₀	12,527	12,525.5
	Zn₉	12,469	12,470.5		M₁₂	12,652	12,650.6
	Cd ₁₀	13,002	13,004.1		M₁₃	12,718	12,713.2
	Cd₁₁S	13,150	13,146.5		M ₁₄	12,775	12,775.7
	Cd ₁₂ S ₂	13,292	13,288.9		pH 2.4:		
	Cd ₁₃ S ₂	13,405	13,399.3		Cu₈	12,399	12,400.4
Cd-CRP1	Cd ₁₅	13,551	13,556.2		Cu ₉	12,458	12,463.0
					Cu ₁₀	12,524	12,525.5
					Cu ₁₁	12,585	12,588.1
					Cu ₁₂	12,651	12,650.6
					pH 7.0:		
					M₁₃	12,525	12,530.6
					M₁₄	12,590	12,593.1
Zn-CRP2	Zn ₆	12,093	12,097.7	Cu-CRP2	M ₁₅	12,648	12,655.7
	Zn ₇	12,157	12,161.1		M ₁₆	12,709	12,718.2
	Zn ₈	12,220	12,224.5		pH 2.4:		
	Zn₉	12,288	12,287.9		Cu ₆	12,089	12,092.7
	Zn₁₀	12,353	12,351.3		Cu ₇	12,149	12,155.3
	Zn ₁₁	12,420	12,414.7		Cu ₈	12,217	12,217.8
	ns	ns	ns		Cu₉	12,279	12,280.4
Cd-CRP2					Cu ₁₀	12,335	12,342.9
					Cu ₁₁	12,410	12,405.5
					pH 7.0:		
					M ₈	14,986	13,993.9
					M₉	14,052	14,056.5
					M₁₀	14,115	14,119.0
					M ₁₁	14,185	14,181.6
Zn-CRP3	Zn ₅	13,802	13,810.5	Cu-CRP3	pH 2.4:		
	Zn ₆	13,865	13,873.8		Cu ₄	13,738	13,743.7
	Zn ₇	13,935	13,937.2		Cu₈	13,989	13,993.9
	Zn ₈	14,001	14,000.6		Cu₉	14,055	14,056.5
	Zn₉	14,064	14,064.0				
	Zn ₁₀	14,128	14,127.4		pH 7.0:		
	Zn ₁₁	14,193	14,190.8		M ₁₁	14,615	14,607.1
Cd-CRP3	Cd ₁₁ S	14,742	14,740.1	Cu-CRP4	M ₁₄	14,675	14,669.6
	Cd ₁₂	14,817	14,818.4		M₁₆	14,793	14,794.7
	Cd₁₂S	14,853	14,850.5		pH 2.4:		
	Cd ₁₃ S ₂	14,999	14,993.0		Cu ₉	14,348	14,356.9
	Zn ₆	14,169	14,174.2		Cu ₁₁	14,486	14,482.0
	Zn ₇	14,235	14,237.6		Cu ₁₂	14,540	14,544.5
	Zn₈	14,302	14,301.0		Cu₁₃	14,608	14,607.1
Zn-CRP4	Zn₉	14,365	14,364.4		Cu ₁₄	14,660	14,669.6
	Zn ₁₀	14,423	14,427.8		Cu ₁₅	14,725	14,732.2
	Cd₉	14,787	14,787.6		pH 7.0:		
	Cd ₉ S ₂	14,850	14,851.7		M ₈	13,315	13,321.2
	Cd ₁₂	15,120	15,118.8		M₉	13,379	13,383.8
	Cd ₁₂ S	15,155	15,150.9		M ₁₀	13,439	13,446.3
					pH 2.4:		
Cd-CRP4					Cu₈	13,323	13,321.2
					Cu ₉	13,385	13,383.8
					Cu ₁₀	13,451	13,446.3
Zn-CRP5	Zn ₆	13,200	13,201.1	Cu-CRP5	pH 7.0:		
	Zn ₇	13,264	13,264.5		M ₈	13,315	13,321.2
	Zn₈	13,327	13,327.9		M₉	13,379	13,383.8
	Zn₉	13,392	13,391.3		M ₁₀	13,439	13,446.3
	Zn ₁₀	13,452	13,454.7		pH 2.4:		
	Cd ₁₂	14,140	14,145.7		Cu₈	13,323	13,321.2
	Cd₁₂S	14,177	14,175.8		Cu ₉	13,385	13,383.8
Cd-CRP5	Cd ₁₄ S ₂	14,425	14,430.7		Cu ₁₀	13,451	13,446.3
	Cd ₁₅ S	14,502	14,509.0				
	Cd ₁₆	14,591	14,587.4				

reported from protists, unicellular ciliates (*Tetrahymena* sp. [30] and *Paramecium* sp. [31,32]), as well as from Basidiomycota fungi of the Tremellales order (*Cryptococcus neoformans* [33] and *Tremella mesenterica* [34]). Ciliates MTs (from MT family 7) have a modular architecture with a combination of repetitive structures that could explain their length [35,36]. A potential origin for the unusual length of CRPs could be that they originated from the fusion of two ancestral and shorter MTs that constitute the N and C-terminal domains in today's CRPs. These shorter MTs would have been around 50 amino acids-long, which correspond to sizes commonly observed for MTs from many taxonomic groups (Fig. 1C). This hypothesis predicts that the CRP's N and C-terminal domains may function independently from the other in conferring metal resistance. This was tested by expressing separately the N and C-terminal domains from CRP5 in yeast Cd-sensitive mutants. As expected both CRP5N and CRP5C domains restored Cd-tolerance in *ycf1Δ* as did the corresponding full-length CRP5 protein (Fig. 2B). This again reinforces the similarity between CRPs and plant MTs, since a similar behavior was shown for the two domains of the *Quercus suber* MT2 protein [37].

At a more detailed scale, cysteine motifs and distribution have also been used in MT comparison [35,38]. At that level, regarding the cysteine motifs in the sequences from different MT families (Fig. 1D), CRPs contain five different cysteine motifs, conserved in all CRPs, including an unusual Cys triplet near the N-terminus of the protein. Cys triplets are very rare among MTs as they have only been reported in MTs with different metal preferences from five unrelated organisms, one crustacean Cu-specific MT [39], one from the mollusk *Crassostrea virginica* (AAZ94898), one from the annelid *Eisenia foetida* (P81695), several very similar Cd-specific ciliate MTs [35,36,40] and four *Yarrowia* (a yeast species) MT isoforms [41] which have been adopted as a model for the study of bioremediation of metal contaminated soils [42]. CRPs N-terminal domains also contain 4–5 Cys doublets (two CXCC and two or three XCCX motifs) and few isolated Cys residues. As for the C-terminal domain, it contains more conventional Cys motifs compared to other known MTs. This is the case for four CXC motifs, present in all MTs, and from one to three isolated cysteine(s) (Fig. 1A and D).

Between two and four His residues are also present in CRPs. Two of them are fully conserved, one in the central spacer, close to the C-terminal domain, and one after the last conserved Cys residue (Fig. 1A). His residues have been suggested to participate to metal coordination in MTs [43]. The presence of His residues is worth noting, because it is a distinctive, although not exclusive, trait of plant MTs where they are also exclusively located either in the spacer or in the C-terminal protein region [26]. Fungal MTs are also characterized by the frequent presence of His [42]. Finally, it is worth noting the frequent presence of prolines exactly after the cysteine residues, which is a constant in several long fungal MTs [33,34].

In conclusion, CRP protein sequences exhibit peculiarities among all MT sequences reported so far that are reminiscent, to some extent, of MTs from unrelated taxa: plants (MT family 15), crustaceans (MT family 3), fungi (MT family 11) and ciliates (MT family 7). To compare these 4 families with CRP sequences, we computed HMM (hidden Markov models) profile for each of these families, thus allowing an accurate sequence comparison. For this purpose, all MT sequences from families 3, 7, 11 and 15 were recovered from databases using PSI-BLAST searches and HMM profiles were built for each of these MT families. According to HMM profile comparison, no hits were found between these MT families and CRP sequences demonstrating that CRPs do represent a completely new MT family. However, metal binding properties needed to be experimentally established to definitely qualify CRPs as MTs [1–3].

3.2. Production and identity of recombinant CRP polypeptides

The cDNAs of CRP1, -2, -3, -4 and -5 were fused to a GST coding gene and cloned separately. The sequencing of the corresponding expression plasmids pGEX-CRPs confirmed the accuracy and integrity of the fusion

constructs. Each GST-CRP protein synthesis was induced by IPTG (isopropyl β-D-thiogalactopyranoside) in Zn-, Cd-, or Cu-supplemented *E. coli* cultures, and the resulting recombinant Zn-, Cd- and Cu-protein complexes were purified and analyzed by electrospray ionization mass spectrometry (ESI-MS) and inductively coupled plasma atomic emission spectroscopy (ICP-AES).

SDS-PAGE of the 15 transformed *E. coli* protein extracts (5 CRPs × 3 metal ions) revealed the presence of bands corresponding to the expected GST-CRP molecular weights for each fusion protein (data not shown). After GST-CRP cleavage with thrombin, to confirm the identity of the apo-polypeptides, aliquots of the Zn- and Cd-CRP preparations were acidified to pH 2.4 (an acidic pH that should lead to demetalation of the divalent metal ion-CRP complexes) and analyzed by ESI-MS. The molecular weights of the recombinant apo-CRP1, -2, -3, -4 and -5 were found to be 11,898; 11,716; 13,490; 13,789 and 12,817 Da, respectively. These are in accordance with the expected molecular weights corresponding to the different CRP sequences with additional N-terminal Gly and Ser residues resulting from their synthesis as thrombin-cleaved GST fusions (11,900.0; 11,717.4; 13,493.5; 13,793.9 and 12,820.8 Da, respectively), which supports the identity and integrity of the recombinant peptides. Nevertheless, in almost all of the acidified Zn-, but not those of the Cd-CRP samples, demetalation was not complete at pH 2.4. Hence, minor Zn₂- and Zn₄- complexes accompanied the apo-peptides for CRP1, -2, -3 and -5, while Zn₂-, Zn₄- and Zn₆-CRP4 complexes were clearly detectable together with the major apo-CRP4 form (Table 1). Similar results were obtained for the *Tetrahymena thermophila* MTT1 and MTT3 isoforms, for which Zn₄-MTT1, Cd₁₂- and Cd₁₁-MTT1 and Cd₈-MTT3 complexes were also detected after acidification at pH 2.4 [40]. Another noticeable feature was that in all of the acidified CRP samples, minor dimer species (mainly dimeric apo-species) were detected, as shown in the results of the metal-CRP complexes presented in the next sections.

Except for CRP5, higher quantities of recombinant proteins were obtained from Zn-supplemented cultures compared to Cu- and Cd-enriched ones (Table 2). This suggests that CRP1, -2, -3 and -4 proteins could have a stronger ability to bind Zn(II) than to bind Cu(I) or Cd(II). As for CRP5, similar amounts of proteins were obtained from both Cu and Zn-supplemented cultures. For all the five CRPs, Cd-supplemented cultures resulted in the lowest quantity of purified protein. This result can hardly be ascribed to a specific effect of Cd on bacterial cells. Indeed, all three metals were added to identical bacterial cultures at the end of the culture period, after the induction of protein synthesis and only 3 h before protein extraction. Furthermore, the same production protocol, including the same plasmid background, the same bacterial strain and the same cultivation time, when used with other recombinant MTs, resulted in higher protein yields in presence of Cd compared to other metals [6].

Table 2

Recombinant protein synthesis yields and metal-to-protein ratios for the different metal-CRP products (as deduced from the ICP-AES data). Nd, Non-detectable metal-CRP product.

Products	Protein yield (g·L ⁻¹)	Zn/prot	Cd/prot	Cu/prot
Zn-CRP1	0.2	5.9	Nd	Nd
Cd-CRP1	0.02	Nd	9.5	Nd
Cu-CRP1	0.05	1.1	Nd	6.6
Zn-CRP2	0.8	8.7	Nd	Nd
Cd-CRP2	0.2	Nd	7.1	Nd
Cu-CRP2	0.3	2.0	Nd	9.9
Zn-CRP3	1.2	7.2	Nd	Nd
Cd-CRP3	0.03	Nd	11.3	Nd
Cu-CRP3	0.2	3.1	Nd	8.4
Zn-CRP4	0.8	7.2	Nd	Nd
Cd-CRP4	0.04	Nd	9.0	Nd
Cu-CRP4	0.08	2.1	Nd	13.5
Zn-CRP5	0.5	8.2	Nd	Nd
Cd-CRP5	0.06	Nd	13	Nd
Cu-CRP5	0.5	1.9	Nd	7.3

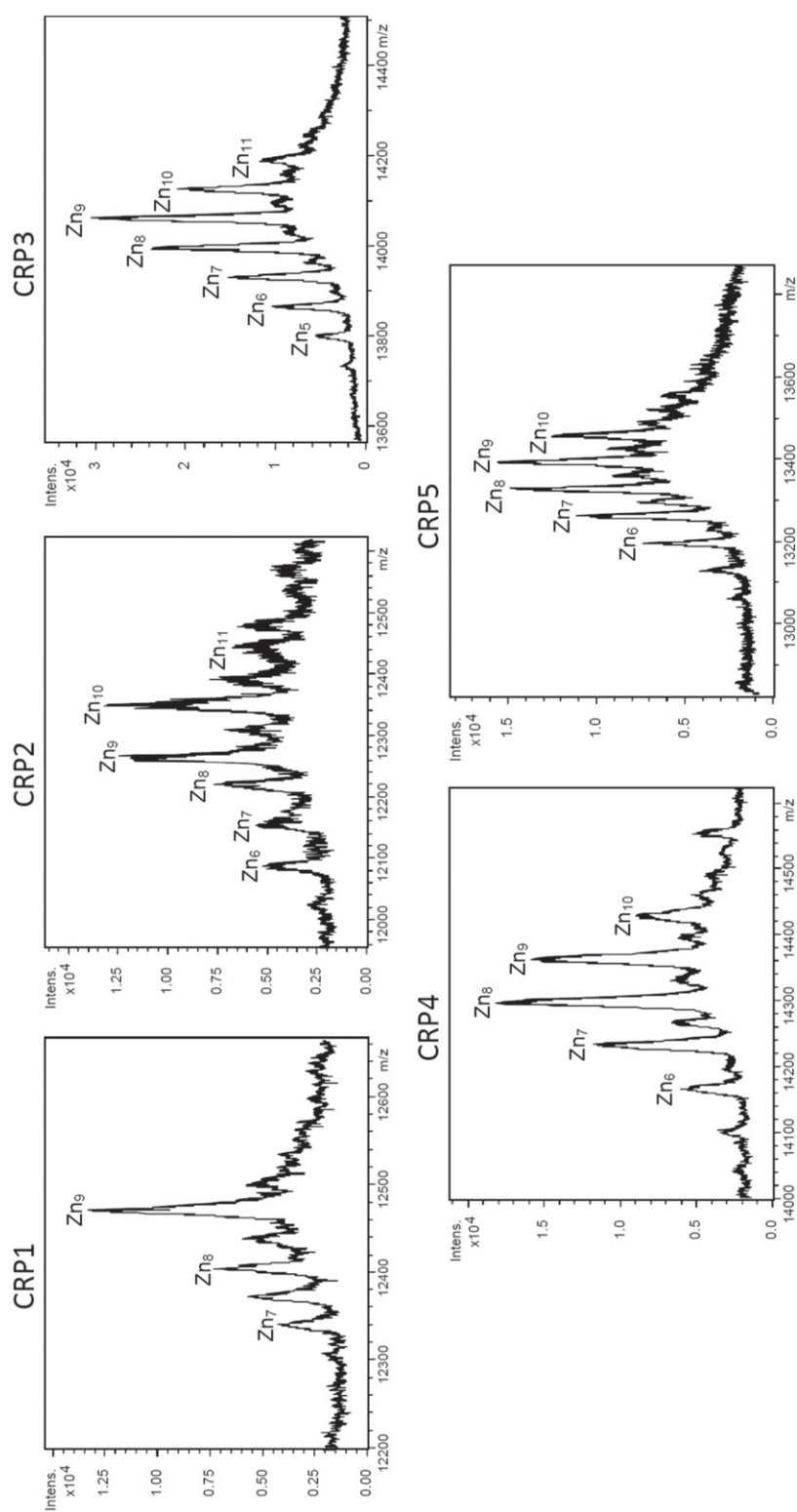


Fig. 3. Deconvoluted ESI-MS spectra of the recombinant Zn-CRP preparations run at neutral pH.

We also observed that all metal-CRP complexes eluted from the FPLC columns as two partially overlapping peaks (data not shown), which were kept as separated fraction pools. These two peaks could have resulted from the systematic presence of dimers in the preparations. Invariably, ESI-MS spectra of the first peaks were less intense, and showed the presence of a great number of metal complex species, some of them dimers and other monomers, of difficult interpretation. For this reason, samples corresponding to the first peaks were discarded, and all the results reported below refer to the second peaks of the FPLC chromatography, which yielded cleaner, definite spectra more easily interpretable (Figs. 3, 4 and 5). The presence of dimeric forms have been extensively reported when studying the metal binding properties of the separate moieties of the *Q. suber* MT2 isoform [37], as well as several fungal MT forms (data not shown).

3.3. Characterization of the *in vivo* metal-binding abilities of recombinant CRP proteins

Following methodologies already established [44], all recombinant metal-CRP preparations were subjected to ESI-MS and ICP analyses to determine each CRP-Zn(II), -Cd(II) and -Cu(I) metal-binding features.

3.3.1. Zn(II)-binding features

ESI-MS analyses of the five CRP proteins purified from Zn-enriched cultures, revealed in all cases the presence of a mixture of several Zn-loaded complexes (Fig. 3 and Table 1). Whatever the Cys and His content of the polypeptides, these complexes comprise the same range of metal species, from Zn₅- to Zn₁₁-CRPs, but with different relative abundances. Thus, while Zn₉-CRP1 was clearly the major species for CRP1, an almost equimolar mixture of Zn₉-CRP and Zn₁₀-CRP was detected for CRP2, a difference potentially attributable to the additional Cys that this polypeptide harbors compared to the former. For the other three CRPs, all containing 27 Cys residues, the main species identified were, in decreasing order of ESI-MS peak intensity: Zn₉- > Zn₈- > Zn₁₀-

> Zn₇-CRP3; Zn₈- > Zn₉- > Zn₇- > Zn₁₀-CRP4; and Zn₉- > Zn₈- > Zn₁₀- > Zn₇-CRP5 (Fig. 3). The lower Zn(II) content observed for CRP4 may be attributed to the fact that it only contains two His residues, compared to CRP3 and CRP5, which contain three and four His, respectively.

3.3.2. Cd(II)-binding features

As stated above, the Cd-supplemented cultures resulted in very low quantities of Cd-CRP complexes, and for CRP2, despite several attempts, no protein could be recovered at all from the Cd-enriched *E. coli* cultures. This already suggested a very poor Cd-binding ability of these peptides, which was corroborated by the characterization of all the four recovered CRP samples. All of them consisted in mixtures of complexes, some of them devoid of, and some others including, inorganic S²⁻ ligands (Fig. 4, Table 1), a typical feature of MTs with a Cu-thionein character when binding Cd(II) ions [6]. CRP1 yielded as major species Cd₁₁S-CRP1, followed by Cd₁₂S₂-, Cd₁₅-, Cd₁₃S₂- and Cd₁₀-CRP1, and other minor complexes. For CRP3, the major species also contained sulfide anions (Cd₁₂S-CRP3), and the other major complexes detected were Cd₁₂-, Cd₁₁S-, and Cd₁₃S₂-CRP3. Strikingly for CRP4, despite its higher Cys content if compared to CRP1, a highly predominant Cd₉-CRP4 complex was purified, accompanied by minor Cd₁₂S-, Cd₉S₂-, and Cd₁₂-CRP4. Finally, CRP5 preparation resulted in the most complex mixture of complexes (abundances in decreasing order: Cd₁₂S- > Cd₁₅S- > Cd₁₄S₂- > Cd₁₆- > Cd₁₂-CRP5) but that rendered the species with the highest Cd content (Cd₁₆-CRP5). The ICP analyses of all these preparations gave mean Cd-per-CRP ratios that were fully concordant with the results of ESI-MS analyses (Tables 1 and 2).

When comprehensively considering these results, it is obvious that the presence of different number of Cys and/or His is not proportionally related with the maximum Cd(II) content of the complexes as it was for Zn(II). This observation can be explained if assuming that the importance of the S²⁻ ligand is more determinant for constructing stable Cd-CRP complexes than the more variable contribution of His bonds, as shown for plant MTs [37,45].

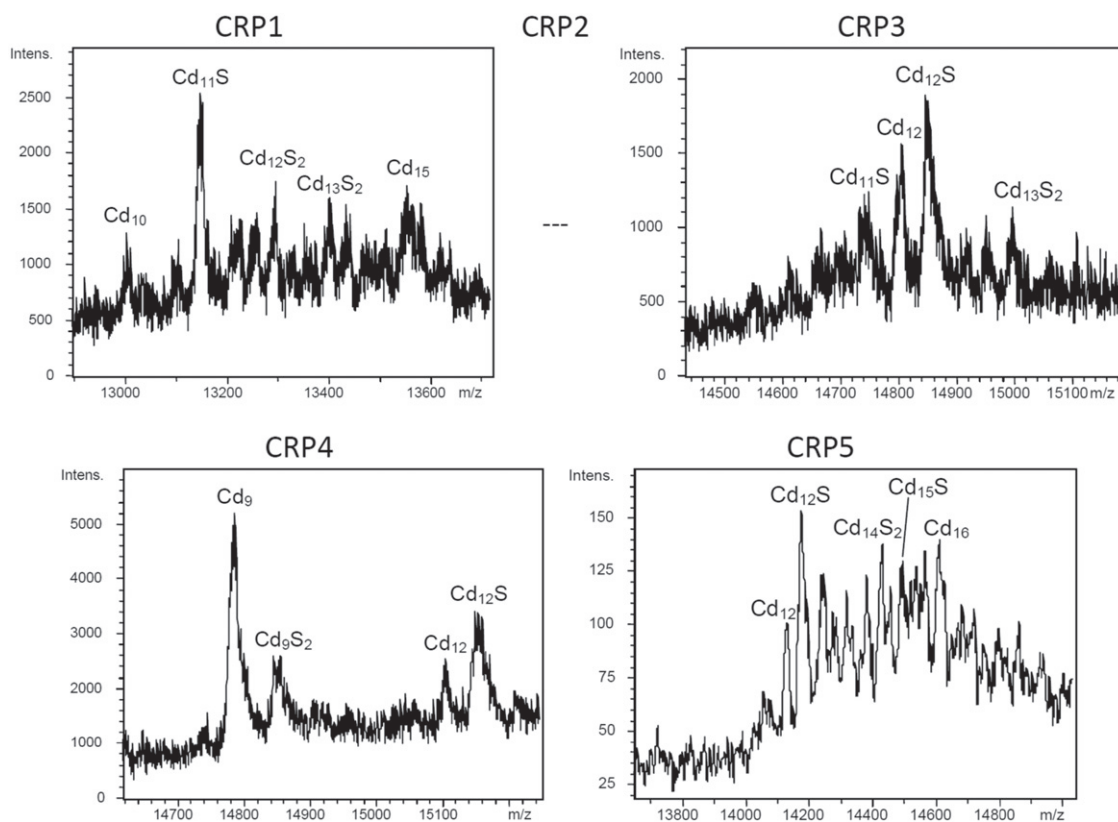


Fig. 4. Deconvoluted ESI-MS spectra of the recombinant Cd-CRP preparations run at neutral pH. Spectra were slightly smoothed when necessary.

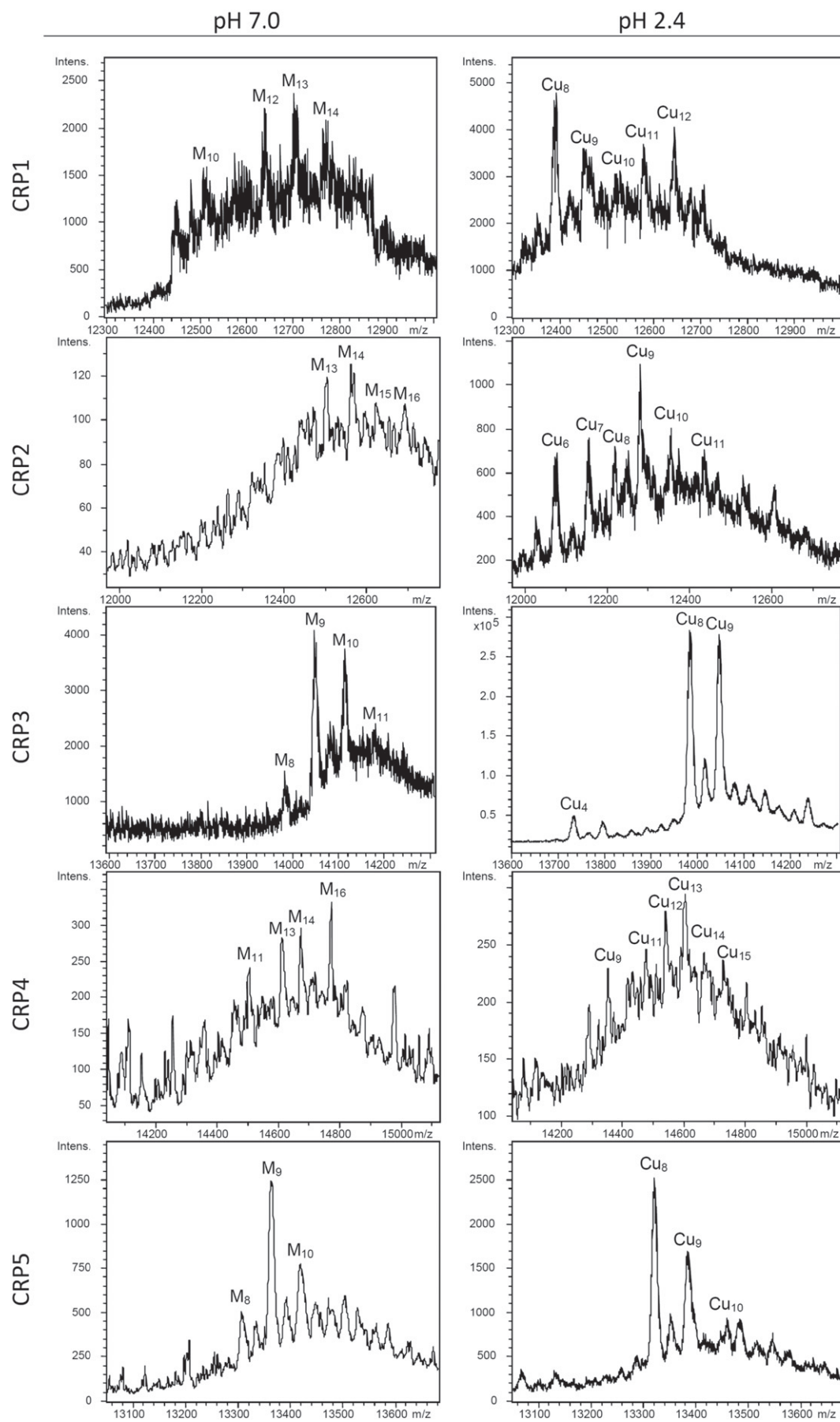


Fig. 5. Deconvoluted ESI-MS spectra of the recombinant Cu-CRP preparations run at neutral and acidic pH. Spectra were slightly smoothed when necessary. M = Zn + Cu.

3.3.3. Cu(I)-binding features

It is common that, unless exhibiting an extreme Cu-thionein character, MTs synthesized in Cu-supplemented media include some Zn(II) in their corresponding complexes, due to the permanent presence of Zn in culture media and in the intracellular environment [44]. In order to determine the Zn and Cu content of the metal-CRP complexes synthesized in Cu-supplemented media, two strategies were followed. First, all of the samples were analyzed by ESI-MS at both neutral (7.0) and acidic (2.4) pH, because, at acidic pH, only the Cu(I) metal ions remain bound to the proteins, but usually not the divalent metal ions, such as Zn(II). Therefore, differences between neutral and acidic ESI-MS detected species result from the presence of Zn(II) in the complexes (Fig. 5, Table 1). Second, ICP-AES was also used to confirm the simultaneous presence of Zn and Cu ions in the same sample, as suspected from ESI-MS measurements (Table 2).

It is clear that the five CRP polypeptides preparations were mixtures of heterometallic complexes, but, in some cases, also included homometallic Cu-species. CRP3 and CRP5 were the two proteins which resulted in samples with a higher Cu:Zn ratio. This is in accordance with the following ESI-MS results. For CRP3, at neutral pH, ESI-MS detected M_9^- and M_{10}^- -CRP3 complexes ($M = \text{Zn(II)}$ or Cu(I)) in approximately equal abundance, which were resolved as major Cu_8^- and Cu_9^- -containing species at acidic pH. For CRP5, M_9^- , M_{10}^- and M_8^- -CRP5 in decreasing abundance order were detected at neutral pH, and they contained major Cu_8^- and minor Cu_9^- -cores, as revealed by acid ESI-MS analysis. Paradoxically, CRP4, with fewer His than CRP3 and CRP5, resulted in a mixture of complexes with higher nuclearity (M_{16}^- , M_{14}^- , M_{13}^- and M_{11}^- -CRP4, in this order of relative abundance) and concordant higher Cu contents (Cu_{13}^- , Cu_{12}^- , Cu_{11}^- , Cu_{14}^- , Cu_{15}^- and Cu_9^- -CRP4). On the other hand, the CRP1 and -2 peptides, gave similar results with the metal stoichiometry slightly enhanced for CRP2 probably owing to its additional Cys compared to CRP1. Hence, for CRP1 the species identified by ESI-MS were M_{13}^- , M_{12}^- and M_{14}^- -CRP1 (with Cu_8^- , Cu_{12}^- , Cu_{11}^- and Cu_9^- -cores); and for CRP2, they were M_{14}^- , M_{13}^- , M_{16}^- and M_{15}^- -CRP2 (with Cu_9^- , Cu_{10}^- , Cu_7^- , Cu_8^- , Cu_6^- and Cu_{11}^- -cores), always in decreasing order of intensity of the corresponding ESI-MS peaks.

3.3.4. Concluding remarks on metal binding features

Overall biochemical data revealed that CRP recombinant proteins are able to chelate Cd(II), Zn(II) and Cu(I), but with significant differences between the three metal ions. The best coordinating ability is exhibited towards Zn(II), as suggested by the observation that Zn-supplemented cultures were those yielding the highest amounts of recombinant metal-CRP species and by the fact that these complexes were hardly demetalated at pH 2.4, which is concordant with a strong and/or buried binding of the metal ions. It is also supported by the observation that the ESI-MS spectra of the Zn-CRPs species were those with the highest quality (Fig. 3). At the opposite, all five CRP polypeptides revealed relatively low propensity to form Cd-CRP complexes. Cultures for recombinant CRP production gave the lowest amount of protein in the presence of Cd and resulted in a heterogeneous mixture of Cd-species, most of them containing S^{2-} ligands, which is distinctive of Cu-thioneins when coordinating Cd^{2+} . In other words, this is indicative of a poor ability of a peptide to fold into Cd-complexes [6]. Finally, the analysis of the Cu-supplemented synthesis of the five CRPs did not indicate that they behave as extreme Cu-thioneins, since all of them were recovered as heterometallic Zn-Cu-species.

Paradoxically, although all CRPs are conferring Cd-resistance in yeast, surprisingly they had difficulties in building Cd-complexes. However, these two features are not as contradictory as it may seem, since, although not optimally, all CRPs are able to bind Cd. Moreover, it was already reported that there is no an absolute correlation between the capacity of a MT to bind ions and its capacity at conferring metal tolerance in yeast cells. This was the case for two plant MT peptides, which despite producing metal complexes with exactly the same stoichiometry, conferred different metal resistance patterns in yeast, depending on the

presence or not of the non-metal coordinating spacer in their sequence [37]. Moreover, the ability of a recombinant MT to confer metal tolerance in yeast is not necessarily related to the biological function that this particular MT plays in its original native environment. In the study of Palacios et al. [46], it was shown that the HpCuMT which participates to metal homeostasis but not detoxication in the snail from which it originates could however confers Cu tolerance in yeast. It is also worth noting that the ability to confer resistance to Zn or Cd is not the same for the 5 CRP isoforms. A higher direct relationship between metal ion coordination ability and capacity to confer metal ion tolerance is observed for Zn than for Cd (this work (Fig. 2A Section 3.1) and [13]), which correlates with the hypothesis that these proteins may be more related to Zn metabolism than to Cd detoxification.

4. General conclusions

The present study represents the first in-depth biochemical characterization of “environmental proteins” encoded by eukaryotic genes of unknown taxonomic origin, selected through an environmental metagenomics approach. These Cysteine-Rich Proteins (CRPs), like known MTs (i) exhibit definite Cys motives in their sequences, which contain few aromatic amino acids, (ii) are able to confer Cd and Zn tolerance in a heterologous host (the yeast *S. cerevisiae*), and (iii) can bind *in vitro* Zn, Cd and Cu ions, with a preference towards Zn(II). They exhibit peculiarities among all MT sequences reported so far that are mainly reminiscent of MTs from three unrelated taxa: plants, unicellular fungi and ciliates. However, as CRPs cannot be assigned by similarity analysis to any known MT family, we propose that they should be considered a new family that could be called EMT for “environmental metallothioneins”, until its taxonomic origin is unveiled.

Acknowledgements

The authors are thankful to the Indo-French Centre for the Promotion of Advanced Research (IFCPAR) for sponsoring the project 4709-1, to the Rhône-Alpes region for a CMIRA travel grant (to AZ), to the French *Ministère de l'Enseignement Supérieur et de la Recherche* for salary grant (to AZ) and to the Spanish *Ministerio de Economía y Competitividad* (MINECO), grants BIO2015-67358-C2-1-P (to SA) and -C2-2-P (to MC), which are co-financed by the European Union through the FEDER program. Authors from Barcelona universities are members of the 2014SGR-423 *Grup de Recerca de la Generalitat de Catalunya*. We thank the *Centres Científics i Tecnològics (CCiT)* de la *Universitat de Barcelona* (ICP-AES, DNA sequencing) and the *Servei d'Anàlisi Química (SAQ)* de la *Universitat Autònoma de Barcelona* (CD, UV-vis, ESI-MS) for allocating instrument time. We kindly thank Agnès Rodrigue (UMR CMRS 5240 – Microbiology, Adaptation and Pathogeny, University Lyon 1) for useful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jinorgbio.2016.11.017.

References

- [1] M. Capdevila, R. Bofill, O. Palacios, S. Atrian, State-of-the-art of metallothioneins at the beginning of the 21st century, *Coord. Chem. Rev.* 256 (2012) 46–62.
- [2] C.A. Blindauer, Metallothioneins, in: W. Maret, A. Wedd (Eds.), *RSC Metallobiology Series No. 2: Binding, Transport and Storage of Metal Ions in Biological Cells*, RSC Publishing, Cambridge, UK 2014, pp. 594–653.
- [3] P.A. Binz, J.H.R. Kagi, Metallothionein: molecular evolution and classification, in: C. Klaassen (Ed.), *Metallothionein IV*, Birkhauser Verlag, Basel, CH 1999, pp. 7–13.
- [4] M. Capdevila, S. Atrian, Metallothionein protein evolution: a miniassay, *J. Biol. Inorg. Chem.* 16 (2011) 977–989.
- [5] M. Valls, R. Bofill, R. Gonzalez-Duarte, P. Gonzalez-Duarte, M. Capdevila, S. Atrian, A new insight into metallothionein (MT) classification and evolution. The *in vivo* and *in vitro* metal binding features of *Homarus americanus* recombinant MT, *J. Biol. Chem.* 276 (2001) 32835–32843.

- [6] R. Boffill, M. Capdevila, S. Atrian, Independent metal-binding features of recombinant metallothioneins convergently draw a step gradation between Zn- and Cu-thioneins, *Metallomics* 1 (2009) 229–234.
- [7] S.M. Adl, A.G.B. Simpson, C.E. Lane, J. Lukes, D. Bass, S.S. Bowser, M.W. Brown, F. Burki, M. Dunthorn, V. Hampl, A. Heiss, M. Hoppenrath, E. Lara, L. Le Gall, D.H. Lynn, H. McManus, E.A.D. Mitchell, S.E. Mozley-Stanridge, L.W. Parfrey, J. Pawlowski, S. Rueckert, L. Shadwick, C.L. Schoch, A. Smirnov, F.W. Spiegel, The revised classification of eukaryotes, *J. Eukaryot. Microbiol.* 59 (2012) 426–493.
- [8] L.A. Katz, J.R. Grant, Taxon-rich phylogenomic analyses resolve the eukaryotic tree of life and reveal the power of subsampling by sites, *Syst. Biol.* 64 (2015) 406–415.
- [9] C.A. Blindauer, O.I. Leszczyszyn, Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more, *Nat. Prod. Rep.* 27 (2010) 720–741.
- [10] M. Taupp, K. Mewis, S.J. Hallam, The art and design of functional metagenomic screens, *Curr. Opin. Biotechnol.* 22 (2011) 465–472.
- [11] S.Y. Moon-van der Staay, V.A. Tzeneva, G.W. van der Staay, W.M. de Vos, H. Smidt, J.H. Hackstein, Eukaryotic diversity in historical soil samples, *FEMS Microbiol. Ecol.* 57 (2006) 420–428.
- [12] S.T. Bates, J.C. Clemente, G.E. Flores, W.A. Walters, L.W. Perfrey, R. Knight, N. Fierer, Global biogeography of highly diverse protistan communities in soil, *ISME J.* 7 (2013) 652–659.
- [13] F. Lehembre, D. Doillon, E. David, S. Perrotto, J. Baude, J. Foulon, L. Harfouche, L. Vallon, J. Poulain, J. Da Silva, P. Wincker, C. Oger-Desfeux, P. Richaud, J. Colpaert, M. Chalot, L. Fraissinet-Tachet, D. Blaudez, R. Marmeisse, Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes, *Environ. Microbiol.* 15 (2013) 2829–2840.
- [14] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (2004) 1792–1797.
- [15] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton, Jalview version 2—a multiple sequence alignment editor and analysis workbench, *Bioinformatics* 25 (2009) 1189–1191.
- [16] M.S. Szczypka, J.A. Wemmie, W.S. Moye-Rowley, D.J. Thiele, A yeast metal resistance protein similar to human cystic fibrosis transmembrane conductance regulator (CFTR) and multidrug resistance-associated protein, *J. Biol. Chem.* 269 (1994) 22853–22857.
- [17] Z.S. Li, Y.P. Lu, R.G. Zhen, M. Szczypka, D.J. Thiele, P.A. Rea, A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae*: YCF1-catalyzed transport of bis(glutathionato)cadmium, *Proc. Natl. Acad. Sci. U. S. A.* 94 (1997) 42–47.
- [18] L. Li, J. Kaplan, Defects in the yeast high affinity iron transport system result in increased metal sensitivity because of the increased expression of transporters with a broad transition metal specificity, *J. Biol. Chem.* 273 (1998) 22181–22187.
- [19] A. Kamizono, M. Nishizaw, Y. Teranishi, K. Murata, A. Kimura, Identification of a gene conferring resistance to zinc and cadmium ions in the yeast *Saccharomyces cerevisiae*, *Mol. Gen. Genet.* 219 (1989) 161–167.
- [20] D. Gietz, A. St Jean, R.A. Woods, R.H. Schiestl, Improved method for high-efficiency transformation of intact yeast-cells, *Nucleic Acids Res.* 20 (1992) 1425.
- [21] N. Cols, N. Romero-Isart, M. Capdevila, B. Oliva, P. Gonzalez-Duarte, R. Gonzalez-Duarte, S. Atrian, Binding of excess cadmium(II) to Cd₇-metallothionein from recombinant mouse Zn₇-metallothionein 1. UV-VIS absorption and circular dichroism studies and theoretical location approach by surface accessibility analysis, *J. Inorg. Biochem.* 68 (1997) 157–166.
- [22] J. Bongers, C.D. Walton, D.E. Richardson, J.U. Bell, Micromolar protein concentrations and metalloprotein stoichiometries obtained by inductively coupled plasma. Atomic emission spectrometric determination of sulfur, *Anal. Chem.* 60 (1988) 2683–2686.
- [23] D. Fabris, J. Zaia, Y. Hathout, C. Fenselau, Retention of thiol protons in two classes of protein zinc ion coordination centers, *J. Am. Chem. Soc.* 118 (1996) 12242–12243.
- [24] E. Freisinger, Plant MTs—long neglected members of the metallothionein superfamily, *Dalton Trans.* 47 (2008) 6663–6675.
- [25] E. Freisinger, Structural features specific to plant metallothioneins, *J. Biol. Inorg. Chem.* 16 (2011) 1035–1045.
- [26] O.I. Leszczyszyn, H.T. Imam, C.A. Blindauer, Diversity and distribution of plant metallothioneins: a review of structure, properties and functions, *Metallomics* 5 (2013) 1146–1169.
- [27] T. Dudev, C. Lim, Metal binding affinity and selectivity in metalloproteins: insights from computational studies, *Annu. Rev. Biophys.* 37 (2008) 97–116.
- [28] E. Freisinger, Cadmium in metallothioneins, *Met. Ions Life Sci.* 11 (2013) 339–371.
- [29] P. Palumaa, Copper chaperones. The concept of conformational control in the metabolism of copper, *FEBS Lett.* 587 (2013) 1902–1910.
- [30] S. Diaz, F. Amaro, D. Rico, V. Campos, L. Benitez, A. Martin-Gonzalez, E.P. Hamilton, E. Orias, J.C. Gutiérrez, *Tetrahymena* metallothioneins fall into two discrete subfamilies, *Plos One* 3 (2007) 1–14.
- [31] R.N. Shuja, A.R. Shakoori, Identification and cloning of first cadmium metallothionein like gene from locally isolated ciliate, *Paramecium* sp, *Mol. Biol. Rep.* 36 (2009) 549–560.
- [32] S. Dar, R.N. Shuja, A.R. Shakoori, A synthetic cadmium metallothionein gene (PMCd1syn) of *Paramecium* species: expression, purification and characteristics of metallothionein protein, *Mol. Biol. Rep.* 40 (2013) 983–997.
- [33] C. Ding, R.A. Festa, Y.L. Chen, A. Espart, O. Palacios, J. Espin, M. Capdevila, S. Atrian, J. Heitman, D. Thiele, *Cryptococcus neoformans* copper detoxification machinery is critical for copper virulence, *Cell Host Microbe* 13 (2013) 265–276.
- [34] P. Iturbe-Espinoza, S. Gil-Moreno, W. Lin, S. Calatayud, O. Palacios, M. Capdevila, S. Atrian, The fungus *Tremella mesenterica* encodes the longest metallothionein currently known: gene, protein and metal binding characterization, *Plos One* 11 (2016), <http://dx.doi.org/10.1371/journal.pone.0148651>.
- [35] J.C. Gutiérrez, F. Amaro, S. Diaz, P. de Francisco, L.L. Cubas, A. Martín-González, Ciliate metallothioneins: unique microbial eukaryotic heavy-metal-binder molecules, *J. Biol. Inorg. Chem.* 16 (2011) 1025–1034.
- [36] P. de Francisco, L.M. Melgar, S. Díaz, A. Martín-González, J.C. Gutiérrez, The *Tetrahymena* metallothionein gene family: twenty-one new cDNAs, molecular characterization, phylogenetic study and comparative analysis of the gene expression under different abiotic stressors, *BMC Genomics* 17 (2016) 346–368.
- [37] J. Domenech, R. Orihuela, G. Mir, M. Molinas, S. Atrian, M. Capdevila, The Cd^{II}-binding abilities of recombinant *Quercus suber* metallothionein: bridging the gap between phytochelatins and metallothioneins, *J. Biol. Inorg. Chem.* 12 (2007) 867–882.
- [38] J.C. Gutiérrez, F. Amaro, A. Martín-González, From heavy metal-binders to biosensors: ciliate metallothioneins discussed, *BioEssays* 31 (2009) 805–816.
- [39] R.A. Syring, T.H. Hoexum-Brouwer, M. Brouwer, Cloning and sequencing of cDNAs encoding for a novel copper-specific metallothioneins and two cadmium-inducible metallothioneins from the blue crab *Callinectes sapidus*, *Comp. Biochem. Physiol.* 125C (2000) 325–332.
- [40] A. Espart, M. Marin, S. Gil-Moreno, O. Palacios, F. Amaro, A. Martín-González, J.C. Gutiérrez, M. Capdevila, S. Atrian, Hints for metal-preference protein sequence determinants: different metal binding features of the five *Tetrahymena thermophila* metallothioneins, *Int. J. Biol. Sci.* 11 (2015) 456–471.
- [41] S. Garcia, M. Prado, R. Degano, A. Dominguez, A copper-responsive transcription factor, CFR1, mediates copper and cadmium resistance in *Yarrowia lipolytica*, *J. Biol. Chem.* 277 (2002) 37359–37368.
- [42] H. Dolderer, J. Hartmann, U. Weser, Metallothioneins and Related Chelators, in: A. Sigel, H. Sigel, R.K.O. Sigel (Eds.), *Metal Ions in Life Sciences*, RSC Publishing, Cambridge, UK 2009, pp. 83–106.
- [43] O.I. Leszczyszyn, R. Schmid, C.A. Blindauer, Toward a property/function relationship for metallothioneins: histidine coordination and unusual cluster composition in a zinc-metallothionein from plants, *Proteins* 68 (2007) 922–935.
- [44] O. Palacios, S. Atrian, M. Capdevila, Zn- and Cu-thioneins: a functional classification for metallothioneins? *J. Biol. Inorg. Chem.* 16 (2011) 991–1009.
- [45] T. Huber, E. Freisinger, Sulfide ions as modulators of metal–thiolate cluster size in a plant metallothionein, *Dalton Trans.* 42 (2013) 8878–8889.
- [46] O. Palacios, A. Pagani, S. Perez-Rafael, M. Egg, M. Hockner, A. Brandstatter, M. Capdevila, S. Atrian, R. Dallinger, Shaping mechanisms of metal specificity in a family of metazoan metallothioneins: evolutionary differentiation of mollusk metallothioneins, *BMC Biol.* 9 (4) (2011) 1–20.

Supplementary material

Table S1: MT sequences used to build PSSM matrices, MT data sets and HMM profiles.

Sequence names	Accession Number	Reference	No. of sequences used for PSSM matrix building	No. of Sequences used in MT protein datasets	Sequence No. after manual corrections	Sequences used to build HMM Profiles
Family 3 (Crustaceans)			PSI-BLAST threshold = 0.005	PSI-BLAST threshold = 0.05		
CsMT-1	P55949	[47]	8 (the same as CmMT)	49	49	53
CmMT	P55948		8 (the same as CsMT-1)			
CsCuMT-2	Q9U620		1	4	4	
Family 7 (Ciliates)			PSI-BLAST threshold = 10^{-5}	PSI-BLAST threshold = 10^{-10}		
TtheMTT1	ABF68755	[35]	4	20	20	30
TtheMTT2	AY204351		0	10	10	
Family 11 (Fungi IV)			PSI-BLAST threshold = 0.005	PSI-BLAST threshold = 10^{-10}		
MT1	P41927	[41]	4 (the same as MT2, 3 and 4)	4	4	4
MT2	P41928		4 (the same as MT1, 3 and 4)			
MT3	Q9HFC9		4 (the same as MT1, 2 and 4)			
MT4	Q9HFD0		4 (the same as MT1, 2 and 3)			
Family 15 (Plants)			PSI-BLAST threshold = 0.01	PSI-BLAST threshold = 10^{-10}		
MT1	CAA65008	[48]	41	84	65	210
MT2	CAA65009	[49]	39	93	102	
MT3	Q40256	[50]	11	28	16	
EC-1/MT4	P30569	[51]	6	26	27	

Table S2: Primers used for the amplification of partial N- or C-terminal CRP5 open reading frames. Each of these primers contained *NotI* restriction site (in *italics*) and in-frame start or stop codons (underlined) in forward (F) or reverse (R) primers, respectively.

Protein name	Primers (5'-3')
CRP5NF	TAAGCAGCGGCCGCATGACTACTTGTTGCTGCAAA
CRP5NR	TGCTTAGCGGCCGCTCAGGACACACAGCAAGGACA
CRP5CF	TAAGCAGCGGCCGCATGCCACAATGTGAGTGTGCA
CRP5CR	TGCTTAGCGGCCGCTCATCCTTCTCTTGGGGGACA

Table S3: Primer pairs used for amplification of the five CRP full-length open reading frames from their original cDNA-SfiIpFL61 plasmids [12] for the construction of GST-CRPs fusion reading frames. The two restrictions sites *Bam*HI and *Sal*I, added before the start codon and after the stop codon respectively, are in italics. The start and stop codons are underlined.

Protein name	Upstream primers (5'-3')	Downstream primers (5'-3')	Original cDNA plasmids
CRP1	CCGCGTGGATCCATGACTACTTGTGTTGC	CTCGAGTCGACTCACTTCACCTTCTCATG	Ycf06L1-84
CRP2	CCGCGTGGATCCATGACTACTTGTGTTGC	CTCGAGTCGACTCACTTTTCATGCTTACA	Ycf06L1-57
CRP3	CCGCGTGGATCCATGGCTACTTGTGTTGT	CTCGAGTCGACCTACTCACCTTGGGACA	Ycf06Lp-67
CRP4	CCGCGTGGATCCATGGCTACTTGTGTTGC	CTCGAGTCGACTCAATCACCACCATCTTG	YcfL1-96
CRP5	CCGCGTGGATCCATGACTACTTGTGCTGC	CTCGAGTCGACTCATCCTTCTCTTGGGGG	YcfL1-94

Supplementary references

- [47] M. Serra-Batiste, N. Cols, L.A. Alcaraz, A. Donaire, P. Gonzalez-Duarte, M. Vasak, The metal-binding properties of the blue crab copper specific CuMT-2: a crustacean metallothionein with two cysteine triplets, *J. Biol. Inorg. Chem.* 15 (2010) 759-776.
- [48] O. Schicht, E. Freisinger, Spectroscopic characterization of *Cicer arietinum* metallothionein 1, *Inorg. Chim. Acta* 362 (2009) 714-724.
- [49] X. Wan, E. Freisinger, The plant metallothionein 2 from *Cicer arietinum* forms a single metal-thiolate cluster, *Metallomics* 1 (2009) 489-500.
- [50] E. Freisinger, Spectroscopic characterization of a fruit-specific metallothionein: *M. acuminata* MT3, *Inorg. Chim. Acta* 360 (2007) 369-380.
- [51] E.A. Peroza, E. Freisinger, Metal ion binding properties of *Tricium aestivum* Ec-1 metallothionein: evidence supporting two separate metal thiolate clusters, *J. Biol. Inorg. Chem.* 12 (2007) 377-391.

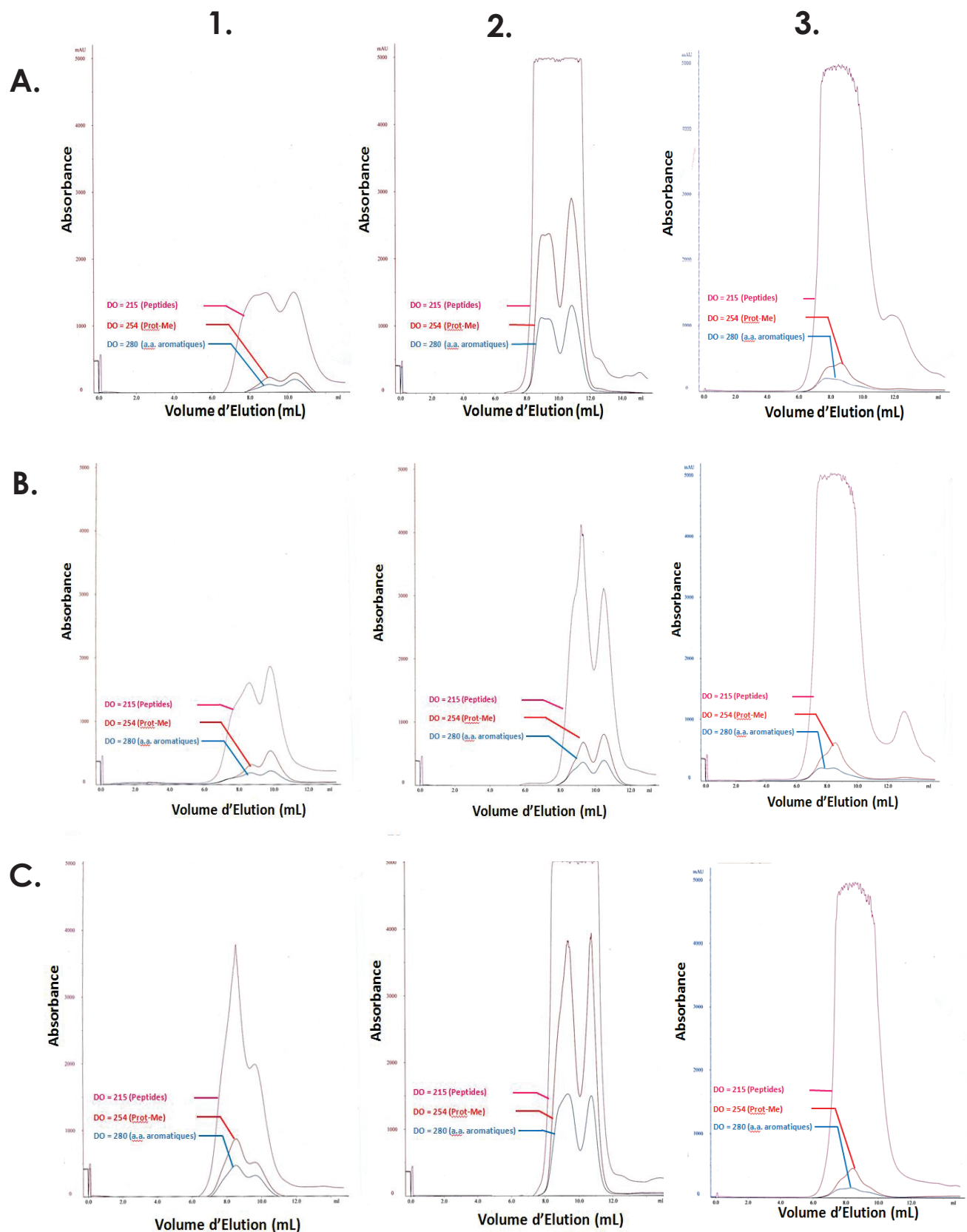


Figure II.2 : Absorbances des éluats après la dernière étape de chromatographie d'exclusion stérique (Colonne Superdex 75) à partir des productions de : **A.** CRP2, **B.** CRP3 et **C.** CRP5 supplémentés en **1.** Cd, **2.** Cu et **3.** Zn. Les chromatographes montrent les absorbances à 280 nm, 254 nm et 215 nm pour chaque fraction.



Données Supplémentaires

La **figure II.2** correspond aux quantités de protéines présentes (estimées par spectrophotométrie à différentes longueur d'ondes) dans les différentes fractions recueillies après chromatographie d'exclusion stérique lors de la purification de 3 CRP sur 5 purifiées. Ces données nous donnent deux indications :

- Les courbes sont souvent bimodales ce qui suggère la présence de dimères ;
- L'intensité des pics dépend des protéines et des métaux considérés. D'une manière générale, les quantités de CRP obtenues en présence de Zn sont plus importantes qu'en présence de Cu qui sont elles-mêmes en quantité plus élevée qu'en présence de Cd.

Conclusion

Les études réalisées ont permis de confirmer l'hypothèse, à savoir que les CRP sont des métallothionéines. En effet, elles sont capables de se lier *in vitro* à différents métaux (Cu, Zn et Cd) et permettent à des levures sensibles au Cd et au Zn de redevenir tolérantes à ces composés. De plus, étant donné qu'elles ne sont pas homologues aux séquences de MT déjà connues, elles constituent une toute nouvelle famille de métallothionéines que nous avons appelées EMT pour « *Environmental MetalloThioneins* ».

Ces EMT semblent être répandues dans l'environnement, ce qui nous amène à penser que les organismes qui les portent pourraient avoir un rôle fonctionnel important dans les sols, notamment dans des milieux pollués aux métaux. Aussi, les prochaines étapes consistent à déterminer si ces gènes peuvent être induits par des métaux et à tenter d'identifier les organismes qui portent ces gènes.





CHAPITRE 3

Mise au point de la quantification des gènes *emt* dans les sols

Introduction

Les métallothionéines (MT) constituent une superfamille de protéines chélatrices de métaux qui interviennent par exemple dans « la gestion » du stock d'éléments métalliques au sein de la cellule (**Bofill et al. 2009**). La caractérisation biochimique de ces protéines donne des indications sur leur capacité de liaison aux métaux et leur implication possible dans les réseaux métaboliques impliquant les métaux. Cependant, cette caractérisation ne donne que des informations sur l'implication potentielle des MT dans les processus cellulaires. L'analyse des niveaux d'expression des gènes correspondants, de leurs régions promotrices (**Stuart et al. 1984**), et des facteurs induisant leur transcription (conditions environnementales, états cellulaires,...) permet de préciser ce rôle fonctionnel. En effet, des organismes peuvent posséder plusieurs gènes codant des MT très similaires mais répondant à des signaux différents. Par exemple, *Mytilus galloprovincialis* (la moule) possède une MT (MT10) induite en présence de Cu, de Zn ou de Cd et une autre MT (MT20) seulement induite en présence de Cd dans l'environnement (**Dondero et al. 2005**).

Dans le cadre de l'étude des EMT, bien que nous sachions que ces métallothionéines sont capables de lier des métaux (**Cf chapitre 2**), nous ignorons toujours de quels organismes elles proviennent et si chez ces organismes leurs niveaux de transcription répondent à la présence d'un ou de plusieurs métaux dans l'environnement. Une approche indirecte pour répondre à cette dernière question serait de quantifier les transcrits *emt* dans



des microcosmes de sol traités ou non par différentes solutions de sels métalliques comme rapporté dans la littérature notamment pour des transcrits *trx* (thiorédoxine), quantifiés à partir de microcosmes de sol, induits en réponse à un stress oxydatif suite à des traitements aux nanoparticules de fer (**Fajardo et al. 2016**).

L'approche expérimentale privilégiée pour la quantification du nombre de copies de séquences cibles au sein d'un pool d'acides nucléiques environnementaux (ADN, ADNc) est la PCR quantitative dont la mise en œuvre nécessite de standardiser et de normaliser les mesures. Actuellement, deux techniques de PCR quantitative sont utilisées : la PCR en temps réel (qPCR) et la PCR « numérique » (*digital* en anglais) (dPCR).

La PCR en temps réel (**Heid et al. 1996**) est une méthode de PCR quantitative avec une mesure en continu du processus d'amplification qui est couramment utilisée pour détecter et quantifier des séquences d'ADN présentes dans un échantillon ou des séquences d'ARN après leur rétro-transcription en ADNc. Cependant, elle nécessite l'utilisation d'une gamme standard de concentrations connues d'ADN pour étalonner la mesure ainsi que des méthodes de calibration statistiques fiables pour s'assurer que la quantification soit correcte. Son principe réside dans : 1) l'ajout d'un fluorophore dans le milieu réactionnel avant l'amplification qui est soit spécifique aux séquences d'intérêt (sonde marquée) soit généraliste, ciblant tout ADN double brin (SYBR® Green, EvaGreen®,...) ; et 2) l'utilisation d'un thermocycleur couplé à un fluoromètre lors de l'étape d'amplification permettant le suivi de la réaction en temps réel par l'intermédiaire du niveau de fluorescence ($\approx$ proportionnel au nombre d'amplicons) en fonction du nombre de cycles. Ensuite, le nombre de cycles de PCR nécessaire pour dépasser un niveau de fluorescence de base (nommé Ct ou Cq) est mesuré pour l'échantillon et pour une gamme de concentration connue de la séquence cible. La concentration initiale en acides nucléiques de l'échantillon



est obtenue en rapportant les Cq mesurés de l'échantillon sur la gamme standard mesurée en parallèle (**Pfaffl 2006**).

La dPCR (**Pohl & Shih 2004**) est une méthode de PCR quantitative plus récente, basée sur la statistique de Poisson, pour laquelle une seule mesure finale du nombre d'amplicons est réalisée. Elle consiste en la séparation du mix PCR en un grand nombre de partitions au sein desquels les séquences d'intérêt à amplifier sont réparties aléatoirement. Ensuite, l'ADN est amplifié et le nombre de partitions positives (contenant des amplicons) et négatives sont comptées et converties en données numériques binaires (*digital*). La loi de Poisson permet de rapporter le nombre de réactions positives à la concentration initiale moyenne d'ADN. Cette technique est de plus en plus utilisée depuis l'automatisation très récente de l'étape de partition des échantillons via une puce microfluidique emprisonnant les partitions de l'échantillon dans de l'huile (jusqu'à 20 000 nanogouttelettes par échantillon) ce qui facilite grandement son utilisation et augmente sa sensibilité. On parle dans ce cas de ddPCR pour *droplet digital PCR* (**Hindson et al. 2011, Pinheiro et al. 2012**).

Plusieurs études ont comparé les avantages de la ddPCR par rapport à la qPCR. La ddPCR est en général plus sensible et plus reproductible que la qPCR pour les « faibles » concentrations en acides nucléiques (**Hindson et al. 2013**). En effet, lorsque l'on quantifie une séquence d'ADN rare dans un échantillon par qPCR il est nécessaire d'utiliser un nombre de cycles élevé. Mais plus le nombre de cycles est important et moins la qPCR est fiable et reproductible (**Buttigieg & Ramette 2014**). En séparant les éléments du milieu réactionnel dans des nanogouttelettes, la ddPCR est aussi moins sensible que la qPCR aux substances inhibitrices de l'ADN polymérase qui peuvent être en nombre important dans les extraits d'ADN environnementaux. Enfin, la ddPCR permet une quantification absolue contrairement à la qPCR qui nécessite une gamme standard pour rapporter le nombre de cycles quantifiés à la quantité d'acides nucléiques initiale.





Aussi, la ddPCR semble représenter une méthode de choix pour quantifier des extraits d'acides nucléiques environnementaux qui : 1) sont en faible abondance et 2) contiennent généralement des inhibiteurs de PCR. Toutefois, la supériorité de la ddPCR par rapport à la qPCR n'est pas garantie, notamment lorsqu'on utilise des extraits d'ADN de sol (**Kim et al. 2014**). Il semble impératif de comparer l'efficacité de la qPCR et de la ddPCR pour chaque type de sol étudié avant de choisir l'une des deux méthodes (**Kim et al. 2014**).

En PCR quantitative, dans le cas de la comparaison du niveau d'induction de gènes exprimés dans différentes conditions, il est essentiel de normaliser tous les échantillons pour pouvoir les comparer (**Bustin & Nolan 2004**). Pour cela deux méthodes sont possibles : 1) la quantification, à partir des mêmes extraits d'acides nucléiques de plusieurs gènes de référence qui doivent impérativement être exprimés à des niveaux similaires dans toutes les conditions (**Bustin et al. 2009**) ; 2) la quantification simultanée dans chaque condition du nombre de molécules d'ARN (ADNc) et d'ADN du gène étudié pour comparer le ratio $\frac{ADNc}{ADN}$ cibles dans chaque échantillon (**Baelum et al. 2008**, **Almario et al. 2013**). Cette quantification n'est possible que si le ratio $\frac{ADNc}{ADN}$ reste linéaire au cours de l'extraction quelle que soit les quantités d'ADN ou d'ARN extraites.

Les objectifs de ce chapitre ont été d'une part de mettre en place un système expérimental de microcosmes de sol traités par différentes concentrations en métaux afin d'étudier la réponse transcriptionnelle des gènes *emt*, et d'autre part de mettre au point une méthode de PCR quantitative (qPCR ou ddPCR) permettant de quantifier les séquences *emt* présentes au sein des pools d'ADN métagénomiques et d'ARN métatranscriptomiques extraits directement de ces microcosmes.



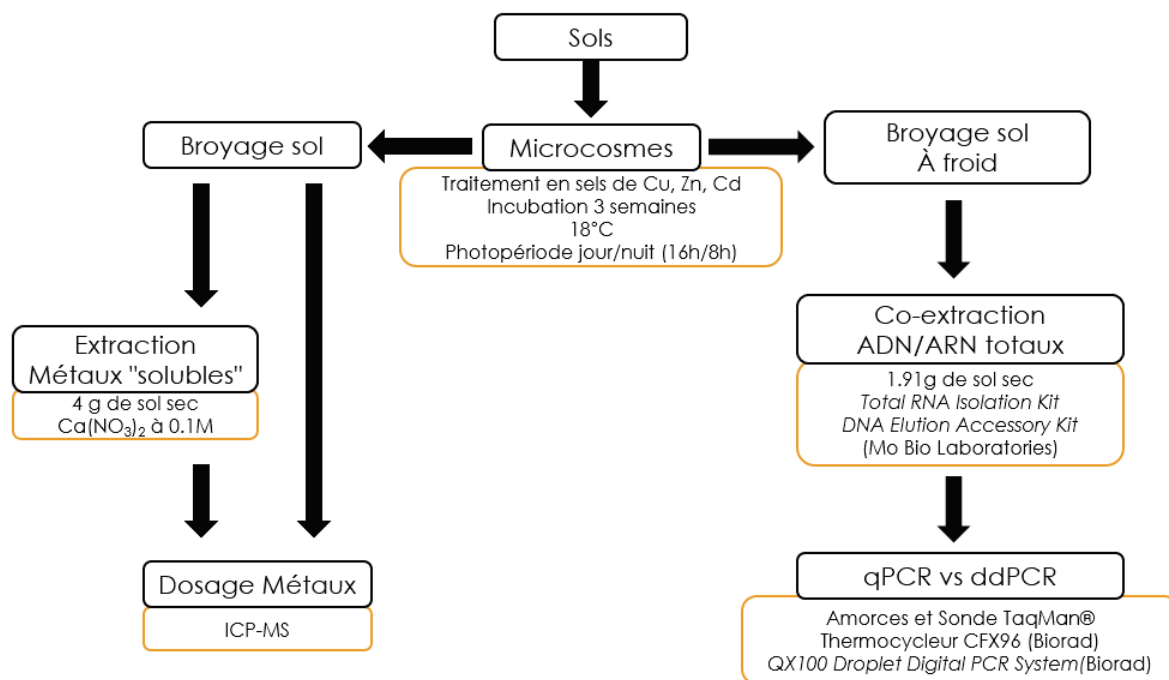


Figure III.1 : Schéma de la stratégie expérimentale.

Tableau III.1 : Caractéristiques du sol utilisé dans cette étude prélevé sur le site de Lommel (51° 14' N, 5° 15' E)

Type de végétation	Pinus sylvestris
Texture	Faiblement pozolisé, sableux
pH	4.7
% Matière organique	1.58
C/N	20
% d'humidité	4
Température du sol (°C)	11

Adapté de la thèse de **F. Lehembre, 2009**



Ce second volet a notamment nécessité la validation d'une méthode de co-extraction des ADN et ARN de sol, la conception de jeux d'amorces et de sondes spécifiques à une sous-famille adaptées à la PCR quantitative. Les différentes expériences réalisées sont schématisées **figure III.1**.

Matériels et méthodes

Sol étudié et préparation des microcosmes.

Les échantillons de sol proviennent du site de Lommel (51° 14' N, 5° 15' E) en Belgique (**Sonke et al. 2002**), à partir duquel les séquences *emt* des sous-familles 1 (*emt1*), 3 (*emt3*), 4 (*emt4*) et 5 (*emt5*) ont été initialement isolées (**Lehembre et al. 2013**). Les caractéristiques de ce sol, prélevé par Jan Colpaert, sont indiquées **tableau III.1**. Deux séries de microcosmes ont été réalisées successivement.

La première série a été réalisée dans le but de tester d'une part la procédure de traitement aux métaux et d'autre part la méthode d'extraction d'acides nucléiques. Pour concevoir ces microcosmes, le sol a été tamisé à 2 mm puis réparti en 5 lots de 1 800 g auxquels ont été ajoutés 100 mL d'une solution contenant soit, de l'eau distillée (témoin) soit du CdCl_2 permettant d'obtenir au final des sols à 3, 6, 12 et 24 ppm de Cd. La deuxième série de microcosmes a été réalisée dans le but de mettre au point la méthode de quantification des gènes *emt* par PCR et d'étudier leur induction en présence ou non de Cd, de Cu et de Zn. Pour la 2^{ème} série de microcosmes, le sol a été tamisé à 2 mm puis réparti en 6 lots de 1 800 g auxquels ont été ajoutés 100 mL d'une solution contenant soit de l'eau distillée (témoin) soit des solutions de sels métalliques (CdCl_2 , CuSO_4 ou ZnSO_4) permettant d'obtenir des sols enrichis en Cd à 3 et 6 ppm, en Cu à 60 et 120 ppm ou en Zn à 120 et 240 ppm.

Une fois les solutions ajoutées aux sols, ils ont été homogénéisés et chaque lot a été réparti en 12 microcosmes (pots) de 150 g chacun. Les microcosmes contenant chacun une plantule de *Pinus pinaster* âgée d'un



mois, ont été incubés à 18°C (photopériode 16 h/8 h lumière/obscurité) pendant 3 semaines avec 2 arrosages par semaine avec de l'eau distillée stérile. En fin d'incubation, quatre microcosmes d'une même condition expérimentale ont été rassemblés, mélangés et congelés à -80°C. Ceci a permis d'obtenir au final 3 échantillons composites (assemblage de 4 microcosmes) pour chaque condition.

Extraction et quantification des métaux.

L'estimation des métaux "solubles" a été faite *via* une extraction au nitrate de calcium. Cinq grammes de sol provenant de la première série de microcosmes (témoin, CdCl_2 à 3 ppm, 6 ppm, 12 ppm et 24 ppm) ont été déshydratés par lyophilisation toute la nuit. Les sols lyophilisés ont été broyés avec un broyeur de roches en agate pendant 3 min puis 4 grammes de sol sec ont été placés dans un erlenmeyer préalablement traité avec de l' HNO_3 10% toute la nuit puis rincé 3 fois à l'eau distillée afin d'éliminer les éléments métalliques résiduels. Après une incubation sous agitation de 2 h à température ambiante en présence de 20 mL de $\text{Ca}(\text{NO}_3)_2$ à 0,1M, les suspensions de sol ont été filtrées sur des filtres d'acétate de cellulose à 0,2 μm . Les filtrats acidifiés par 100 μL d' HNO_3 pur sont conservés à 4°C avant dosage.

Le dosage des métaux par ICP-MS a été réalisé par le Service d'Analyse des Roches et des Minéraux du Centre de Recherches Pétrographiques et Géochimiques de Nancy (CNRS), à partir des extraits au nitrate de calcium (pour le dosage des métaux solubles) ou à partir de 1 gramme de sol après un micro-broyage à 0,8 -1 mm et un traitement acide pour le dosage des métaux totaux afin de dissocier toutes les molécules pour obtenir un mélange d'éléments.



Co-extraction des ADN et des ARN.

Les sols ont été broyés pendant 3 min en présence d'azote liquide dans un broyeur de roches en agate préalablement nettoyé au SDS et à l'alcool à 70° puis refroidi à -80°C pendant une nuit (**Bailly et al. 2007**).

Les ADN et ARN totaux ont été co-extraits en utilisant le protocole du kit *RNA PowerSoil® (Total RNA Isolation Kit et DNA Elution Accessory Kit ; MoBio laboratories, Carlsbad, CA)* modifié. Ce protocole consiste dans un premier temps à séparer les différents constituants du sol (matière organique, acides nucléiques, protéines, matière minérale,...) par un broyage mécanique et un traitement chimique en mélangeant vigoureusement au vortex dans un tube de 15 mL pendant 15 min le sol avec du SDS, une solution tampon, du phénol/chloroforme/alcool isoamylique (25:24:1 ; vol:vol:vol) et des microbilles de carbure de silice. Après une centrifugation de 2 500 g à 4 °C pendant 10 min, les acides nucléiques présents dans la phase supérieure aqueuse sont précipités pendant 30 min en présence d'isopropanol. Après centrifugation, le culot d'acides nucléiques est repris dans de l'eau stérile. La solution est déposée dans une colonne de purification composée d'une membrane de silice, puis les acides nucléiques sont élués après un lavage avec une solution saline. En fonction de la solution utilisée pour l'élution finale, il est possible d'éluer successivement les ARN puis les ADN. Pour chaque condition, 2 réplicas de 1,91 g de sol sec chacun ont été extraits en parallèle, puis les acides nucléiques extraits ont été mélangés et placés à -20°C pour les ADN et -80°C pour les ARN.

La qualité et la quantité d'acides nucléiques ont été évaluées par séparation électrophorétique sur gel d'agarose 1% en parallèle avec une gamme d'acides nucléiques de quantités connues. La quantité des ADN a aussi été estimée par fluorométrie au Qubit 2.0® (ThermoFisher Scientific) en utilisant le kit « *Qubit dsDNA BR Assay Kit* » et la quantité d'ARN a été mesurée par spectrophotométrie (à 230, 260 et 280 nm) au Nanodrop® (ThermoFisher Scientific).



Tableau III.2 : Amorces et sonde utilisées en qPCR et ddPCR

Nom	Séquences (5' - 3')	Taille	Tm (°C)	Taille des amplicons (pb)
EMT1F	CGGAGAATGTCACGAGTGAGGTTTGCTG	28	62.6	276
EMT1R	ACTTGCTCTCCACACTCACACTTGGCC	26	62.6	
EMT1P	TGCTGTGCATCCGAGTCGGGAAAGAAGGTGGT	32	71.6	
EMT3F	GTTACGAGCCAGACATGCTGCCC	23	62.4	284
EMT3R	CTCCACACCCACACATCACTACCAAGG	27	62.4	
EMT4F	GAAGCGTGCTGCCCACCTACATGTTG	26	63.9	288
EMT4R	CACACTCTGCTCCACATCCACACACACGA	29	63.9	
EMT5F	AAAGTCTTGCTGTGTCACCGAGTGCA	26	62.5	249
EMT5R	CCACACTCTGCTCCACATCCACATACAAC	29	62.3	



Mise au point des amorces et sondes.

Quatre paires d'amorces spécifiques respectivement des sous-familles 1, 3, 4 et 5 ont été conçues à partir d'un alignement multiple de séquences appartenant à chacune de ces sous-familles. La spécificité de ces amorces a été testée *in silico* par BLASTn contre nr (NCBI) et *in vitro* par PCR à partir de 2 ng d'ADN plasmidique contenant une séquence des sous-familles 1 (HE716999), 3 (HE716994), 4 (HE716992) et 5 (HE716995) puis visualisée par électrophorèse sur gel d'agarose 1%. Les PCR ont été réalisées dans des milieux réactionnels de 25 µL de tampon PCR 1x (20 mM Tris-HCl pH 8.4, 50 mM KCl) contenant : 0,2 mM de dNTP ; 0,25 mg/mL de BSA ; 1,5 mM de MgCl₂ ; 0,2 µM de chaque amorce et 0,5 U de *Taq DNA polymerase* recombinante (ThermoFisher) avec le programme suivant : une pré-dénaturation à 94°C de 5 min, une dénaturation à 94°C d'1 min, une hybridation à 58°C d'1 min, une élongation à 72°C de 30 sec et une élongation finale à 72°C de 10 min. Les paires d'amorces testées sont EMT1F/EMT1R, EMT3F/EMT3R, EMT4F/EMT4R, EMT5F/EMT5R, EMT5F/EMT5R (**Tableau III.2**). Le produit de la PCR EMT1F/R a été cloné en utilisant le kit « *TOPO® TA cloning® kit for sequencing* ». La sonde TaqMan EMT1P (**Tableau III.2**) spécifique de la sous-famille 1 a été conçue à partir de l'alignement multiple de séquences *emt* de la sous-famille 1 en utilisant la procédure décrite par **Thornton & Basu (2011)** combinant plusieurs logiciels libres (*Primer3*, *Beacon Designer* et *mFold*) et vérifiée *in silico* par BLASTn contre nr (NCBI).

Préparation du standard d'ADN plasmidique pour les réactions de PCR quantitative.

La gamme standard a été réalisée en utilisant un plasmide contenant une séquence *emt* de la sous-famille 1 (HE716999). Le plasmide a été linéarisé par l'enzyme *NotI* qui ne coupe qu'une fois dans l'insert du plasmide. Puis une gamme de 6 dilutions en série a été réalisée dans de l'eau ultra pure afin



d'obtenir des nombres de copies théoriques de : 10^4 , 10^3 , 100 et 10 copies par réactions. Ce nombre de copies a été calculé en présumant que la masse moléculaire moyenne d'une paire de bases correspond à 660 Da.

Réactions d'amplification des ADN par PCR, qPCR et ddPCR

Les amplifications des séquences *emt* par PCR classique ont été réalisées en ajoutant 2 μ L d'extrait d'ADN de sol témoin (≈ 160 ng) au 25 μ L de milieu réactionnel de tampon PCR 1X (20 mM Tris-HCl pH 8.4, 50 mM KCl) contenant : 6,25 μ g de BSA ; 1,5 μ g de protéine *T4 Gene 32* ; 0,2 μ M de chaque amorce EMT1F/EMT1R ; 140 μ M de chaque dNTP ; 0,5 U de *Taq DNA Polymerase* ; 1,5 mM de $MgCl_2$ (Invitrogen). Les conditions d'amplification étaient : une dénaturation initiale à 94°C pendant 5 min, puis 40 cycles de dénaturation à 94°C pendant 1 min, d'hybridation à 59°C pendant 1 min et d'élongation à 72°C pendant 1 min. Après l'élongation finale à 72°C pendant 10 min, les produits PCR ont été déposés sur un gel d'agarose 1% pour visualiser le résultat de l'amplification.

Les qPCR et ddPCR ont été réalisées en utilisant les amorces EMT1F/R et la sonde TaqMan® EMT1P toutes trois spécifiques de la sous-famille 1 (**Tableau III.2**). Pour les deux méthodes, les réactions d'amplification ont été réalisées en utilisant un *thermocycler* Biorad CFX9, des consommables interagissant faiblement avec les molécules du mix PCR (*low binding*) et une quantité variable de matrice ADN plasmidique (ou d'eau dans le cas du control négatif) a été ajoutée pour obtenir des quantités initiales de 10^4 , 10^3 , 100 et 10 copies par réaction.

La qPCR, a été réalisée dans un volume de 20 μ L contenant : 1X de prémix PCR *Sso advance supermix* (Biorad) pour sondes (*Sso7d*-fusion polymérase, dNTP, $MgCl_2$, stabilisateurs de réaction) ; 0,3 μ M de chaque amorce ; 0,2 μ M de sonde TaqMan® EMT1P ; 1 μ g de protéine *T4 Gene 32* (Roche). La réaction d'amplification a été réalisée avec une étape de



dénaturation initiale à 95°C pendant 3 min suivie de 45 cycles de trois étapes : 95°C pendant 15 sec, puis 62,6°C pendant 30 sec, puis 72°C pendant 30 sec.

Pour la ddPCR, le milieu réactionnel d'un volume de 22 µL contient : 1X de prémix *ddPCR™ Supermix for Probes (No dUTP)* (Biorad) ; 0,9 µM de chaque amorce et 0,2 µM de sonde TaqMan® EMT1P. Vingt microlitres de mix et 70 µL d'huile de génération de gouttelettes ont ensuite été transférés dans les puits prévus à cet effet sur la puce microfluidique (*droplet generation cartridge* (Biorad)), puis celle-ci a été placée dans le générateur de gouttelettes (*QX100™ droplet generator* (Biorad)) permettant de réaliser la partition des échantillons en mélangeant l'échantillon avec l'huile pour obtenir une émulsion. Trente-neuf microlitres de chaque émulsion ont ensuite été transférés dans les puits d'une microplaque de 96 puits, avant l'amplification. Celle-ci a été réalisée avec une étape de dénaturation initiale à 95°C pendant 10 min suivie d'un cycle : dénaturation à 94°C (30 sec), hybridation/élongation à 63,8°C (1 min), répété 45 fois et une étape finale à 98°C pendant 10 min pour stabiliser les gouttelettes. Une fois l'amplification terminée, la plaque de 96 puits a été placée dans le *QX100™ droplet reader* (Biorad), qui est un cytomètre de flux comptant le nombre de gouttelettes positives et négatives. Le nombre de copies absolues d'ADN par échantillon ainsi que les échantillons acceptables (c'est-à-dire ayant un nombre de partitions supérieur à 8 000) ont été obtenus en traitant les données avec le logiciel *QuantaSoft™* (Biorad).

La sensibilité à l'inhibition des réactions de PCR quantitative a été testée en ajoutant au milieu réactionnel 1 ng d'extrait d'ADN de sol provenant des microcosmes témoins.



Synthèse des ADNc.

Suite à la co-extraction des acides nucléiques de sol, les éluats contenant des ARN ont été traités à la *DNase I RNase-free* (Fermentas) puis précipités à l'éthanol et repris dans 9,5 µL d'eau ultra pure stérile. La transcription inverse a été réalisée dans 21,1 µL de milieu réactionnel de tampon 1X (47 mM Tris-HCl pH 8.3 ; 47 mM KCl ; 3,8 mM MgCl₂ ; 9,5 mM DTT) contenant : 900 µM d'un mix de dNTP ; 10 U d'inhibiteur de RNase ; 200 U de *RevertAid Reverse Transcriptase* (M-MuLV, Thermo Scientific) et 0,2 µg de *Random Hexamers*. Le mélange a été incubé à 42°C pendant 1h puis à 70°C pendant 10 min pour inactiver l'enzyme avant d'être placé à -20°C. Les ADNc ont été quantifiés en ddPCR selon la procédure décrite dans la partie précédente.

Analyses statistiques des données

L'efficacité des réactions d'amplification (E) et le coefficient de détermination (R^2) ont été obtenus à partir des paramètres de la droite de régression linéaire ($y = ax + b$) où dans le cas de la qPCR, « y » est le nombre de C_q mesurés en fonction du nombre « x » de copies théoriques de plasmides exprimé en $\log_{10}$. Dans le cas de la ddPCR, « y » correspond au nombre de copies mesurées exprimé en $\log_{10}$ en fonction du nombre « x » de copies théoriques de plasmides exprimé en $\log_{10}$. L'efficacité a été calculée d'après les formules suivantes : $E_{qPCR} = 100 \times (10^{-1/a} - 1)$ et $E_{ddPCR} = 100 \times a$ où a représente le coefficient directeur de la pente du modèle linéaire de chaque courbe (**Rebrikov & Trofimov 2006**). La précision des mesures a été estimée en utilisant le coefficient de variation obtenu d'après la formule suivante : $CV(\%) = \left(1 + \frac{1}{4n}\right) \times 100 \left(\frac{\sigma}{\mu}\right)$ (**Scherrer 2007**) où n correspond au nombre de réplicas, σ à l'écart-type et μ à la moyenne du nombre de copies de gènes *emt* mesurées, exprimé en $\log_{10}$. Plus le CV est faible et moins les valeurs des mesures sont dispersées autour de la moyenne, donc plus la répétabilité des mesures est



bonne (**Cavé et al. 2016**). Un test de Student a été réalisé pour vérifier la significativité des estimations du coefficient directeur (a) et de l'ordonnée à l'origine (b) des droites de régression linéaire des PCR quantitatives où les signes *, ** et *** correspondent respectivement à des p -value de 0,05 ; 0,01 et 0,001. Toutes les analyses statistiques et les graphiques des **figures III.6 et 7** ont été réalisés à l'aide du logiciel R et du package « *tseries* » (<https://www.r-project.org/>).



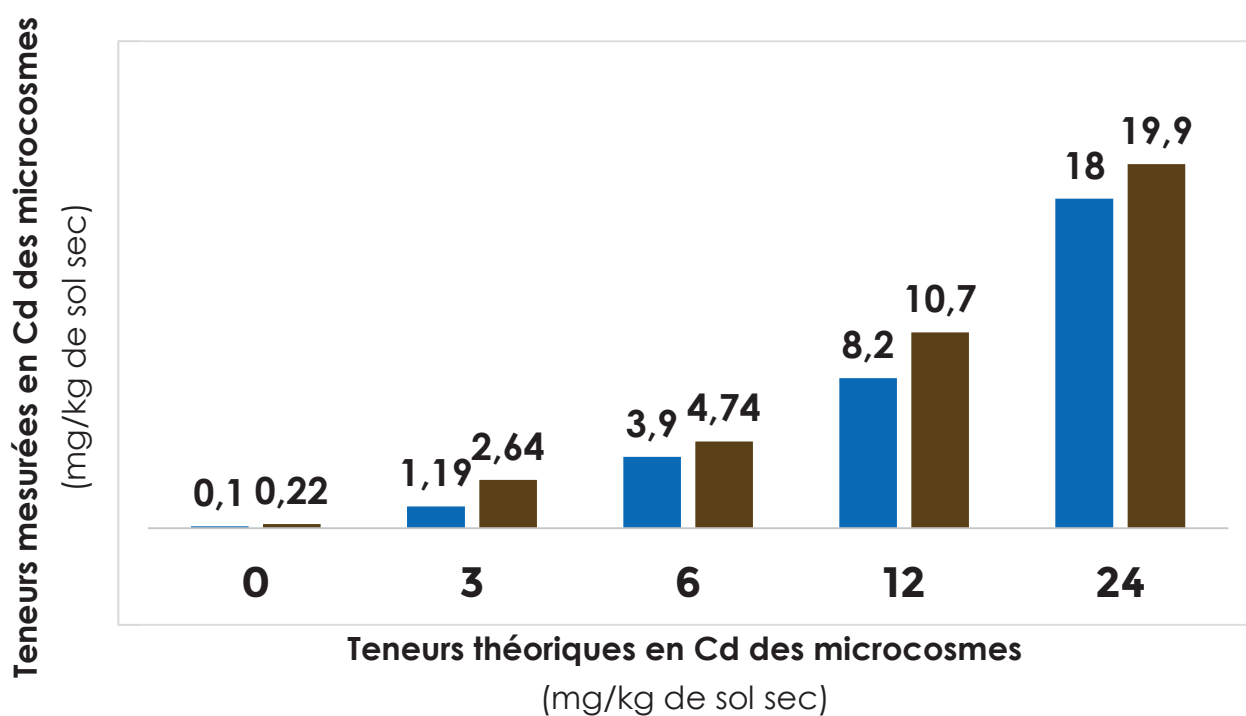


Figure III.2 : Diagramme représentant les différentes teneurs en Cd des microcosmes mesurées par ICP-MS en fonction de la teneur théorique en Cd (mg de Cd ajouté/kg de sol sec). Les barres bleues correspondent à la teneur en Cd "soluble" extrait par du $\text{Ca}(\text{NO}_3)_2$ (en mg/kg de sol sec) alors que les barres marron correspondent à la teneur en Cd total (en mg/kg de sol sec).



Résultats et discussion

Nous souhaitons mettre au point une méthode de quantification des transcrits *emt* à partir d'ARN et d'ADN co-extraits de sol de microcosmes afin d'évaluer leur réponse transcriptionnelle à l'apport de métaux. Pour cela, nous avons : (i) réalisé des microcosmes de sol pollués par différents métaux, (ii) conçu et testé un jeu d'amorces adapté à la réalisation de PCR quantitatives à partir d'extraits d'acides nucléiques de sol, (iii) développé une méthode d'amplification des gènes *emt* par PCR en 40 cycles, puis (iv) mis au point deux méthodes de PCR quantitative (par PCR en temps réel et par PCR numérique) afin de choisir celle qui serait la plus adaptée à notre objectif.

Réalisation de microcosmes de sol amendés en éléments métalliques

Deux séries de microcosmes ont été réalisées : la première amendée dans le but d'obtenir une gamme de concentrations croissantes de Cd à 3, 6, 12 et 24 ppm ; la seconde contenant 3 ou 6 ppm de Cd, 60 ou 120 ppm de Cu, 120 ou 240 ppm de Zn. Ces concentrations ont été choisies pour qu'elles soient physiologiquement actives sur la composante microbienne du sol (Alloway *et al.* 2013).

• Quantification des métaux totaux et "solubles" dans les sols amendés par différentes concentrations en Cd.

La première série de microcosmes a servi de test pour la procédure d'amendement en métaux. Les concentrations de métaux totaux et "solubles" ont été dosées pour chaque condition de microcosmes (témoin, 3 ppm, 6 ppm, 12 ppm et 24 ppm de Cd) par ICP-MS en fin d'expérience (**Figure III.2**).

Après avoir ajouté 3, 6, 12 et 24 ppm de Cd, on dose respectivement après 3 semaines d'incubation 2,64 ; 4,74 ; 10,7 et 19,9 ppm de Cd total (**Figure**



III.2). Il y a donc eu pendant les 3 semaines d'incubation des phénomènes de lessivage du Cd au sein des microcosmes puisque en moyenne 17% du Cd initial amendé a disparu. Les teneurs en Cd obtenues après traitement au nitrate de calcium sont de 1,19 ; 3,9 ; 8,2 et 18 ppm, respectivement pour les microcosmes amendés avec 3, 6, 12 et 24 ppm de Cd. Comparées aux teneurs en Cd total mesurées, ces valeurs nous suggèrent qu'une majeure partie du Cd total présent dans les microcosmes est mobilisable et surtout pour les fortes concentrations avec des teneurs de Cd "soluble" comprises entre 77 et 90% du Cd total mesuré dans les microcosmes amendés avec 6, 12 ou 24 ppm, alors que seul 45 % du Cd total des microcosmes témoin et amendé avec 3 ppm de Cd est mobilisable.

Les concentrations en Cd total et "soluble" rajoutées dans le milieu augmentent d'un facteur de 2 d'une condition à l'autre. Elles ne semblent donc pas trop élevées car il n'y a pas de saturation (**Figure III.2**). Néanmoins, il semble que la mobilité des métaux soit importante dans le sol. Cette mobilité est probablement due aux propriétés du sol. En effet, le sol du site Lommel est légèrement acide et lessivé, avec une texture sableuse et une faible teneur en matière organique. Toutes ces conditions favorisent la mobilisation des éléments métalliques et leur lessivage (**Tableau III.1**) (**Tack 2000**). Nous aurons donc potentiellement une grande part de ces métaux mobilisables pour les communautés d'organismes vivant au sein de cet écosystème. Il aurait été intéressant d'extraire aussi les métaux avec de l'eau et de doser également les métaux contenus dans les plantules de pin ce qui aurait indiqué la fraction de métaux biodisponibles. Ceci devra être fait ultérieurement.

Une même procédure a été utilisée pour réaliser la deuxième série de microcosmes avec des amendements en Cd, en Cu et en Zn. Le dosage des métaux totaux, solubles et absorbés par les plantules de pins, au sein de ces derniers microcosmes n'a pas encore été réalisé.



• **Significations biologiques des concentrations en métaux retrouvées dans les microcosmes.**

En Europe, les concentrations naturelles moyennes du Cu, Zn et Cd dans les sols considérés comme « sains » sont autour de 10 à 40 ppm de Cu (**Scientific Committee on Health and Environmental Risks 2009**), de 10 à 100 ppm de Zn (**Mertens & Smolders 2013**) et de 0,1 à 1 ppm de Cd (**Smolders & Mertens 2013**). Il n'existe pas encore de législation européenne sur les valeurs seuils de métaux des sols. En France, la teneur moyenne de Cu, Zn et Cd dans les sols est respectivement de 14,9 ; 149 et 0,42 ppm. Les valeurs seuils ne sont fixées que pour des cas particuliers comme par exemple pour les sites d'épandage de boues où ces valeurs sont de 100, 300 et 2 ppm de matière sèche respectivement pour le Cu, le Zn et le Cd (**G.Miquel 2000-2001**).

Il a été montré que le Cu a des effets sur la composante biologique des sols à partir d'une concentration totale comprise entre 20 - 200 ppm en fonction des types de sol considérés avec des effets sur certaines activités enzymatiques à partir 150 - 200 ppm (**Ruyters et al. 2013**). Le Cd a des effets sur les plantes, les invertébrés et les espèces microbiologiques à partir de 3 - 10 ppm selon les sols (**Smolders 2002**), et le Zn à partir de 100 ppm (**Mertens & Smolders 2013**), ce qui est au-delà des concentrations seuils pour le Cu et le Zn.

Etant donné que les taux de contamination des microcosmes amendés sont proches des seuils où des effets sur les organismes vivants sont observés, nous nous attendons à une modification de l'expression des gènes *emt* en fonction du taux de pollution apporté. Par ailleurs, nous supposons que ces taux de contamination sont sub-létaux pour les organismes qui portent les gènes *emt*, leurs permettant ainsi d'être actif dans les microcosmes.





Co-extraction ADN/ARN

Après avoir réalisé les microcosmes, nous avons adapté un protocole de co-extraction des ADN et ARN à partir de sol. La co-extraction d'acides nucléiques nous a permis d'obtenir à partir de tous les microcosmes des rendements de $1 \pm 0,5 \mu\text{g}$ et $1,4 \pm 0,3 \mu\text{g}$ respectivement d'ADN et d'ARN par gramme de sol sec nous permettant de les utiliser pour développer les techniques de PCR classique et de PCR quantitatives.

Mise au point de la PCR.

Avant mon arrivée au laboratoire (**Lehembre et al. 2013**), les séquences *emt* étaient amplifiées à partir d'extraits d'ADN de sol en utilisant une procédure de PCR nichée dont les caractéristiques étaient incompatibles avec les méthodes de PCR quantitative (**Thornton & basu 2010, Wang & Seed 2006**). En effet, cette PCR nichée utilise un jeu d'amorces dégénérées ciblant toutes les sous-familles d'*emt* avec des T_m trop bas (56°C contre 60°C minimum pour les PCR quantitatives), donne des amplicons d'une taille trop élevée (450 pb contre 300 pb attendues en PCR quantitative) et une amplification de 60 cycles si on cumule les cycles des deux réactions de la PCR nichée. Aussi, dans un premier temps, nous avons mis au point une méthode d'amplification par PCR classique des séquences *emt* à partir d'extraits d'ADN de sol en concevant de nouveaux jeux d'amorces spécifiques compatibles avec les PCR quantitatives et en amplifiant les séquences *emt* en moins de 45 cycles, conditions préalables nécessaires aux PCR quantitatives.



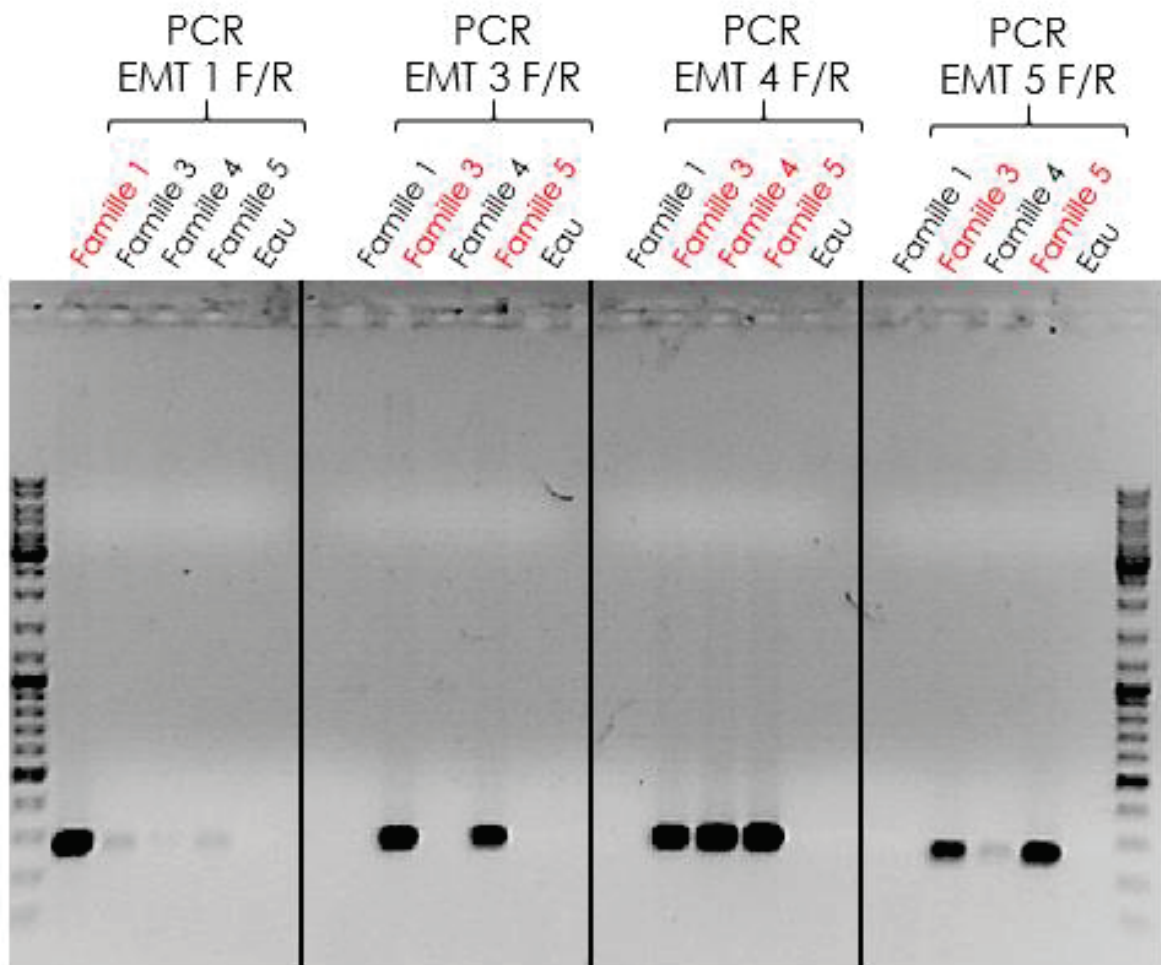


Figure III.3 : Spécificité des couples d'amorces conçus pour amplifier les gènes *emt* des sous-familles 1 (amorces EMT1F/R), 3 (EMT3F/R), 4 (EMT4F/R) et 5 (EMT5F/R). Chacun de ces couples a été utilisé en PCR avec des gènes des sous-familles 1, 3, 4 et 5 et les produits d'amplification déposés sur gel d'agarose à 1%. Seul le couple EMT1F/R s'avère spécifique de la sous-famille 1. En rouge sont représentées les séquences nettement amplifiées par les amorces.



• Développement d'amorces spécifiques

Les amorces EMT1F/R, EMT3F/R EMT4F/R, EMT5F/R (**Tableau III.2**) ont été conçues pour amplifier spécifiquement et respectivement les séquences des sous-familles 1, 3, 4 et 5. Leur spécificité vis-à-vis de chacune de leur famille respective a été testée par PCR, sur une matrice d'ADN plasmidique contenant une séquence de chaque sous-famille (**Figure III.3**). Seul le couple d'amorces EMT1F/R paraît spécifique de la sous-famille 1. En effet, une forte amplification ne s'observe qu'à partir du gène *emt1* (bien qu'un fragment soit très légèrement amplifié avec les plasmides contenant un gène *emt* des sous-familles 3, 4 et 5). Les autres couples d'amorces ne sont pas spécifiques de leur famille puisque les couples EMT3F/R et EMT5F/R amplifient les deux gènes *emt* des sous-familles 3 et 5, et le couple EMT4F/R amplifie 3 gènes *emt* des sous-familles 3, 4 et 5. Nous avons donc choisi pour la suite de ce travail d'utiliser la paire d'amorces EMT1F/R, qui n'amplifie que des séquences de la sous-famille 1.

Une sonde EMT1P TaqMan®, spécifique de la sous-famille 1 a été conçue et testée en qPCR en utilisant les amorces EMT1F/R sur une gamme de concentrations de plasmide contenant une séquence de la sous-famille 1. La dénaturation des produits PCR générés par ce couple d'amorces qPCR montre qu'elles amplifient un seul produit et que les amorces ne forment pas de dimères.



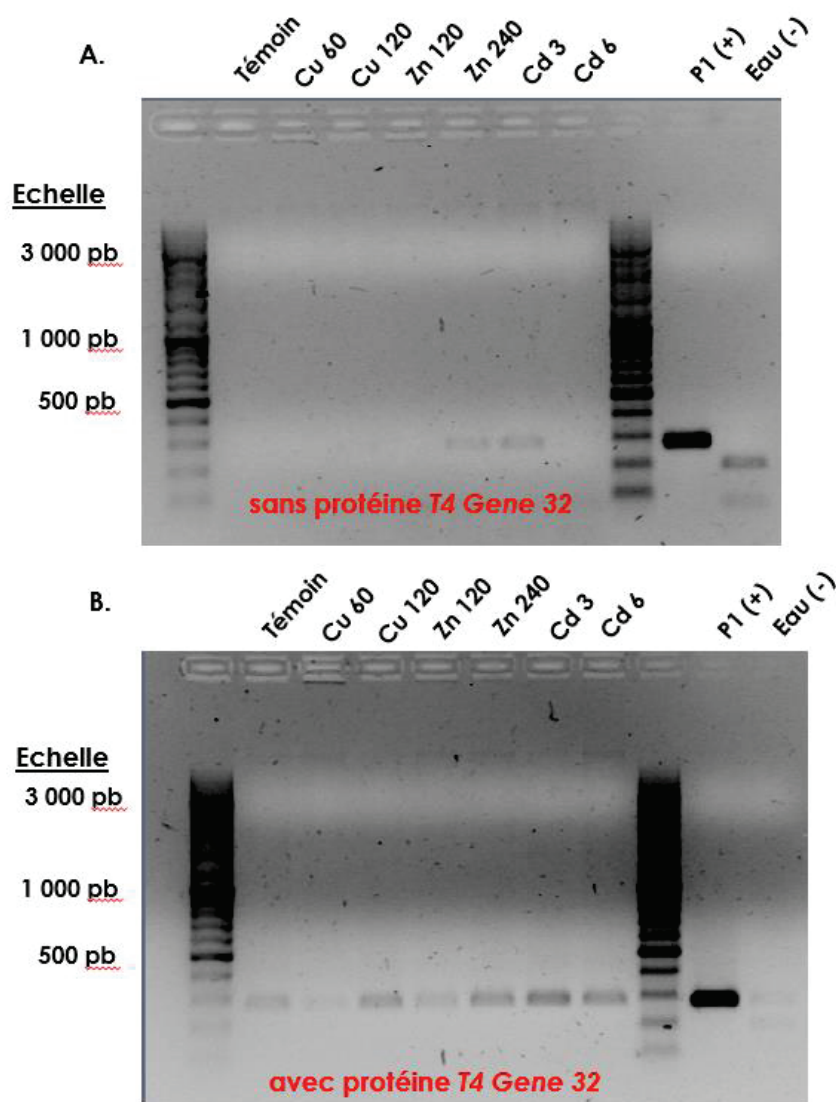


Figure III.4 : Effet positif de la protéine T4 Gene 32 sur la capacité à amplifier un fragment interne du gène *emt1* (amorces EMT1F/R) à partir de 160 ng d'ADN extraits de sols amendés ou non (témoin) par différents métaux.

P1 (+) : témoin positif d'amplification correspondant à l'amplification de 30 ng du plasmide pFL61 contenant une séquence *emt* de la sous-famille 1 (HE716999).

Eau (-) : témoin négatif de la PCR correspondant à l'amplification en absence d'ADN.



• Amplification par PCR classique des séquences *emt1* à partir d'extraits d'ADN de sol

Précédemment, les séquences *emt* ne pouvaient être amplifiées à partir d'extraits d'ADN de sol qu'en PCR nichée (avec 60 cycles au total). Afin d'amplifier spécifiquement les séquences *emt* de la sous-famille 1 (*emt1*), nous avons réalisé une PCR de 40 cycles à l'aide des amorces EMT1F/R en ajoutant ou non la protéine *T4 Gene 32* au milieu réactionnel de PCR.

Pour ce test, les amplifications ont été réalisées à partir d'extraits d'ADN de sol de toutes les conditions de la deuxième série de microcosmes (**Figure III.4**). Les résultats nous montrent que sans la protéine *T4 gene 32*, seules deux légères amplifications des séquences *emt1* sont observées à partir des extraits d'ADN de sol amendés par 3 ppm de Cd et 240 ppm de Zn (**Figure III.4 A**). En présence de la protéine *T4 gene 32*, tous les extraits d'ADN, quels que soient les microcosmes dont ils sont issus, donnent un fragment d'amplification après une PCR de 40 cycles (**Figure III.4 B**).

Il nous est donc possible de quantifier en qPCR les gènes *emt1* avec le jeu d'amorces EMT1F/R en ajoutant au milieu réactionnel la protéine *T4 gene 32*. Pour la ddPCR, nous avons supposé qu'il n'était pas nécessaire d'utiliser la protéine *T4 gene 32* puisqu'en théorie, la ddPCR permet d'amplifier des séquences très faiblement représentées dans des extraits d'ADN environnementaux (**Cavé et al. 2016**).

Pour vérifier la spécificité des amorces, une PCR en temps réel préliminaire a été réalisée à partir d'extraits d'ADN de microcosme témoin. Le clonage et le séquençage de 9 clones contenant les produits de PCR nous a permis de confirmer que les amorces amplifient des fragments de la taille attendue (276 paires de base) appartenant à la sous-famille 1.



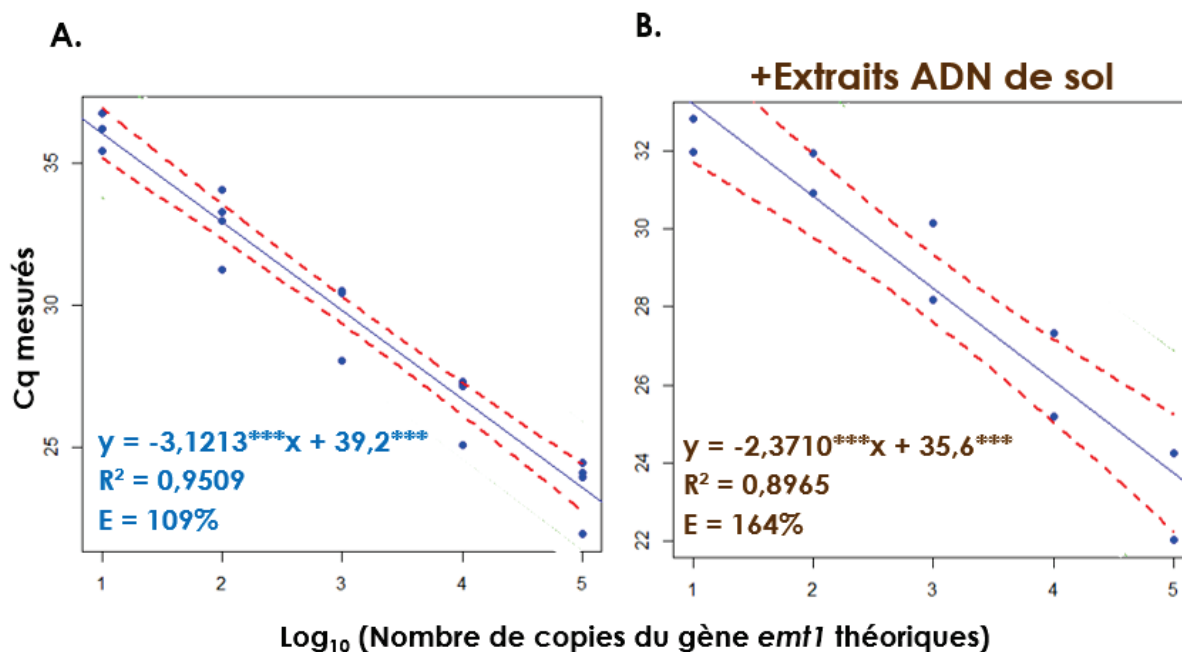


Figure III.5 : Amplification du gène *emt1* par qPCR **A.** en l'absence ou **B.** en présence d'ADN de sol dans le mélange réactionnel. Correspondance entre le nombre de copies du gène ajouté dans le mélange réactionnel et la valeur du Cq mesurée. Les points bleus correspondent aux moyennes des mesures par plaque, la droite bleue correspond au modèle de régression linéaire et l'intervalle de confiance de la valeur moyenne est représenté en pointillé rouge.

R^2 , coefficient de détermination

E, efficacité de l'amplification

Le signe *** correspond à une *p-value* de 0,001.



Amplifications par PCR quantitative : comparaison de la qPCR en temps réel et de la ddPCR

Dans le but de choisir la méthode la plus adaptée à notre modèle, nous avons choisi de mesurer le nombre de copies du gène *emt1* à partir d'une gamme de concentrations d'ADN plasmidique contenant ce gène en présence ou non d'1 ng d'extraits d'ADN métagénomique provenant des échantillons de sol de microcosme témoin. Les résultats de cette expérience permettront de savoir si les extraits de sol contiennent ou non des molécules interférant lors de réactions enzymatiques comme par exemple des acides humiques (**Dong et al. 2006**).

Nous avons tout d'abord quantifié le gène *emt1* par qPCR. Lorsque l'on représente les C_q mesurés en fonction du log₁₀ du nombre de copies théorique de séquences *emt1* (**Figure III.5**). On obtient un coefficient de détermination de 0,95 pour l'amplification du plasmide sans extraits d'ADN métagénomique et de seulement 0,90 pour l'amplification du plasmide en présence d'extraits métagénomiques ce qui est en deçà du seuil d'acceptabilité de 0,95. Par ailleurs, l'efficacité d'amplification est de 109% pour le plasmide seul contre 164% pour le plasmide en présence d'ADN de sol. Pour les deux conditions, l'efficacité d'amplification est supérieure à 100%, ce qui peut s'expliquer par le fait que la méthode de calcul de régression linéaire, utilisant une gamme de dilution standard, a tendance à surestimer l'efficacité d'amplification (**Pfaffl 2006**). Toutefois, l'efficacité est de 164% quand on ajoute des extraits d'ADN de sol ce qui est bien trop élevé par rapport à une erreur acceptable de $\pm 10\%$. Il semble probable que ce problème d'efficacité soit dû à la présence de composés extraits à partir du sol et perturbant la réaction enzymatique.



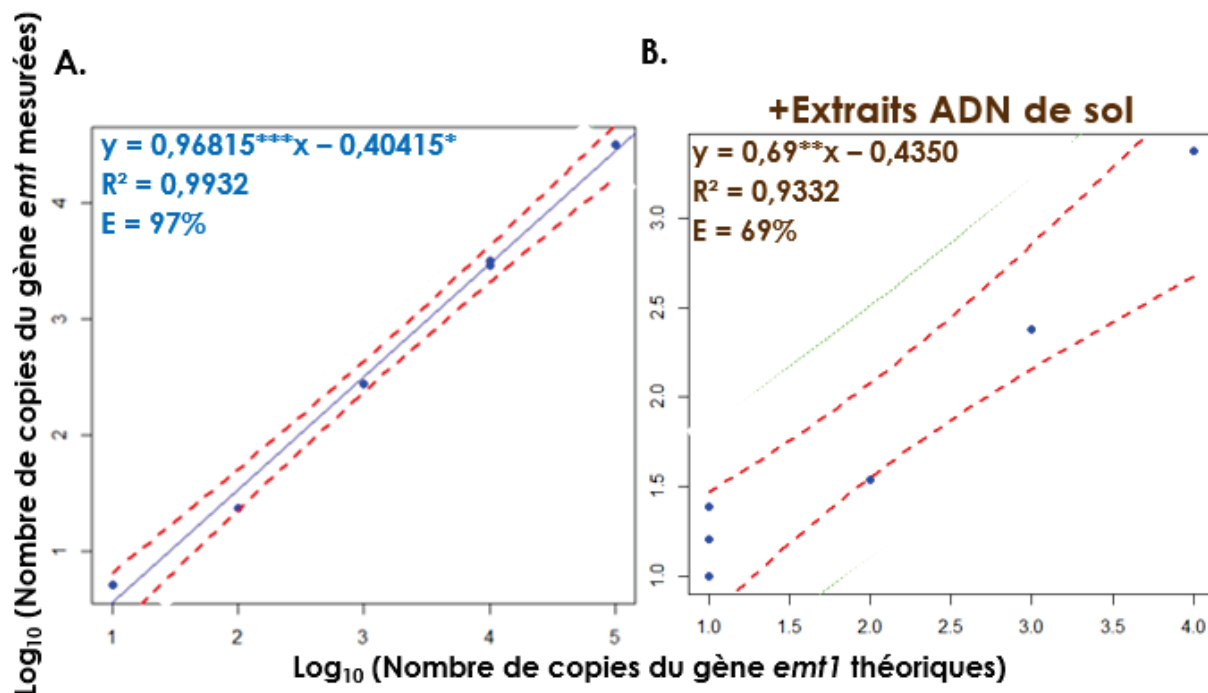


Figure III.6 : Amplification du gène *emt1* par ddPCR **A.** en l'absence ou **B.** en présence d'ADN de sol dans le mélange réactionnel. Correspondance entre le nombre de copies du gène ajouté dans le mélange réactionnel et le nombre de copies mesuré. Les points bleus correspondent aux moyennes des mesures par plaque, la droite bleue correspond au modèle de régression linéaire et l'intervalle de confiance de la valeur moyenne est représenté en pointillé rouge.

R^2 , coefficient de détermination

E, efficacité de l'amplification

Les signes *, **, *** correspondent respectivement à des p -value de 0,05 ; 0,01 ; 0,001.



Nous avons ensuite testé l'amplification du gène *emt1* par *droplet digital* PCR. En ddPCR, l'amplification du plasmide seul ou en présence d'extraits d'ADN de sol (**Figure III.6**) donne des coefficients de détermination R^2 respectivement de 0,99 et 0,93. Alors que le coefficient obtenu avec le plasmide seul est correct, celui calculé pour l'amplification du plasmide en présence d'extrait d'ADN de sol n'est pas bon puisqu'il est légèrement inférieur à 0,95. Les valeurs des coefficients de détermination reflètent les efficacités de la réaction d'amplification qui sont respectivement de 97% et 69% pour les conditions sans et avec des extraits d'ADN de sol. Il semble donc que, comme pour la qPCR, la présence d'extraits de sol perturbe la quantification par ddPCR.

Ainsi, pour la qPCR et la ddPCR, il semble que des composés extraits à partir de sol perturbent la réaction d'amplification. Toutefois, les données portant sur le plasmide seul suggèrent que la ddPCR a une meilleure efficacité lors de l'amplification des gènes *emt1* en utilisant les oligonucléotides EMT1F/R/P comparé à la qPCR. Nous pourrions donc essayer d'améliorer le protocole d'extraction des acides nucléiques afin de réduire plus significativement la présence de ces molécules en ajoutant des étapes de purification en amont et en aval de la co-extraction. Nous pouvons par exemple précipiter les acides humiques des sols avec un traitement aux sels d'aluminium avant de débiter l'extraction (**Peršoh et al. 2008**) et/ou purifier les extraits d'acides nucléiques par une nouvelle filtration sur les colonnes à membrane de silice du kit, une incubation avec des charbons actifs (**Luis et al. 2005**) ou une électrophorèse non linéaire sur gel (**Engel et al. 2012**). La quantité et la qualité des acides nucléiques obtenues seront estimées par électrophorèse en capillaire (Agilent). Nous pourrions aussi, dans le cas de la ddPCR, optimiser le mélange réactionnel d'amplification en rajoutant des « facilitateurs » d'amplification telle que la protéine *T4 Gene 32*.



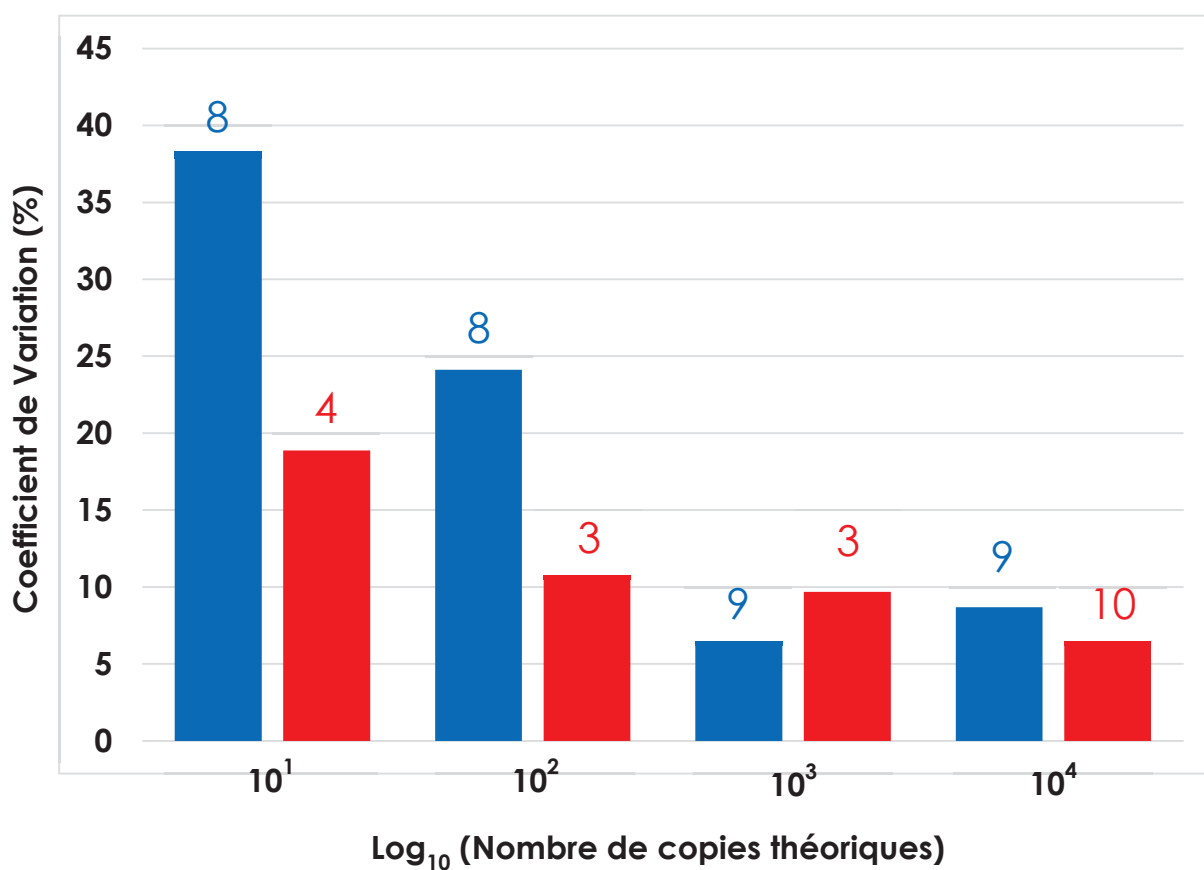


Figure III.7 : Coefficients de variation (%) des mesures de qPCR (en bleu) et de ddPCR (en rouge) (réalisées en présence de plasmide seul), en fonction du nombre de copies théoriques ($\log_{10}$). Le nombre de répliques est indiqué en haut de chacune des barres.



Les coefficients de variation des mesures en qPCR et en ddPCR nous donnent des informations sur la variabilité des mesures autour de la moyenne (**Figure III.7**). Ils ont été calculés à partir de la moyenne de toutes les mesures en présence de plasmide seul et sont présentés **figure III.7**. Pour l'amplification par qPCR (barres bleues de la **figure III.7**), la variabilité décroît à mesure que le nombre de copies augmente passant de 38% pour 10 copies à 9% pour 10^4 copies. Ces résultats nous indiquent qu'en qPCR la variabilité des mesures est dépendante de la quantité initiale d'ADN utilisée. Pour l'amplification par ddPCR (barres rouges de la **figure III.7**), la variabilité est relativement uniforme (entre 6 et 19%). Ces résultats nous montrent qu'en ddPCR la variabilité de la mesure est relativement indépendante de la quantité initiale d'ADN utilisée et de plus, cette variabilité est minimisée quel que soit le nombre de copies considéré (même aux faibles concentrations), ce qui est en accord avec les études réalisées sur des extraits d'ADN de sol (**Cavé et al. 2016**).

Pour finir, un essai préliminaire d'une amplification par RT-ddPCR a également été réalisé et a montré la présence d'environ 220 et 290 copies d'ADNc par gramme de sol sec respectivement dans les extraits de microcosme témoin et contenant 4,74 ppm de Cd (amendé avec 6 ppm) ce qui nous indique que les séquences *emt1* sont exprimées dans les conditions expérimentales et qu'elles sont en quantité suffisante pour pouvoir être quantifiées par les approches développées.



Conclusions et perspectives.

L'objectif de ce chapitre était de mettre au point une méthode de quantification des gènes *emt* à partir d'extraits de sol. Pour cela nous avons : 1) mis en place une expérimentation en microcosmes et 2) étudié la faisabilité de la quantification des gènes *emt* à partir d'extraits d'acides nucléiques de sol.

La première série de microcosmes réalisée à partir d'une gamme de concentrations en Cd a été dosée par ICP-MS ce qui a permis d'observer que les microcosmes ont fixé du Cd en concentrations croissantes permettant de les utiliser pour comparer l'induction des gènes *emt*. La deuxième série de microcosmes réalisée avec du Cd, du Cu et du Zn reste à être dosée.

La mise au point de la PCR quantitative a nécessité la mise en place d'une co-extraction ADN/ARN et la conception de jeux d'amorces spécifiques de certaines sous-familles d'EMT permettant l'amplification d'au moins un de ces gènes par PCR quantitative. Elle a débouché sur l'utilisation du couple d'amorces EMT1F/R spécifique de la sous-famille 1, permettant l'amplification d'un amplicon de 276 pb en association avec la sonde TaqMan® EMT1P. Les paramètres d'amplification retenus sont une température d'hybridation de 62,6°C pour la qPCR, en présence de protéine *T4 Gene 32* au milieu réactionnel et une température d'hybridation de 63,8°C pour la ddPCR. Même en présence de cette topo-isomérase, l'efficacité d'amplification par qPCR n'est pas optimale lorsqu'on rajoute des extraits d'ADN de sol du microcosme témoin dans le milieu réactionnel ce qui met en cause la procédure d'extraction utilisée. D'autres optimisations concernant l'élimination de substances inhibitrices présentes dans les extraits d'ADN de sol sont à prévoir par la suite. La comparaison des efficacités d'amplification des gènes *emt1* en utilisant le système d'amorces EMT1F/R/P et les matrices d'ADN mentionnées ci-dessus suggère que la ddPCR permet de quantifier les séquences *emt1* avec une meilleure précision comparée à la qPCR et en particulier lorsque les



quantités initiales d'ADN ajoutées au milieu réactionnel sont faibles. Toutefois, l'efficacité de la ddPCR n'était pas optimale lorsque des extraits d'ADN de sol sont rajoutés au milieu réactionnel. Celle-ci pourrait être améliorée en rajoutant la protéine *T4 Gene 32* dans le milieu réactionnel et en modifiant la procédure de co-extraction ADN/ARN.

Pour la suite de ce projet, nous devons aussi : 1) mettre au point la rétro-transcription des ARN en ADNc notamment en choisissant une stratégie d'amorçage optimale, 2) choisir la méthode de normalisation des mesures de ddPCR et 3) comparer le nombre de transcrits obtenus en fonction des différents amendements en métaux. La mise au point de l'amorçage sera réalisée en vue d'optimiser l'efficacité de réaction de transcription *in vitro* et de maximiser la diversité des séquences récupérées. Elle sera faite en comparant le rendement de rétro-transcription sur différentes concentrations d'ARN associées à 3 types d'amorçage différents : i) des penta-décamères aléatoires, ii) des amorces spécifiques *emt1*, iii) un mélange d'oligo-d(T) et de penta-décamères aléatoires (**Stangegaard et al. 2006, Sugden & Winter 2008**). De plus, un essai préliminaire d'une amplification des gènes *emt1* par RT-ddPCR a été réalisé en utilisant des *random hexamers* et a montré que plusieurs centaines de copies sont présentes dans les microcosmes témoins et amendés avec 6 ppm de Cd ce qui semble très prometteur pour la suite de cette expérience.

Par ailleurs, étant donné que nous ne connaissons pas les organismes qui portent les gènes *emt1*, nous nous orientons vers une normalisation en utilisant le rapport $\frac{ADNc}{ADN}$ du gène *emt1*. Le prérequis pour cela est de vérifier que la procédure d'extraction permet d'obtenir les ADN et les ARN de manière linéaire (**Bealum et al. 2008, Nicolaisen et al. 2008, Almario et al. 2013,**). Cette vérification consiste à ajouter, à des sols témoins, une dilution en série de cellules de levure exprimant un gène *emt1* cloné dans le vecteur d'expression pFL61. Le nombre de séquences d'ADN et d'ADNc *emt1* sera mesuré par



ddPCR. Si les conditions sont remplies, nous pourrons alors comparer l'expression du gène *emt1* entre chacun des différents microcosmes amendés.





CHAPITRE 4

Recherche de gènes associés aux séquences *emt* par une approche de capture ciblée

Introduction

Les métallothionéines environnementales (EMT), telles que définies dans le chapitre 2, regroupent des protéines chélatrices de métaux appartenant à la superfamille des MT (**Ziller *et al.* 2017**). Contrairement aux autres MT décrites, les EMT ont une origine taxonomique inconnue car elles ont été isolées d'ARN environnementaux par une approche de métatranscriptomique fonctionnelle (**Lehembre *et al.* 2013, Yadav *et al.* 2016**) et ne présentent pas d'homologues dans les bases de données qui puissent les rattacher à un groupe taxonomique. Bien que les EMT puissent se lier au Cu, Zn et Cd, nous ne savons pas si leur transcription peut être induite par la présence d'un de ces trois métaux (**cf chapitre 3**). L'obtention de l'environnement génomique des séquences *emt* est ainsi un enjeu majeur dans leur étude car cela permettrait : i) d'analyser leurs régions promotrices proximales pour avoir accès à des informations concernant leur régulation transcriptionnelle ; et, ii) d'affilier éventuellement ces séquences à un taxon si des biomarqueurs phylogénétiques sont obtenus. En effet, si les régions en amont et en aval des séquences *emt* sont homologues à des séquences présentes dans les bases de données, des analyses phylogénétiques incorporant ces séquences flanquantes pourraient permettre d'inférer l'origine taxonomique des *emt*.



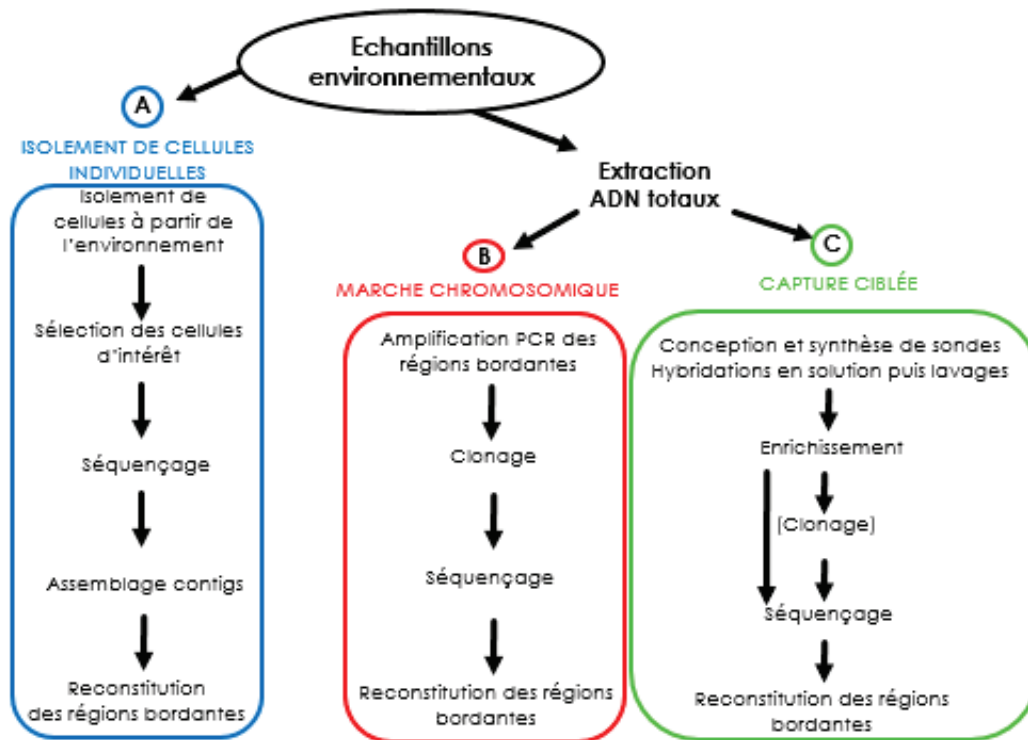


Figure IV.1 : Stratégies pour obtenir des régions bordantes à partir d'une séquence nucléotidique connue. Trois grandes stratégies sont possibles : l'isolement de cellules individuelles, les techniques de marche chromosomique et la capture ciblée de gènes.

L'isolement de cellules individuelles **(A)** consiste à isoler des cellules du sol et à sélectionner les cellules portant les séquences d'intérêt après marquages (marqueurs fluorescents, magnétiques), séquencer leur génome et reconstituer les régions bordantes *in silico*. Les deux autres techniques sont basées dans un premier temps sur l'extraction de l'ADN total du milieu étudié. Ensuite on amplifie spécifiquement les régions bordantes par des techniques de marche chromosomique puis on les identifie par clonage/séquençage **(B)**, soit on enrichit sélectivement l'échantillon de régions génomiques contenant les séquences d'intérêt par capture ciblée puis on les identifie par clonage/séquençage **(C)**.



Trois grands groupes de stratégies sont disponibles pour identifier les régions génomiques bordant une séquence connue (**Figure IV.1**) : les techniques de marche chromosomique, l'isolement de cellules individuelles couplé à un séquençage génomique, et la capture ciblée de gènes.

Les techniques de marche chromosomique (**Figure IV.1 B**) permettent d'obtenir les régions adjacentes de séquences connues par l'utilisation d'une combinaison de techniques classiques de biologie moléculaire directement sur des extraits d'ADN génomiques ou environnementaux : PCR, digestions enzymatiques, clonage et séquençage (**Kotik 2009, Leoni et al. 2011**). Ces techniques peuvent être séparées en deux grands ensembles : 1) celles nécessitant la prédigestion enzymatique des extraits d'ADN, et 2) celles ne nécessitant pas de digestion. Le premier ensemble regroupe des techniques où l'ADN matrice est fragmenté principalement par digestion enzymatique puis soit i) ligaturé sur lui-même afin de circulariser les fragments (pour la réalisation de PCR inverses), soit ii) ligaturé à des adaptateurs (PCR cassettes), iii) soit ligaturé directement à un vecteur de clonage. Dans un second temps, des PCR optimisées pour les longs fragments utilisant des jeux d'amorces spécifiques aux constructions obtenues, après ligature, permettent d'obtenir les régions bordantes. Un second ensemble de méthodes de marche chromosomique, ne nécessitent pas la prédigestion des extraits d'ADN environnementaux mais repose sur la répétition de PCR associées à des jeux d'amorces dégénérées et spécifiques des séquences cibles comme par exemple la TAIL-PCR (*Thermal asymmetric interlaced-PCR*) (**Liu & Whittier 1995, Leoni et al. 2011**). Les limites des méthodes de marche chromosomique sont principalement liées au fait que ces techniques combinent plusieurs réactions enzymatiques successives dont l'efficacité de réaction va être conditionnée par la quantité et la nature des molécules inhibitrices présentes dans le milieu réactionnel. Ces molécules sont particulièrement nombreuses dans le cas d'extrait d'ADN de sol (acides humiques, chélateurs inorganiques,...). De plus,



toutes ces méthodes sont basées sur des amplifications d'ADN par PCR ne permettant d'obtenir que des fragments d'ADN de quelques milliers de pb.

Une autre approche consiste à isoler dans un premier temps les organismes (potentiellement unicellulaires) portant les séquences cibles à partir d'échantillons environnementaux puis à réaliser un séquençage de leurs génomes (**Brehm-Stecher & Johnson 2004, Pivetal et al. 2015, Gasc et al. 2015**) (**Figure IV.1 A**). Différentes méthodes peuvent être utilisées pour isoler des cellules : dilution, micro-manipulation, cytométrie de flux, microfluidique, compartimentation et des combinaisons de plusieurs de ces techniques (**Ishii et al. 2010**). Beaucoup d'entre elles passent par le marquage de gènes spécifiques des organismes étudiés, gènes qui sont utilisés comme sondes pour marquer (et ainsi distinguer et trier) les organismes qui les portent. En fonction du type de marquage utilisé, il est possible d'obtenir en plus des séquences génomiques, des informations quant à l'organisation et la morphologie cellulaire de l'organisme hôte dont la cellule entière est isolée. Le marquage des cellules peut être réalisé par exemple en utilisant des sondes fluorescentes complémentaires des ARNm codés par le gène cible ou des *loci* recherchés (**Brauns & Goos 2005, Ozawa et al. 2016**). Une autre possibilité consiste à utiliser une sonde magnétique couplée à des puces aimantées permettant de récupérer les cellules marquées (**Pivetal et al. 2015**). Les limitations de ce groupe de techniques sont : 1) le protocole d'extraction de cellules de la matrice environnementale qui doit être optimisé ; 2) la présence de molécules naturellement fluorescentes dans les pools cellulaires qui rendent leur utilisation délicate dans le cas de cellules provenant d'une matrice complexe telle que du sol ou des sédiments ; et, 3) l'effectif des cellules cibles dans les échantillons environnementaux. Aussi ces techniques sont surtout utilisées sur des échantillons aquatiques.

La capture de gènes est une dernière technique récente permettant d'enrichir sélectivement des extraits d'ADN environnementaux en séquences d'intérêt (**Figure IV.1 C**) (**Gnirke et al. 2009**). Elle repose sur le même principe



que la chromatographie d'affinité en phase liquide utilisée dans la purification de protéines : 1) des sondes ARN biotinylées complémentaires des séquences d'intérêt sont ajoutées à un extrait d'ADN environnemental fragmenté et s'hybrident aux séquences cibles ; 2) les hybrides entre sondes biotinylées et séquences environnementales cibles sont spécifiquement capturées à la surface de billes magnétiques recouvertes de streptavidine, protéine présentant une très forte affinité pour la biotine ; et, 3) les billes magnétiques recouvertes des séquences cibles sont récupérées grâce à un aimant. Enfin, après « décrochage » des séquences de la streptavidine recouvrant les billes, celles-ci peuvent être directement séquencées par une approche haut-débit, moyen-débit ou les deux (**Gasc et al. 2015**). L'utilisation de la capture de gènes, couplée au séquençage massif, a permis de reconstruire *in silico* des fragments de 301 à 1.639 pb, dont une région promotrice d'environ 200 pb, d'un gène d'une archée peu abondante d'un lac français (**Denonfoux et al. 2013**).

Les méthodes d'isolement de cellules individuelles nécessitent à la fois des moyens techniques (e.g. systèmes de tri cellulaire) et une expertise qui n'étaient pas disponibles dans mon laboratoire, nous nous sommes donc tournés vers des méthodes de marche chromosomique et de capture de gènes. Préalablement à mon travail, des essais d'identification des *loci* adjacents des gènes *emt* avaient été réalisés au laboratoire. Ces essais de marche chromosomique par TAIL-PCR et PCR après ligature d'adaptateurs se sont avérés infructueux. La très forte dilution des séquences *emt* au sein des extraits d'ADN environnementaux, associés à la présence d'inhibiteurs, ont été envisagées comme causes de cet échec.





Aussi, nous avons choisi de développer, en collaboration avec le groupe du professeur Pierre Peyret de l'Université d'Auvergne, une approche de capture de gènes par séquençage bas-débit après présélection des séquences d'intérêt. L'équipe de Pierre Peyret est à l'origine de la première utilisation de cette méthode sur des extraits d'ADN environnementaux (**Denonfoux et al. 2013**). Après capture, nous avons tenté de cloner les fragments d'ADN obtenus, de cribler ceux-ci par PCR en utilisant des amorces ciblant les gènes *emt* et de séquencer les clones positifs par la méthode Sanger.

L'objectif de cette étude est donc de mettre au point une procédure hybride entre la marche chromosomique (clonage, séquençage bas-débit) et la capture de gènes (hybridation en solution) pour obtenir les séquences d'ADN qui bordent les régions codantes des gènes *emt* afin de permettre à la fois la détermination du groupe taxonomique auquel appartiennent ces gènes, et d'avoir des indications sur les régions génomiques régulatrices des gènes *emt*. Pour atteindre cet objectif, deux captures de gènes successives ont été réalisées, avec chacune un jeu différent de sondes.



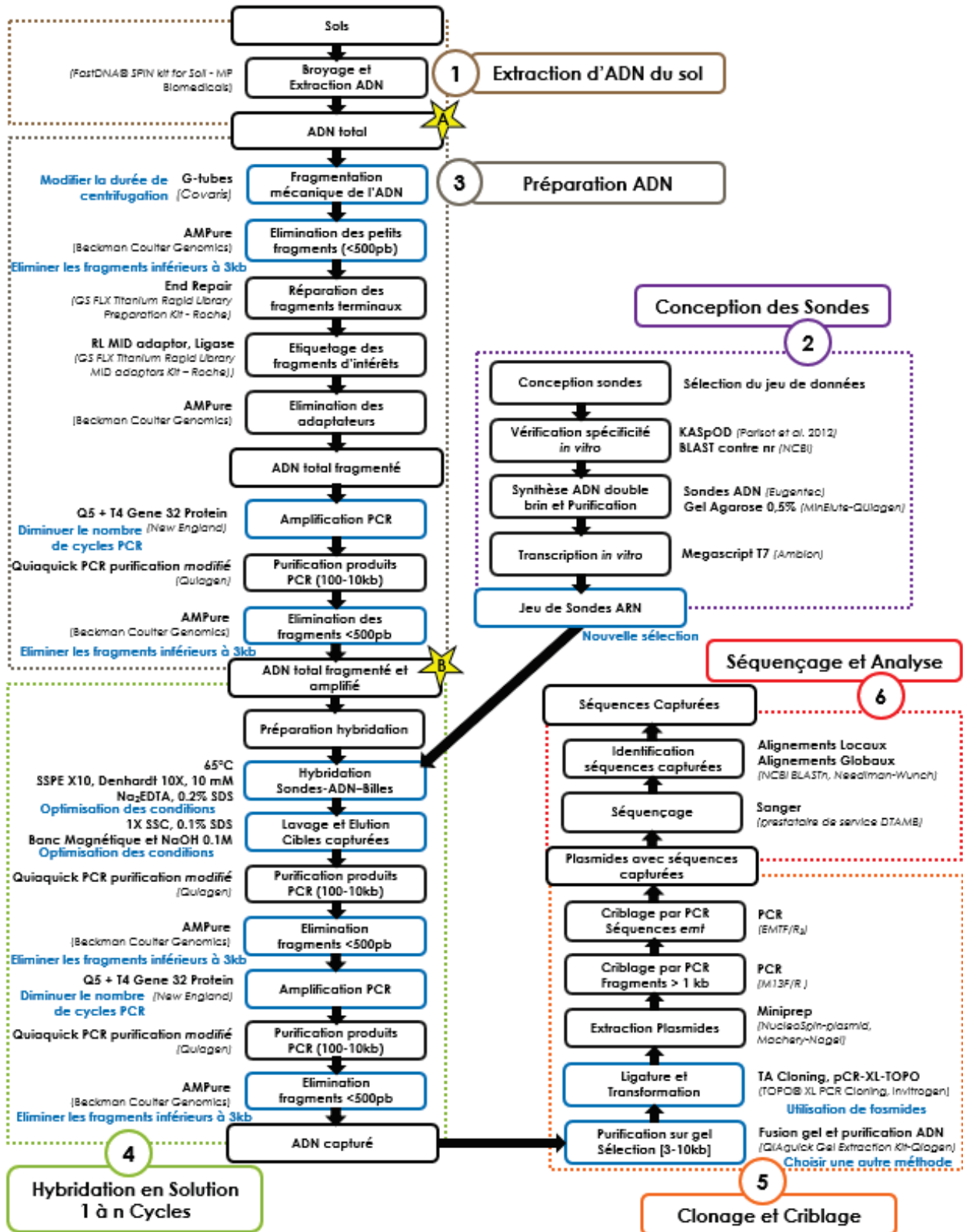


Figure IV.2 : Protocole expérimental de la capture de gènes. La procédure écrite en noir est composée de six grands ensembles d'étapes. Encadrées en bleu sont indiquées les étapes pour lesquelles nous pouvons apporter des modifications en vue d'améliorer la technique. Les étoiles jaunes correspondent aux vérifications de la taille des ADN par électrophorèse en capillaire.

Matériels et méthodes

L'ensemble du protocole expérimental de la capture de gènes est schématisé dans la **figure IV.2**.

Site d'étude et extraction d'ADN à partir d'échantillons de sol

L'ADN métagénomique a été extrait à partir d'échantillons de sol prélevés sur le site de Lommel situé en lisière de forêt de Pin au nord-est de la Belgique (51° 14' N, 5° 15' E). Ce site anciennement contaminé par des métaux était à proximité d'une ancienne raffinerie de Zinc et a été reboisé en 1975. C'est à partir d'ARNm extraits de sol de ce site qu'ont été isolées la plupart des séquences *emt* (**Lehembre et al. 2013**).

• Caractéristiques des échantillons de sol

La parcelle échantillonnée est en bordure d'une forêt de *Pinus sylvestris*. Le sol est sableux, légèrement acide et oligotrophe. L'échantillonnage a été réalisé en 2006. Vingt carottes cylindriques de 20 cm de longueur et de 5 cm de diamètre ont été collectées le long de deux transects de 10 m chacun. Ces carottes ont été rassemblées, tamisées à 2 mm puis stockées à -80°C après une congélation rapide dans l'azote liquide.

• Extraction de l'ADN total des échantillons de sol

Quarante grammes de sol ont tout d'abord été broyés pendant 3 min dans un broyeur de roche en agate préalablement placé à -80°C (**Bailly et al. 2007**). L'ADN total a été extrait à partir de 12 g de sol broyé en utilisant le kit « *FastDNA® SPIN kit for Soil* » (MP Biomedicals) dont le protocole a été modifié afin de limiter la fragmentation des ADN (**Figure IV.2, étape 1**). Pour cela, l'étape de « *bead beating* » a été réalisée avec des porte-tubes en céramique préalablement placés à 4°C ; les différentes étapes de centrifugation ont été



réalisées à 4°C et l'ensemble des pipetages s'est effectué avec des cônes stériles dont les extrémités ont été coupées pour réduire la fragmentation de l'ADN. La qualité des ADN extraits (taille moyenne des fragments obtenus) a été vérifiée par électrophorèse sur gel d'agarose à 1% coloré au bromure d'éthidium. La quantité d'ADN a été estimée par spectrophotométrie ($DO_{230:260:280}$; NanoDrop 3300®, Thermo Scientific) et la taille moyenne des fragments d'ADN a été vérifiée par électrophorèse en capillaire (Agilent DNA 7500 and DNA 12 000 Kit – Agilent). L'ADN métagénomique ainsi extrait a été stocké à -20°C.

Conception et synthèse des sondes de capture

• Première expérience (capture A), conception de sondes oligonucléotidiques courtes

L'étape de conception a été réalisée par le groupe de Pierre Peyret à Clermont-Ferrand et la synthèse des sondes biotinylées par Laurence Fraissinet-Tachet. Quarante sondes dont 23 dégénérées de 50 pb ont été conçues en utilisant le logiciel KASpOD (**Parisot et al. 2012**) à partir de l'alignement multiple de 45 séquences ADNc d'EMT, représentant toute la diversité des séquences EMT connues (**Figure IV.2, étape 2**). Ce lot est composé d'un ensemble de sondes couvrant les extrémités 5' et 3' des régions codantes et ciblant les sous-familles d'EMT 1, 2, 3, 4, 5 et 7 (**Tableau 1 de l'annexe IV.1**). Lors de leur synthèse sous forme d'ADN simple brin, à chacune de ces sondes de 50 nucléotides (S_{50}) spécifiques des EMT a été ajoutée à l'extrémité 5' la séquence promotrice reconnue par l'ARN polymérase T7 associée à l'adaptateur A, et à l'extrémité 3' l'adaptateur B (T7-A- S_{50} -B) permettant leur amplification par PCR (**Tableau 3 de l'annexe IV.1**). La spécificité des sondes vis-à-vis des séquences EMT a été testée *in silico* par BLASTn contre la base de données INSD (« International Nucleotide Sequence Databases ») sur le serveur NCBI.



• Deuxième expérience (capture B), conception de sondes longues

Pour cette capture, nous avons choisi de synthétiser 16 sondes ARN (**Tableau 2 de l'annexe IV.1**) de grande taille (entre 203 et 349 pb) couvrant toute la région codante de 16 gènes *emt* appartenant aux sous-familles EMT 1, 3, 4 et 5.

• Synthèse des sondes

Les sondes courtes ont été synthétisées (Eurogentec) sous forme d'ADN simple brin (**Figure IV.2, étape 2**). Elles ont ensuite été synthétisées sous forme d'ADN double brin en les amplifiant par PCR avec les amorces T7-A et B (**Tableau 1 (PCR 1) de l'annexe IV.2**). La matrice d'ADN utilisée pour la synthèse des sondes longues a été obtenue par PCR (**Tableau 2(PCR 2) de l'annexe IV.2**) en utilisant comme matrice des plasmides contenant des séquences *emt* des familles précédemment mentionnées. Les amorces «sens» utilisées pour l'amplification contenaient en 5' la séquence du promoteur T7.

Pour les deux captures, après amplification des sondes sous forme d'ADN double brin, les produits PCR ont été purifiés sur gel d'agarose *low melting point* à 4% (*Qiaquick gel extraction kit, MinElute-Qiagen*), puis transcrits *in vitro* en ARN en présence d'UTP biotinylé et d'une ARN polymérase T7 haute-fidélité (*Megascript T7, Ambion*). Les ADN ont été éliminés en utilisant de la *TurboDNase*® fournie dans le kit *Megascript T7* et les ARN biotinylés ont été stockés à -80°C après une précipitation à l'éthanol. Pour les utiliser, les ARN ont été repris dans 100 µL de tampon TE.



Préparation des extraits d'ADN de sol

• Fragmentation de l'extrait d'ADN de sol

Quatre microgrammes d'ADN métagénomique ont été fragmentés mécaniquement pour obtenir une taille moyenne théorique des fragments de 10 kb (**Figure IV.2, étape 3**). Pour cela, 4 µg d'ADN ont été déposés dans le réservoir supérieur de G-tubes (Covaris) qui ont été centrifugés à 8300 g dans une microcentrifugeuse Eppendorf® 5424R pendant 1 min, puis re-centrifugés 30 sec de sorte que tout l'échantillon se retrouve dans le réservoir inférieur après être passé au travers d'un capillaire permettant la fragmentation régulière de l'ADN. En théorie, une telle centrifugation devrait produire des fragments d'ADN de 10 kb.

• Réparation des extrémités de l'ADN fragmenté et pré-amplification

Les extrémités des fragments d'ADN ont été rendues franches et phosphorylées (GS FLX Titanium Rapid Library Preparation Kit - Roche) en les incubant en présence de *T4 DNA polymerase* (activité 5'-3' polymérase), de *Taq DNA polymerase* (activité 3'-5' exonucléase), et de *T4 Polynucleotide kinase* (3' phosphorylase), d'ATP et des 4 dNTP (**Figure IV.2, étape 3**). Par la suite, des adaptateurs MID RL 006 (GS FLX Titanium Rapid Library MID adaptors Kit – Roche) contenant des séquences TiA et TiB ont été ligaturés aux extrémités et les fragments d'ADN de petite taille (<500 pb) ont été éliminés à l'aide des billes magnétiques AMPure (Beckman Coulter Genomics). Les fragments d'ADN ainsi produits ont été amplifiés à l'aide des amorces TiA et TiB spécifiques des adaptateurs MID en utilisant un temps d'élongation de 6 min permettant d'amplifier de longs fragments (**Tableau 2 (PCR 3) de l'annexe IV.2**). Une fois le produit PCR purifié (Quiaquick PCR purification, Qiagen et AMPure, Beckman Coulter Genomics), l'extrait d'ADN environnementaux fragmenté obtenu a été quantifié par spectrophotométrie (DO_{230:260:280} ; NanoDrop 3300®, Thermo Scientific) puis la taille moyenne des fragments



d'ADN a été vérifiée par électrophorèse en capillaire (Agilent DNA 7500 and DNA 12 000 Kit – Agilent).

Hybridation en solution

Pour chaque capture (captures A et B), le même extrait d'ADN environnemental a été capturé deux fois consécutivement pour augmenter l'enrichissement (capture 1 et capture 2) (**Figure IV.2, étape 4**). Les expériences de capture A et B diffèrent l'une de l'autre seulement par le mélange de sondes utilisé qui est composé de 40 sondes dégénérées de 50 pb pour la capture A et de 16 sondes non dégénérées de 203 à 349 pb pour la capture B. Ces captures ont été réalisées par Laurent Vallon à Clermont-Ferrand pour la capture A et à Lyon pour la capture B.

• Etape 1 de la capture 1 : hybridation

2,5 µg d'ADN de sperme de saumon (Ambion) ont été mélangés à 2 µg d'extrait d'ADN environnementaux fragmentés et le tout a été évaporé à sec sous vide (au speedvac). Les ADN ont été repris dans 7 µL d'eau et dénaturés pendant 5 min à 95°C. Un tampon d'hybridation (SSPE X10, solution de Denhardt 10X, 10 mM Na₂EDTA et 0.2% SDS) (**Annexe IV.3**) préchauffé à 65°C a été ajouté puis le mélange a été incubé pendant 2 min à 65 °C. Six microlitres (=500 ng) de sondes ARN biotinylées ont alors été ajoutés après un préchauffage à 65°C pendant 2 min puis le mélange (ADN, tampon d'hybridation et sondes ARN) a été incubé à 65°C pendant 24h pour permettre l'hybridation des sondes aux séquences cibles.

• Etape 2 de la capture 1 : élution et ré-amplification

Deux microlitres de billes magnétiques recouvertes de streptavidine (Dynabeads® M-280 Streptatvidin – Invitrogen) ont été équilibrés dans un tampon (1M NaCl, 10 mM Tris-HCl pH 7,5, 1mM EDTA) par 3 lavages successifs, puis ajoutés à la solution d'hybridation. Le tout est incubé 30 min à température ambiante pour permettre à la streptavidine de se lier à la biotine présente sur



les sondes d'ARN. Les billes sont alors attirées sur un banc magnétique pendant 30 min puis une fois le surnageant enlevé, lavées successivement une fois par 500 µL de tampon de rinçage peu stringent (1X SSC/0.1% SDS) pendant 10 min à température ambiante et 3 fois par 500 µL de tampon de rinçage plus stringent (0.1X SSC/0.1% SDS) pendant 10 min à 65°C. Les ADN hybridés sont alors remis en solution en plaçant les tubes sur un banc magnétique et en ajoutant 50 µL de NaOH 0,1M. Après 10 min à température ambiante, la solution est neutralisée par ajout de 70 µL de Tris-HCl 1 M, pH 7,5. Le surnageant contenant les ADN métagénomiques capturés est alors purifié (Qiaquick PCR purification, Qiagen et AMPure, Beckman Coulter Genomics) puis réamplifié (**Tableaux 1 et 2 (PCR 3) de l'annexe IV.2**) en utilisant 25 cycles de PCR.

- **Capture 2 et amplification finale.**

Le second cycle de capture (=capture 2), utilisant les ADN métagénomiques obtenus à la suite de la capture 1, a été réalisé selon la même procédure que la capture 1. En fin de capture 2 une amplification finale a été réalisée (**Tableaux 1 et 2 (PCR 3) de l'annexe IV.2**). Le produit PCR purifié (Qiaquick PCR purification kit - Qiagen) a été quantifié par spectrophotométrie (DO_{230:260:280} ; NanoDrop 3300®, Thermo Scientific) et la taille des fragments d'ADN obtenus a été appréciée par électrophorèse en capillaire (Agilent DNA 7500 and DNA 12 000 Kit – Agilent).

Clonage et séquençage des produits de capture.

- **Vérification de l'enrichissement.**

L'enrichissement en séquences d'intérêt a été vérifié par PCR. La première PCR est une PCR nichée utilisant 2 couples d'amorces dégénérées EMTF/R puis EMTF/R₃ (**Tableaux 1 et 2 (PCR 4) de l'annexe IV.2**) et la deuxième est une PCR standard utilisant les amorces spécifiques de la famille 1 EMT1F/R (**Tableaux 1 et 2 (PCR 5) de l'annexe IV.2**).



• Clonage et criblage des produits de capture

Les fragments capturés compris entre 3 et 10kbp ont été découpés d'un gel d'agarose 1%, purifiés (QIAquick Gel Extraction Kit – QIAGEN), puis ligaturés au plasmide pCR-XL-TOPO (TOPO® XL PCR Cloning, Invitrogen) permettant le clonage de longs fragments d'ADN (**Figure IV.2, étape 5**). Les produits de ligature ont été transformés dans des cellules d'*E. coli* (One Shot® TOP10 Chemically Competent *E. coli*, Invitrogen) et les cellules étalées sur milieu sélectif LB solide contenant de la kanamycine à 50 µg/mL et incubées la nuit à 37°C. Les inserts des plasmides ont été amplifiés par PCR sur colonies en utilisant les amorces M13F et M13R s'hybridant de part et d'autre du site de clonage (**Tableaux 1 et 2 (PCR 6) de l'annexe IV.2**). La taille des amplicons a été visualisée par électrophorèse sur gel d'agarose 1%. Les clones ayant des inserts d'une taille supérieure à 1kb et étant positifs à la PCR EMTF/R₃ (**Tableaux 1 et 2 (PCR 4) de l'annexe IV.2**) ont été choisis pour être séquencés.

• Séquençage des clones

Les plasmides sélectionnés précédemment ont été extraits (NucleoSpin-plasmid, Macherey-Nagel) et leurs inserts séquencés par la méthode Sanger via un prestataire extérieur (Biofidal) (**Figure IV.2, étape 6**). La qualité des séquences a été évaluée en visualisant les chromatogrammes des séquences grâce au logiciel Chromas lite. Les séquences ont été corrigées et modifiées manuellement sous Word (Microsoft®). Après avoir éliminé les séquences du plasmide pCR-XL-TOPO et les adaptateurs A et B, les séquences obtenues ont été alignées aux séquences EMT. Par ailleurs, des homologues de séquences ont été recherchés parmi les séquences de la base de données nr de GenBank en utilisant l'outil BLASTn et Global Alignment (<http://www.ncbi.nlm.nih.gov/BLAST/>).





Résultats

Deux expériences de capture ont été réalisées : les captures A et B. Ces 2 expériences diffèrent l'une de l'autre par la longueur des sondes utilisées ce qui peut affecter la stabilité des hybrides sondes / séquences environnementales ciblées et la spécificité de l'hybridation. La capture A a été réalisée en utilisant 40 sondes d'ARN biotinylées de 50 pb spécifiques des parties 5' et 3' des régions codantes des gènes *emt* des sous-familles 1, 2, 3, 4, 5 et 7. Cette capture n'ayant pas donné de résultats probants, la nature des sondes utilisées a été remise en question. Ainsi, pour la capture B, 16 sondes beaucoup plus longues, d'une taille de 203 à 349 pb (couvrant chacune entre 55 et 95% la région codante *emt*), identiques à des séquences de 4 sous-familles d'EMT (1, 3, 4 et 5), ont été conçues, synthétisées et utilisées sur la même matrice d'ADN utilisée pour la capture A.

La première étape après capture consiste à vérifier que celle-ci a effectivement conduit à l'enrichissement en séquences cibles. Cette vérification de l'enrichissement en séquences EMT a été réalisée d'une part, par PCR nichée à l'aide d'amorces dégénérées capables d'amplifier l'ensemble des gènes *emt* et d'autre part, par PCR conventionnelle en utilisant des amorces spécifiques des gènes *emt* de la sous-famille 1. Ensuite, les produits de capture ont été clonés et les inserts de plus grandes tailles qui contiennent des séquences d'*emt* ont été séquencés.



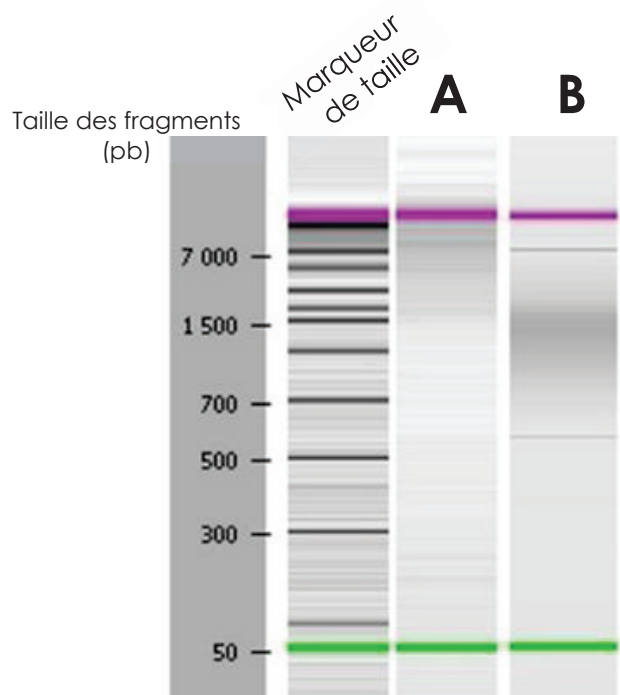


Figure IV.3 : Visualisation de la taille des extraits d'ADN par électrophorèse en capillaire (Agilent DNA1200) (A) Après extraction d'ADN métagénomique de sol et (B) après la fragmentation et l'amplification de cet ADN. Ce contrôle de la taille des fragments d'ADN métagénomique est indiqué par une étoile sur la **figure IV.2**.



Analyse du pool d'ADN fragmenté.

La taille des fragments d'ADN contenus dans l'extrait de sol a été vérifiée par électrophorèse en capillaire à deux reprises lors de la procédure : A) après l'extraction d'ADN total du sol et B) après la fragmentation de l'ADN total du sol par les G-tubes et l'amplification de ces ADN par l'ADN polymérase Q5 (**Figures IV.2 et 3**). Après l'extraction d'ADN de sol (**Figures IV.3A**), le profil d'électrophorèse que nous observons est sous forme de « traînée » de 50 pb à 10 kb qui est plus intense entre 1,5 et 10 kb nous indiquant que cette fraction contient plus d'ADN que le reste de la traînée. Après fragmentation et amplification (**Figures IV.3B**), nous observons une traînée qui cette fois est beaucoup plus intense entre 0,7 et 2 kb nous indiquant que cette fraction contient plus d'ADN. Nous pensons que l'amplification des ADN par PCR a enrichi les extraits d'ADN de sol en fragments de tailles contenues dans cette gamme.

Malgré cet enrichissement en fragments d'ADN de tailles comprises entre 0,7 et 2 kb, nous avons continué la procédure car une étape de purification et de sélection sur gel d'agarose de la fraction qui nous intéresse de séquences comprises entre 3 et 10 kb est prévue avant clonage.



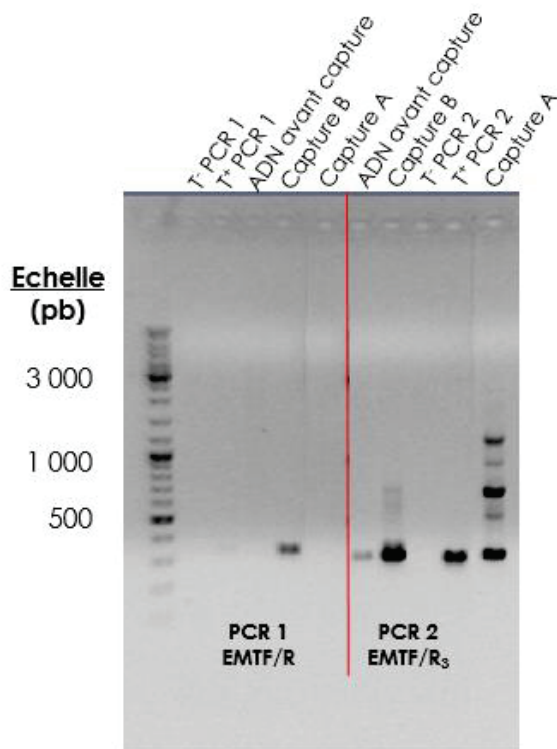


Figure IV.4 : Vérification de l'enrichissement en séquences cibles *emt* après les captures A et B par PCR nichée en utilisant les couples d'amorces dégénérées EMTF/R (PCR 1) puis EMTF/R₃. (PCR 2).

T⁻ : Témoin négatif de PCR (eau)

T⁺ : Témoin positif de PCR (plasmide portant une séquence *emt* de la sous-famille 1 (1 ng))

ADN avant capture : Extraits d'ADN de sol Lommel (1 ng par PCR)

Captures A et B : Extraits d'ADN de sol obtenus après capture A ou B (1 ng par PCR).



Analyse de l'enrichissement en séquences cibles

Cette analyse nous a permis de vérifier que la capture a fonctionné afin de pouvoir passer à l'étape de clonage.

Le résultat de la PCR nichée (**Figure IV.4**) nous montre que lors de la première PCR, seuls les extraits d'ADN issus de la capture B s'amplifient contrairement aux extraits d'ADN avant capture ou issus de la capture A. En reprenant 2 µL de produits PCR de la première PCR comme matrice pour la deuxième PCR utilisant les amorces EMTF/R₃, tous les extraits sont amplifiés. L'ADN avant capture s'amplifie très faiblement alors que l'ADN après les captures A et B donne un bande d'amplification plus intense ce qui montre que les 2 captures ont bien permis un enrichissement en séquences *emt*. Toutefois, les produits de la capture A donnent également des fragments d'amplification aspécifiques de grandes tailles dont l'origine reste inexplicée.

Etant donné que : 1) les extraits d'ADN issus de la capture B s'amplifient dès la première PCR de la PCR nichée contrairement aux extraits d'ADN avant capture et ceux issus de la capture A, et que 2) la seule différence entre la capture A et la capture B réside dans l'utilisation d'un jeu de sondes différentes, on peut en conclure que le nouveau jeu de sondes de plus grandes tailles utilisé lors de la capture B a permis un meilleur enrichissement de fragments d'ADN métagénomiques porteurs des séquences *emt*.



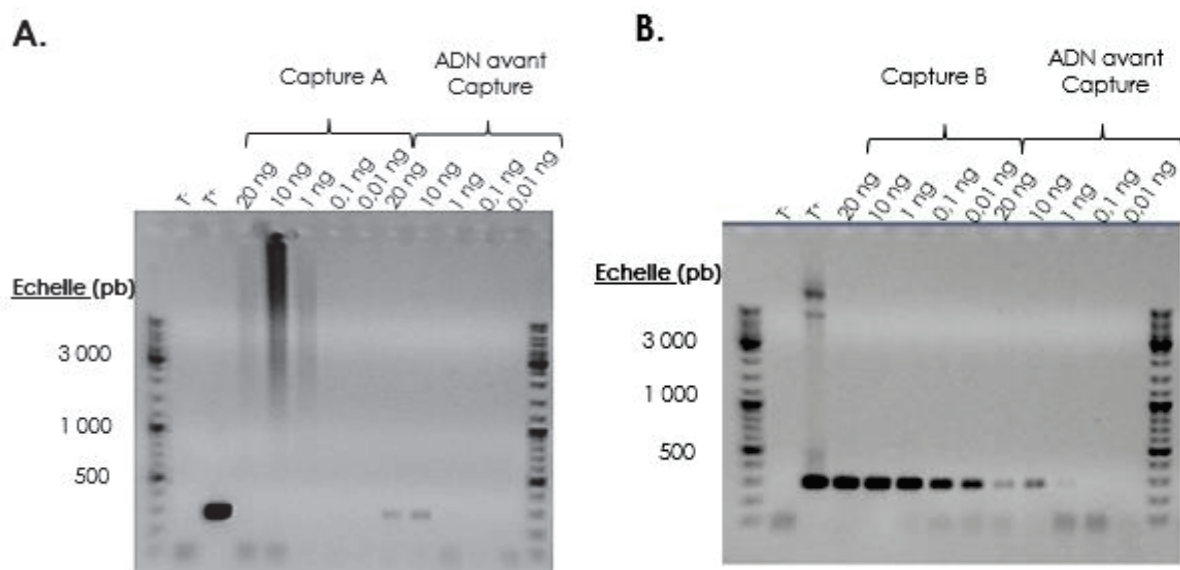


Figure IV.5 : Evaluation de l'enrichissement en séquences *emt* après captures A (A.) et B (B.) par PCR en utilisant les amorces EMT1F/R spécifiques de la sous-famille 1.

Les PCR ont été réalisées en ajoutant dans le mélange réactionnel 20 ; 10, 1 ; 0,1 et 0,01 ng d'extraits d'ADN de sol Lommel avant et après captures,

T⁻ : Eau

T⁺ : Plasmide portant une séquence de la sous-famille 1 (1 ng)

ADN avant capture : Extraits d'ADN de sol de Lommel

Capture A ou B : Extraits d'ADN de sol de Lommel après capture ciblée de séquences *emt*.



Une PCR utilisant des amorces spécifiques de la sous-famille 1 a aussi été réalisée en utilisant une gamme de quantités décroissantes de fragments d'ADN métagénomiques avant et après captures allant de 20 à 0,01 ng (**Figure IV.5**). Cette PCR révèle que les séquences *emt1* présentes dans les extraits d'ADN de sol de Lommel avant capture s'amplifient faiblement lorsqu'on ajoute entre 10 et 20 ng d'extraits d'ADN de sol (**Figure IV.5 A et B**). Or, les séquences *emt1* ne s'amplifient pas lorsque les extraits d'ADN issus de la capture A sont utilisés comme matrice (**Figure IV.5 A**).

Après l'échec de l'enrichissement en séquences *emt1* lors de la capture A, nous avons voulu analyser la spécificité des sondes testées. Pour ce faire, les produits amplifiés par PCR nichée ont été clonés et séquencés. Tous les clones analysés contiennent une séquence *emt* appartenant à la sous-famille 4 ce qui suggère la présence d'un biais dans l'enrichissement des ADN lors de cette capture qui pourrait expliquer l'absence d'amplification avec les amorces PCR spécifiques de la sous-famille 1.

Par contre, les séquences *emt1* s'amplifient fortement avec seulement 0,01 ng d'extraits d'ADN issus de la capture B (**Figure IV.5 B**) ; ceci suggère que la capture B a permis un enrichissement en séquences *emt1* d'un facteur d'au moins 1000 par rapport aux extraits d'ADN avant capture. La longueur des sondes semble donc être un facteur crucial à la technique de capture.



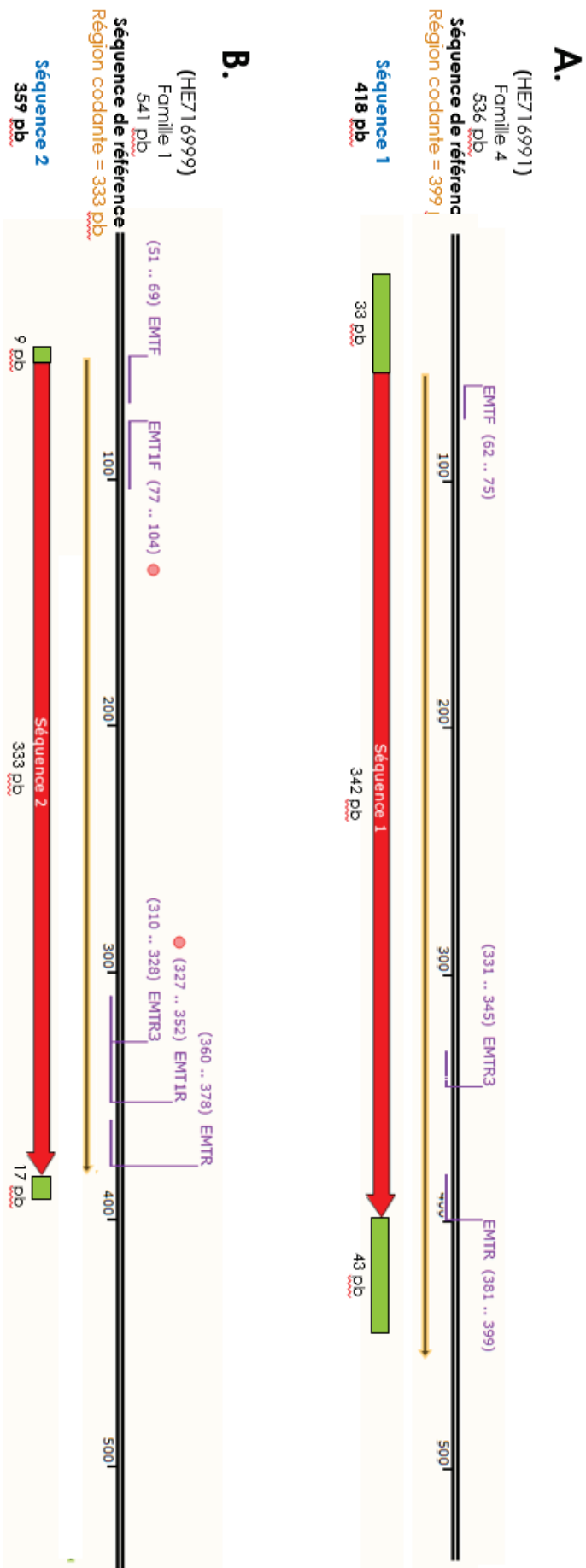


Figure IV.6 : Séquences *emt* obtenues après la capture B sur des extraits d'ADN de sol de Lommel. A. Séquence 1 capturée, B. Séquence 2 capturée. Les amorces EMTF, EMTR, EMTR₃, EMT1F (pastille rouge) et EMT1R (pastille rouge) ont été placées sur les séquences.



Clonage et séquençage des fragments d'ADN issus de la capture B

Le clonage de fragments d'ADN issus de la capture B a permis d'obtenir 650 clones. Sur 395 clones analysés par PCR, seuls 50 clones possédaient un insert d'une taille supérieure ou égale à 1 kb (variant entre 1 et 1,7 kb) ce qui est une taille insuffisante par rapport aux objectifs que nous nous étions fixés de plus de 3kb. Par ailleurs, aucun de ces 50 fragments d'ADN n'a pu être amplifié avec les amorces EMTF/R₃, ils ne semblent donc pas contenir de régions codantes *emt*. Aussi, 19 clones capturés choisis au hasard ont été séquencés pour identifier la nature de ces séquences clonées. En réalisant un BLASTn contre les séquences de la base de données nr, seules 2 séquences clonées sur 19 (10%) présentent des régions homologues à des séquences *emt* (**Figure IV.6**). La séquence 1 de 418 pb (**Figure IV.6 A**) possède 33 et 43 pb supplémentaires respectivement avant l'ATG et après le codon stop d'une région codante. Cette région codante est homologue à des séquences *emt* dont une séquence de la sous-famille 4 obtient le meilleur score (HE716991). Cette région recouvre à 86% la première partie de la région codante de la séquence *emt4* nous indiquant que cette séquence est légèrement plus courte que la région codante *emt4*.

La séquence 2 de 359 pb (**Figure IV.6 B**) possède 9 et 17 pb respectivement avant l'ATG et après le codon stop d'une région codante aussi homologue à des séquences d'*emt* dont le meilleur score de BLAST correspond à une séquence de la sous-famille 1 (HE716999). Les régions codantes de ces deux séquences se recouvrent à 100% avec seulement une différence de 2 nucléotides.

La taille des séquences obtenues après capture est inférieure à la taille de séquences attendues (les régions bordantes ont une taille variant de 9 à 43 pb alors que nous attendions des séquences supérieures à 3 kb) compromettant l'identification de biomarqueurs phylogénétiques. De plus, nous n'avons pas trouvé de séquences initiatrices de transcription comme les



boîtes CCAAT ou TATA dans les régions situées en 5'. Toutefois, cette technique alliant la capture de gènes suivie d'un séquençage direct a tout de même permis d'obtenir des gènes *emt*.

Discussion et Perspectives

L'objectif de cette étude était d'obtenir les régions génomiques flanquant les régions codantes des gènes *emt* correspondant d'une part aux régions promotrices de ces gènes et d'autre part aux gènes proximaux qui leur sont associés. L'analyse des régions promotrices aurait permis d'identifier des motif(s) de liaison à des facteurs de transcription déjà connus donnant ainsi des indications sur les modes de régulation de ces gènes alors que l'analyse phylogénétique des gènes proximaux aurait permis d'identifier l'origine taxonomique des organismes qui portent les *emt*. Etant donné que les gènes *emt* sont rares dans les extraits d'ADN de sol et que leur isolement par deux techniques de marche chromosomique n'avait précédemment pas fonctionné, nous avons choisi de mettre au point une procédure de capture ciblée associée au clonage/séquençage bas-débit des produits de capture pour obtenir ces séquences d'ADN.

La capture A ne nous a pas permis d'obtenir un taux d'enrichissement important en séquences cibles et les séquences obtenues appartenaient exclusivement à la sous-famille 4. Nous avons alors remis en question les sondes utilisées et émis l'hypothèse que la taille des sondes ARN de 50 pb était trop petite par rapport aux longs fragments d'ADN métagénomique à capturer étant donné que les lots de sondes utilisés lors d'hybridation en phase liquide sont un paramètre critique (**Pozhitkov et al. 2006 ; ribière et al. 2016**). De nouvelles sondes spécifiques plus longues (recouvrant toute la région codante) ont donc été utilisées lors de la capture B qui, cette fois ci, a permis effectivement d'enrichir l'extrait d'ADN en séquences *emt*, d'un facteur 1000 dans le cas d'*emt1*. Cependant, la taille et la nature des fragments obtenus



lors de cette capture ne permettent pas de répondre à la question posée car la majorité (90%) des séquences analysées ne présentaient aucun homologue ni dans les bases de données, ni dans le jeu de séquences utilisées pour concevoir les sondes. Ainsi, avec ces sondes, un taux important de faux-positifs a été obtenu. De plus, bien que nous soyons parvenus à cloner et séquencer 2 séquences *emt*, les régions flanquantes obtenues ne dépassaient pas une 50^{aine} de nucléotides, ce qui est bien en deçà des résultats attendus.

La capture de gènes ne nous a pas permis de remplir notre objectif ; aussi, dans cette partie, nous discuterons les causes possibles de cet échec et les moyens d'améliorer cette technique de capture de gènes dont l'objectif est de capturer de longs fragments d'ADN métagénomique. Sont indiquées en bleu sur la **Figure IV.2** les différentes étapes de la capture de gènes que nous pouvons modifier en vue d'une amélioration de cette technique.

Un nombre important de produits de capture aspécifiques ?

La capture ciblée est une technique très récente ayant été utilisée peu de fois sur des matrices d'ADN environnementaux. **Denonfoux et ses collaborateurs (2013)** ont, pour la première fois, adapté cette technique sur des matrices d'ADN environnementaux. Lors de cette étude, ils ont dans un premier temps mis au point une capture de gènes *mcrA*, codant une *Methyl Coenzyme M Reductase* impliquée dans l'utilisation du méthane chez les Archées, à partir d'extrait d'ADN de culture pure de *Methanosarcina acetivorans* (Archée lacustre). Après deux cycles d'hybridation/élution, sur les 6 clones sélectionnés et séquencés par Sanger, tous étaient homologues aux séquences recherchées, donnant ainsi un taux de réussite de 100% contre moins de 10% pour notre étude. Dans un second temps, ils ont capturé les gènes *mcrA* à partir d'extraits d'ADN lacustres et après deux cycles de capture, séquencés les ADN capturés par des méthodes haut-débit. Ils n'ont cependant pas cloné leurs extraits d'ADN environnementaux après capture. Par ailleurs, au sein de mon équipe d'accueil, **Bragalini et ses collaborateurs**



(2014) ont utilisé un jeu de 35 sondes ARN dégénérées de 31 pb, prédites à l'aide du logiciel KASpOD en utilisant un jeu de données contenant des séquences de xylanases GH11 (Glycosides Hydrolases 11, 639-2099 pb) et ciblant le site actif de 540 pb. Ils ont cloné les fragments environnementaux capturés et ont pu obtenir un taux d'enrichissement de 72%. Les xylanases sont exprimées par les communautés fongiques des sols forestiers qui sont des sols riches en matière organique et en biomasse. Ainsi, dans ces sols ces séquences ne sont pas rares (sans doute de l'ordre d'une séquence parmi 10^4 autres séquences environnementales), elles sont aussi mieux représentées dans les bases de données (au moins 342 séquences disponibles) ce qui facilite la conception de sondes de sensibilités plus importantes. Cependant, dans le cas de capture de gènes d'enzymes fongiques, cette technique ne semble pas avoir un taux de réussite élevé puisqu'elle a fonctionné dans le cas des xylanases GH11 (**Bragalini et al. 2014**), des peroxydases DYP mais pas dans le cas des endoglucanases GH45 ni dans le cas des peroxygénases UPO bien que les produits de capture de ces derniers furent séquencés par PacBioRS (**Bragalini et Adamo, Communications personnelles**) qui est une technologie de séquençage moyen-débit produisant de long reads mais un taux d'erreur élevé contrairement au séquençage haut-débit qui produit de courts reads mais avec un taux d'erreurs faible (**Gasc et al. 2015**). Ceci nous laisse penser que selon la séquence cible, il n'est pas rare que la capture ciblée soit difficile à mettre en œuvre.

Selon **Noguera et ses collaborateurs (2014)**, les sondes idéales doivent avoir 2 caractéristiques : i) être spécifiques aux séquences recherchées (spécificité) et iii) arriver à s'hybrider aux séquences cibles même lorsque celles-ci sont en faible quantité dans les extraits d'ADN (sensibilité). Cependant, ces caractéristiques doivent être équilibrées. Si les sondes recouvrent toute la séquence cible, comme pour notre deuxième jeu de sondes, les chances d'obtenir des séquences différentes des séquences recherchées sont faibles et on perd en sensibilité ce qui peut être très



problématique dans le cas de séquences rares dans l'environnement comme les *emt* (**Noguera et al. 2014, Ilie et al. 2013**).

Dans notre cas, le premier jeu de sondes a été synthétisé via le logiciel KASpOD (**Parisot et al. 2012**) et le deuxième jeu de sondes a été synthétisé directement à partir des régions codantes de familles d'*emt*. Une des dernières étapes de la conception des sondes est la vérification de la spécificité des sondes *in silico* par un BLAST ce qui a été fait dans le cas de notre étude. Cependant, cette méthode de vérification ne garantit pas une bonne spécificité des sondes (**Ilie et al. 2013**) et il est possible que les sondes se soient hybridées à des séquences d'ADN différentes des séquences d'intérêt qui sont : 1) présentes en plus grande quantité dans les extraits d'ADN de sol et 2) sans homologues dans les bases de données ce qui empêche leur détection lors des analyses de spécificité *in silico*. Aussi, il semble que dans certains cas, la spécificité ne puisse être connue *a priori* avant l'utilisation de jeu de sondes (**Pozhitkov et al. 2006**).

L'efficacité d'hybridation des sondes à l'ADN cible va aussi grandement dépendre de conditions autres que la conception des sondes telles que les caractéristiques de l'échantillon (l'abondance des séquences cibles, qualité des extraits d'ADN) et des conditions expérimentales qui peuvent être optimisées (la combinaison de sondes utilisées, leur densité dans le milieu réactionnel, la possibilité d'hybridations inter-sondes, la température et le temps d'hybridation *in vitro* et les conditions de lavages (**Pozhitkov et al. 2006, Manoharan et al. 2015**) (**Figure IV.2**, étape 4 : hybridation en solution). Par exemple, **Manoharan et ses collaborateurs (2015)** ont montré que la réduction du temps d'hybridation augmentait le nombre de séquences aspécifiques.



Des produits de capture de trop petite taille ?

Bien que la sensibilité des sondes ait été remise en question et que nous ayons tout de même réussi à obtenir deux séquences *emt*, les séquences capturées avaient toutes une taille bien inférieure aux tailles attendues (1 kb récupéré en moyenne contre 3-10 kb espérés) ; les régions bordantes étaient donc trop courtes. Nous avons donc aussi remis en question les étapes : 1) de fragmentation de l'ADN métagénomique ; 2) d'amplification et d'hybridation des fragments d'ADN capturés, 3) de purification; et, 4) de clonage de l'ADN capturé.

Après l'extraction de l'ADN métagénomique de sol, celui-ci a été fragmenté en utilisant des G-tubes (Covartis) avec une coupure *a priori* tous les 10 kb puis les petits fragments d'ADN (<500 pb) ont été éliminés avec un traitement AMPur (Beckman Coulter Genomics) (**Figure IV.2**, étape 3 : préparation ADN). La fragmentation de l'ADN peut être plus ménagée en réduisant la force de centrifugation de 8 300 g à 3 000 g afin d'obtenir une fragmentation théorique deux fois plus grande.

Ensuite, beaucoup d'amplifications par PCR ont été réalisées avant, pendant et après l'hybridation, avec un nombre important de cycles (25 cycles). Or la PCR est une technique entraînant un certain nombre de biais tels que la sélection préférentielle des fragments de petites tailles (**Kanagawa 2003, Dabney & Meyer 2012**) que nous retrouvons d'ailleurs lorsqu'on observe les pools d'ADN par électrophorèse en capillaire (**Figure IV.3**). Il est envisageable de diminuer le nombre de cycles de PCR de 25 à 15 pour réduire ce biais.

Il est aussi possible que lors des étapes d'hybridation et de lavage, les fragments de grandes tailles ne se soient pas accrochés ou se soient décrochés car les conditions n'étaient pas assez adaptées. Elles pourraient être optimisées à l'aide de plans d'expérience testant différentes températures d'hybridation avec des solutions d'hybridation et de lavage de stringences différentes sur des extraits d'ADN de sol contenant des gammes de concentrations en plasmides linéaires contenant des séquences *emt*.



La sélection de petits fragments est aussi favorisée lors des étapes de clonage (**Figure IV.2**). Pour remédier à cela, nous avons essayé différentes méthodes de ligature avec le kit TOPO-XL-Cloning kit prévu pour cloner des fragments d'une taille de 10 kb fonctionnant sur le principe du *TA cloning* et la T4 DNA ligase d'Invitrogen avec le vecteur du kit mais sans amélioration de la taille des fragments clonés. Nous pouvons envisager l'utilisation de protocoles de fosmides plus adaptés au clonage de grands fragments d'ADN (**Ufarté et al. 2016**).

Les méthodes de purification d'ADN avant clonage peuvent être également remises en cause avec l'utilisation du kit QIAquick (Quiagen) (**Figure IV.2**). Pour cela nous avons essayé une séparation des ADN capturés par électrophorèse bidimensionnelle (**Yadav et al. 2014**) mais sans succès. Il s'avère que le taux de récupération du protocole QIAquick est inférieur à 30% ce qui est faible (**Harnpicharnachai et al. 2007**). Nous pouvons par exemple mettre en place une purification par la méthode de la « gouttière » consistant en une électrophorèse sur gel d'agarose pendant 45 min, puis l'ajout de 30% de PEG 8 000 dans une gouttière « creusée » à la limite recherchée et l'ADN migre pendant 30 min nous permettant ainsi de récupérer la fraction d'ADN voulue (**Harnpicharnachai et al. 2007, Engel et al. 2012**). Au-delà de l'étape de purification des ADN capturé, toutes les étapes d'élimination de petits fragments utilisaient le kit AMPur (Beckman Coulter Genomics) (**Figure IV.2**) qui éliminait les fragments d'ADN inférieurs à 500 pb. Nous pouvons imaginer d'utiliser aussi la méthode de la « gouttière » afin de sélectionner préférentiellement à chaque étape de purification les fragments d'ADN supérieurs à 3 kb.

Perspectives pour obtenir les régions flanquant les gènes *emt*.

Denonfoux et ses collaborateurs (2013) ont utilisé une approche de capture de gènes couplée à du séquençage massif (*Shotgun*) à partir d'extraits d'ADN métagénomique d'une couche anoxique du lac Pavin



(France). Cette technique leur a permis de passer de 0,003% des séquences à environ 41% de séquences *mcrA* dans leur jeu de données de séquençage massif. Nous pouvons envisager d'utiliser des techniques de séquençage massif sur nos extraits d'ADN capturé et d'obtenir les régions promotrices des séquences *emt* par traitement bioinformatique. De plus, l'expérience de **Bragalini et ses collaborateurs (2014)** nous a montré que cette méthode pouvait être versatile et peut nécessiter de lourdes optimisations concernant : le jeu de sondes utilisées, l'étape de fragmentation de l'ADN, le nombre de cycles d'amplification, les conditions d'hybridation et d'élution des fragments d'ADN capturés, le choix de la méthode de purification de l'ADN capturé et le choix de la méthode de clonage de cet ADN.

Par ailleurs, il n'est pas certain qu'un biomarqueur phylogénétique soit présent dans les régions proximales des séquences *emt*. Pour identifier les organismes qui portent ces gènes, il semble que les méthodes d'isolement cellulaire pourraient représenter une alternative à explorer (**Brehm-Stecher & Johnson 2004, Pivetal et al. 2015, Gasc et al. 2015**). Elles permettent d'isoler des cellules entières ce qui nous donnerait « immédiatement » des informations sur l'organisme qui porte ces gènes (biomarqueurs moléculaires phylogénétiques et fonctionnels et caractéristiques morphologiques)





Conclusions générales et perspectives de la thèse

Mes travaux de thèse avaient pour objectif la caractérisation d'une famille de gènes codant des protéines riches en cystéines, dont les seules informations à notre disposition étaient issues du travail de **Lehembre et ses collaborateurs (2013)** à savoir que ces gènes : 1) confèrent une résistance au Cd lorsqu'ils sont exprimés dans la levure, 2) n'ont aucun homologue dans les bases de données, 3) ont été détectés dans des ADN environnementaux extraits de sol et de lac par PCR nichée, 4) codent des protéines riches en cystéines conservées, et 5) se regroupent en 9 sous-familles géniques. Afin de caractériser cette famille de gènes, nous avons essayé de répondre aux trois questions suivantes : 1) Les protéines CRP peuvent-elles être considérées comme appartenant à la superfamille des métallothionéines (**Chapitre 2**) ? 2) L'expression de ces gènes est-elle induite par l'apport exogène de métaux dans le milieu (**Chapitre 3**) ? De quel(s) groupe(s) d'organisme(s) sont issus ces gènes (**Chapitre 4**) ? Ces questions s'inscrivent dans le cadre plus large de l'étude des mécanismes de résistance mis en place à l'échelle moléculaire et cellulaire suite à la présence de métaux dans l'environnement et relèvent de l'écotoxicologie microbienne. Elles s'inscrivent aussi dans le cadre de l'inventaire de la biodiversité fonctionnelle des sols relevant de problématiques d'écologie microbienne.

Dans le **chapitre 1**, nous avons dans un premier temps décrit l'état des connaissances concernant l'écosystème sol et précisé l'importance des communautés de microorganismes eucaryotes pour ses fonctions et services écosystémiques, les origines des apports en éléments métalliques dans les sols, les réponses cellulaires face à un stress métallique et les mécanismes de résistance aux métaux.

Dans un second temps, nous avons vu que les MT sont des protéines riches en cystéines dont les positions sont conservées au sein d'une même



famille. Elles sont capables de chélater les métaux grâce aux atomes de soufre présents dans leurs cystéines. Etant donné que les outils informatiques ne permettent ni de modéliser, ni de prévoir quelle sera l'affinité d'une MT pour un métal donné, il est indispensable de réaliser des analyses fonctionnelles *in vitro* afin d'affirmer qu'une MT putative est réellement une MT. L'expression basale des MT permet notamment le maintien des équilibres en éléments métalliques dans les systèmes cellulaires. Ces propriétés d'induction sont utilisées pour préciser son rôle fonctionnel dans le métabolisme cellulaire. Les MT sont principalement retrouvées chez les organismes eucaryotes mais sont aussi présentes chez quelques bactéries.

Les CRP ont toutes les caractéristiques de MT putatives puisqu'elles contiennent un fort pourcentage en cystéines ayant des positions conservées. Sont présentés ci-dessous les principaux résultats de leur caractérisation et quelques perspectives possibles.

- **Les CRP peuvent-elles être considérées comme des métallothoinéines (Chapitre 2) ?**

Nous avons émis l'hypothèse que les CRP sont des MT. Par un raisonnement déductif, nous avons appliqué les mêmes méthodes d'analyses fonctionnelles des MT aux CRP (**Chapitre 2**). J'ai réalisé une partie de ces analyses en Espagne à Barcelone en collaboration avec les équipes des professeurs Silvia Atrian (Université de Barcelone) et Mercè Capdevila (Université Autonome de Barcelone), toutes deux spécialistes de la caractérisation fonctionnelle des MT (**Palacios et al. 2011b**).

Après la surexpression de 5 CRP dans *Escherichia coli*, nous les avons extraites, purifiées et avons quantifié le nombre d'atomes de Cu, de Zn et de Cd liés aux CRP par ESI-TOF-MS et ICP-AES. En parallèle, nous avons décrit et comparé leurs séquences en acides aminés avec celles des MT connues. Ces protéines ont également été exprimées dans des levures sensibles au Zn et au



Cu dans le but de compléter leur caractérisation par complémentation fonctionnelle débutée avec le Cd (**Lehembre et al. 2013**).

Les expériences de chélation *in vitro* ont montré que les CRP sont capables de se lier, dans les conditions de l'étude, à 4-15 atomes de Cu, 5-11 atomes de Zn et 10-16 atomes de Cd en fonction des isoformes considérées. Dans des levures sensibles aux métaux, les CRP confèrent un phénotype de résistance au Cd (**Lehembre et al. 2013**) et au Zn mais, de manière inexplicée, pas au Cu.

Leurs séquences en acides aminés sont composées de 20-24% de Cys avec une structure en deux domaines comme chez les plantes. Ces domaines riches en cystéines, de 45-52 acides aminés, sont séparés par une séquence de 18-25 acides aminés. Elles possèdent une taille importante de 110-132 acides aminés comparée aux autres métallothionéines (souvent composées de 20-80 acides aminés). Leurs cystéines conservées sont organisées en motifs dont certains (CCC, CXCC) sont assez rares au sein des MT. Les analyses de similarité de séquences par HMM nous ont permis de confirmer que les CRP n'étaient similaires à aucune famille de MT connues et forment ainsi une toute nouvelle famille de MT. Nous avons nommé ces protéines : EMT pour « *environmental metallothioneins* ».

Ces résultats ont fait l'objet d'une publication dans la revue *Journal of Inorganic Biochemistry* (**Ziller et al. 2017**).

Parmi les 9 sous-familles d'EMT, seules deux séquences de la sous-famille 1 et trois séquences respectivement des sous-familles 3, 4 et 5 ont été caractérisées. Nous pouvons envisager de caractériser les séquences des sous-familles restantes selon la même procédure.

De plus, les métallothionéines peuvent se lier à un large spectre de métaux. Il serait intéressant d'étendre leur caractérisation fonctionnelle à d'autres métaux constitutifs de pollutions émergentes comme les terres rares. Les EMT possèdent deux domaines avec des motifs en Cys différents qui peuvent entraîner des affinités particulières pour chacun de ces métaux. Nous



pourrions produire les protéines EMT en présence de mélanges de métaux afin de vérifier si ces deux domaines ont des affinités différentes.

Enfin, très peu de structures 3D sont disponibles pour les métallothionéines. Etant donné les propriétés singulières des EMT (grande taille, deux domaines riches en Cys), la caractérisation de leur structure 3D par dichroïsme circulaire ou par résonance magnétique nucléaire (**Sutherland & Stillman 2014**) permettrait d'améliorer nos connaissances sur les modalités de chélation.

• **L'expression des gènes *emt* est-elle induite par des métaux (Chapitre 3) ?**

Les MT sont des protéines impliquées soit dans le fonctionnement basal des cellules eucaryotes surtout en régulant l'homéostasie des métaux soit dans la réponse à des stress notamment métallique (excès de métal). Elles peuvent donc être exprimées constitutivement et/ou en réponse à des excès de métaux (**Chapitre 1**). Il est ainsi possible que l'expression des gènes *emt* soit accrue par la présence de métaux biodisponibles dans le milieu. Comme les organismes portant ces gènes sont inconnus, cela ne peut être étudié qu'indirectement par une approche de transcriptomique environnementale utilisant des sols au sein desquels les organismes correspondants sont présents.

Aussi, nous avons mis au point une procédure de co-extraction d'ADN et d'ARN à partir de sol provenant de microcosmes amandés ou non en Cd, Cu ou Zn, puis mis au point une procédure de quantification des acides nucléiques de la sous-famille *emt1* par PCR quantitative (qPCR en temps réel et ddPCR).

Dans une première série de microcosmes contenant une gamme de concentration croissante de Cd, après 3 semaines d'incubation, nous avons dosé le Cd total par ICP-MS, observé que 17% du Cd ajouté avait été lessivé mais que la majeure partie du Cd fixé au substrat est mobilisable lorsque le sol est acidifié ce qui est sans doute lié aux propriétés du sol. Les conditions



d'amplification ainsi que le jeu d'amorces et la sonde utilisés en PCR quantitative ont été testés et nous sommes parvenus à la conclusion que la technique de ddPCR était la mieux adaptée à notre modèle. En effet, la ddPCR a une variabilité de mesures moins importante comparée à la qPCR et une marge d'optimisation plus grande (ajout de BSA, T4 Gene 32, co-extraction). Toutefois, ces résultats et hypothèses devront être confirmés par plus de répétitions et de nouveaux essais.

Pour la suite de ce projet, il reste : à réaliser la quantification des métaux totaux, solubles et contenus dans les plantules de pins de la deuxième série de microcosmes amendés en Cd, Cu ou Zn ; à mettre au point la rétro-transcription des ARN en ADNc ; à vérifier la constance du rapport ARN/ADN lors de l'extraction des acides nucléiques du sol ; puis à comparer les microcosmes entre eux afin de pouvoir déterminer si l'apport de métaux dans ceux-ci induit une augmentation de la transcription du gène *emt1*. Cette méthodologie pourra aussi être étendue aux gènes des autres sous-familles d'*emt* qui peuvent réagir différemment à la présence de métaux.

- **De quel(s) groupe(s) d'organisme(s) sont issus les gènes codant les EMT (Chapitre 4) ?**

Selon les organismes considérés, l'expression des gènes de MT est régulée par divers facteurs cellulaires et environnementaux tels que la présence de métaux ou encore un état de stress oxydant. Ces facteurs conduisent au recrutement de facteurs de transcription qui se lient à de courtes séquences RE (*Responsive Element*) dans la région promotrice des MT. Ces séquences, parfois conservées entre organismes, peuvent être considérées comme des biomarqueurs fonctionnelles précisant l'implication de gènes dans certains processus métaboliques (**Chapitre 1**). Il serait donc intéressant d'obtenir et d'étudier les régions promotrices des *emt*. Par ailleurs, les organismes qui portent les séquences *emt* sont inconnus. Il se pourrait que



des biomarqueurs phylogénétiques se trouvent proches des gènes *emt* et leur caractérisation phylogénétique nous permettrait d'affilier les gènes *emt* à un groupe taxonomique connu. Pour ces 2 raisons, nous nous sommes donnés pour objectif de récupérer les séquences des régions en amont et en aval des gènes *emt*.

Les séquences *emt* étant en faible nombre dans l'environnement, nous avons choisi une méthodologie basée sur l'enrichissement en solution des séquences *emt* accompagnée de leurs régions bordantes par une capture de gènes ciblées et leur isolement par clonage, criblage par PCR spécifique et séquençage.

Après la capture de gènes *emt*, nous avons difficilement réussi à identifier des séquences d'ADN métagénomique contenant les gènes *emt*. Au total, seules 2 séquences contenant des gènes *emt* ont pu être isolées. De plus, ces deux séquences *emt* possédaient des régions bordantes très courtes : 33 pb et 43 pb pour la première séquence, 43 pb et 9 pb pour la deuxième séquence, respectivement pour les régions 5' et 3'. Ces courtes régions ne nous ont permis ni d'obtenir des séquences régulatrices, ni d'obtenir des biomarqueurs phylogénétiques nous empêchant ainsi de répondre aux questions que nous nous étions initialement posées.

Plusieurs facteurs ont pu conduire à ce résultat négatif. Ces facteurs peuvent être lors de la capture ciblée : la nature des sondes utilisées, les conditions d'hybridation de celles-ci avec l'ADN ou encore les conditions de « lavage » des billes magnétiques après hybridation. La multiplication d'étapes qui favorise la sélection de petits fragments comme par exemple les PCR, des clonages et des purifications d'ADN sur gels d'agarose ont aussi pu favoriser la sélection de petits fragments d'ADN métagénomique au détriment des plus longs fragments qui étaient recherchés. Afin de répondre à la question posée, il est nécessaire : 1) soit d'optimiser les nombreuses étapes citées précédemment, ce qui représente un travail laborieux; 2) soit de changer de



méthodologie d'enrichissement par exemple en choisissant d'isoler préalablement de l'environnement les cellules microbiennes portant ces gènes par des sondes utilisant un marquage magnétique couplé à de la micromanipulation ou de la microfluidique (**Pivetal et al. 2015**). Cette deuxième possibilité présente l'avantage de pouvoir nous donner aussi accès au génome complet des organismes qui portent les *emt* mais aussi à la morphologie cellulaire qui est indispensable à l'affiliation taxonomique des eucaryotes.

- **Améliorer les connaissances relatives à la diversité des métallothionéines en s'affranchissant des analyses de similarité de séquences.**

L'étude que nous avons conduite sur les *emt* révèle que la diversité des MT est encore loin d'être entièrement connue. Il est très probable que des MT soient présentes chez tous les organismes possédant des métalloenzymes à Cu et à Zn, et/ou vivant dans des environnements riches en ces métaux *a fortiori* chez tous les eucaryotes. Cependant, des MT n'ont été identifiées qu'au sein de 5 grands groupes d'eucaryotes (*alveolata*, *stramenopiles*, *chloroplastidia* (plantes), *metazoa*, *fungi* et *metamonada*) sur 17 au total (**Adl et al. 2012**).

Botella et ses collaborateurs (2012) décrivent que lors d'une infection bactérienne de cellules humaines, les macrophages phagocytent les bactéries qui sont alors emprisonnées dans des vésicules spécialisées. Dans ces vésicules, la stoechiométrie métallique est contrôlée par l'intervention de plusieurs pompes permettant la sortie du Fe et du Mg qui favorisent la croissance bactérienne et l'entrée de Cu, de Zn et de Mn fonctionnant comme bactéricides en concentrations élevées. En réponse à cela, *Mycobacterium tuberculosis* synthétise des pompes à efflux de Cu, des sidérophores et une des rares métallothionéines décrites chez les bactéries. Sans cette métallothionéine, sa virulence est grandement diminuée (**Gold et al. 2008, Ding et al. 2013**). Il est possible que la présence de MT puisse ainsi



conditionner de manière importante les interactions entre hôtes eucaryotes et parasites/symbiontes, en tant que facteurs de virulence ou symbiotiques, en permettant à ces derniers d'être adaptés à « l'environnement intracellulaire ».

Ainsi, dans le cadre de mon année d'Attaché Temporaire d'Enseignement et de Recherche (ATER) à l'université Claude Bernard Lyon 1, j'ai pour projet de recherche l'identification des métallothionéines putatives dans les bases de données pour améliorer nos connaissances sur leur diversité. Il a été en effet rapporté à plusieurs reprises (**Blindauer 2011, Palacios et al. 2014**) que beaucoup de MT ne sont pas correctement annotées dans les bases de données et se retrouvent souvent qualifiées de « protéines inconnues ». En effet, l'identification de protéines se fait par des algorithmes qui ne sont pas adaptés à ce genre de protéines très divergentes entre elles car issues de multiples événements de convergence évolutive.

Il serait possible de les identifier manuellement au sein de protéomes et de métaprotéomes disponibles dans des bases de données.

- Protéomes : 1) d'organismes eucaryotes appartenant à des phyla sans métallothionéines décrites, 2) d'organismes procaryotes pathogènes ou symbiontes d'eucaryotes.
- Métaprotéomes/métatranscriptomes : 1) d'environnements pollués par des métaux, 2) d'environnements acides qui sont des milieux où le Cu et le Zn sont sous leurs formes les plus réactives.

Un script a déjà été développé par Audrey Dubost, bio-informaticienne du Laboratoire d'Ecologie Microbienne, permettant de trier des jeux de séquences d'acides aminés en fonction de leur richesse en Cystéines. Toutes les séquences ayant un taux de cystéines supérieur à 15% seront analysées et le cas échéant considérées comme des MT putatives. J'ai déjà testé cette démarche sur des protéomes de 7 microorganismes eucaryotes chez lesquels aucune MT n'a été décrite et j'ai identifié trois MT putatives préalablement annotées comme « protéines hypothétiques » chez l'amibe *Dictyostelium*



discoideum AX4. Jusqu'à présent, aucune MT n'a été décrite chez le groupe des amibes ce qui est encourageant.

De plus, **Vita et ses collaborateurs (2015, 2016)** ont identifié chez des bactéries méthanotrophes (caractérisées par un fort besoin en cuivre) plusieurs thiolates tétramériques nommées Csp (*Copper Storage Protein*). Chaque monomère contient 122 acides aminés avec une composition d'environ 10% de Cys. Ces Cys sont impliquées dans le stockage de 80 atomes de Cu. Les positions des cystéines sont conservées et il se pourrait que ces protéines aient des propriétés proches des MT. Elles pourraient justifier la création d'une famille de protéines incluant dans un premier temps les MT et les CSP et ayant les mêmes propriétés que le « système thiol ».

A plus long terme, les MT putatives présentant les caractéristiques les plus exotiques pourront être caractérisées fonctionnellement après expression et purification dans un hôte hétérologue par des approches spectrométriques.

• Pour une synthèse entre la génomique environnementale et biologie des systèmes ?

Actuellement, l'écologie microbienne et l'écotoxicologie microbienne sont des disciplines qui ont largement hérité des concepts et des techniques de la biologie moléculaire. Cela se retrouve dans les problématiques qu'elles abordent comme par exemple l'étude des réponses de communautés fonctionnelles de microorganismes d'un environnement face à différentes perturbations (naturelles ou anthropiques) à travers l'analyse des constituants cellulaires (ADN ribosomiques, gènes de fonction,...). Etant donné la difficulté de l'étude des microorganismes dans leur environnement, ce lien s'est renforcé suite à l'amélioration constante des outils bioinformatiques associés à l'émergence et au développement des technologies de séquençage massif de plus en plus performantes utilisées en génomique environnementale (métagénomique, métatranscriptomique et métaprotéomique). Cependant,



ces méthodologies ne sont pas sans biais. Par exemple, l'extraction de l'ensemble des acides nucléiques de sol favorise les molécules de certaines communautés en fonction de la procédure utilisée (**Delmont et al. 2011**), ou encore lors de l'étape d'assemblage des génomes microbiens à partir des mélanges d'ADN environnementaux ne pouvant être réalisée que pour quelques populations de la communauté et ne donnant que des génomes chimériques de populations d'une espèce voire de plusieurs espèces (**Prosser 2015**). Appliquées à des extraits de molécules environnementales, ces techniques ne permettent d'accéder qu'à des informations sur une partie des populations microbiennes présentes dans les écosystèmes (**Prosser 2015**).

Cependant, la plus petite unité fonctionnelle du vivant est la cellule (les individus) et lorsque ces techniques sont appliquées à des organismes unicellulaires en culture pure, les inventaires moléculaires qu'elles fournissent ne permettent pas de comprendre la physiologie cellulaire, car elles ne donnent pas accès aux réseaux métaboliques qui reposent sur la plasticité et le dynamisme des interactions moléculaires (**Barabási & Oltvai 2004**). Ainsi, en parallèle à ces approches «-omiques», une vision systémique du fonctionnement cellulaire s'est développée depuis une dizaine d'années et a donné naissance à la biologie des systèmes (*systems biology*) (**Boogard et al. 2007**). Cette discipline, née du constat que les propriétés et le fonctionnement cellulaire ne peuvent se réduire à l'inventaire de ses constituants et que de nouvelles propriétés émergent de réseaux d'interactions moléculaires qui peuvent changer en fonction des circonstances et au cours du temps. Les observations des processus métaboliques sont ainsi le fruit des interactions complexes entre les populations de molécules dans la cellule. Cette discipline « post-génomique », s'appuie largement sur les inventaires moléculaires des technologies de génomique fonctionnelle mais leur associe des méthodes plus anciennes et fastidieuses de caractérisation des constituants cellulaires ainsi que de leur organisation (**O'Malley & Dupré 2005, Boogard et al. 2007**). Cette thèse avait pour objectif de caractériser l'origine et la fonction d'une nouvelle



famille de protéines riches en cystéine issues d'un métatranscriptome eucaryote de sol. Elle s'inscrit dans l'ère « post-génomique » qui consiste à exploiter les données massives obtenues par les technologies de séquençage pour comprendre le fonctionnement du monde vivant à tous les niveaux d'organisation. Nous pensons qu'un des enjeux majeurs de l'écologie microbienne et de l'écotoxicologie microbienne sera de faire la synthèse entre les concepts et les approches « descendantes » de la génomique environnementale, étudiant les écosystèmes et les communautés d'organismes qui leur sont associées (**Prosser 2015, Jansson & Baker 2016**) avec les approches « ascendantes » de la biologie des systèmes étudiant les cellules et les communautés de molécules qui leur sont associées (**Kitano 2002, O'Malley & Dupré 2005, Boogard et al. 2007**) afin de relier les modifications des réseaux métaboliques, les modifications des assemblages des communautés et les grandes perturbations de l'anthropocène entre elles au travers de l'étude "des individus cellulaires".





Remerciements

La recherche scientifique est pour moi une formidable aventure humaine qui vise à entreprendre collectivement la tâche dantesque de "comprendre" l'univers grâce à la mise en place de multiples manières de faire (méthodes scientifiques) et manières de voir (théories, concepts, objets). Elle est basée, ni complètement sur des moyens financiers, ni complètement sur des moyens techniques, mais surtout sur la capacité de ses acteurs à croiser leurs regards, interagir ensemble, sur plusieurs générations. La recherche scientifique est une institution qui dépend avant tout du facteur humain. Aussi, je tenais en tout premier lieu à remercier toutes les personnes ayant contribué à l'acquisition, la critique et la diffusion des savoirs au cours des générations qui m'ont précédé et me font me sentir comme un « nain sur les épaules de géants ». A titre personnel, ces personnes m'ont permis de connaître toutes les choses que je connais et donc d'être celui que je suis.

A toutes les personnes qui m'ont permis de développer ma réflexivité, je dis « merci » : Alain Chalmers pour « Qu' est-ce que la Science ? » (La Découverte, 1990), Leo Coutellec pour « La science au pluriel » (Quae, 2015), Bruno Latour pour « chronique d'un amoureux des sciences » (Broché, 2006), Michel Serre pour ses idées sur le « Grand Récit », Kant pour ses critiques, Socrate et Platon pour la philosophie pratique, Christophe Bonneuil et Jean Baptiste Fressoz pour « l'Événement Anthropocène » (Seuil, 2013), Stephen Hawking pour « une brève histoire du temps » (Poche, 2007), Pierre Bourdieu pour ses idées sur le « capital culturel », Jean-Jacques Kupiec pour « ni dieu, ni gènes » (Poche, 2003) , Serge Halimi pour la gestion du Monde Diplomatique et le documentaire « les nouveaux chiens de gardes », Howard Gardner pour sa théorie sur « les intelligences multiples », Annie Thébaud-Mony pour « la science asservie » (La Découverte, 2014) ; Stéphane Horel pour « Intoxication » (la Découverte, 2015), Dominique Pestre pour « Le gouvernement des technosciences » (La Découverte, 2014) et les « Vulgarizateurs » ainsi que les

collègues de l'Université Populaire de Lyon pour la démocratisation des connaissances scientifiques.

Je remercie aussi toutes les personnes avec qui j'ai pu quotidiennement travailler, rire, discuter, partager, et qui ont fait que les heures passaient comme des secondes.

Anna, Carlès, Jordi, Katia, Maria, Mercè, Oscar, Quim, Sarah, Sébas, Selene et Weiyu en Espagne.

Anne, B.M., Blasko, Christophe, Christeline et Jean, Desislava, Doudou, Eric et Marine, Kevin R., Magali, Manon, Sil et Sou, So et Tom(la guinche), Sophie A., Stéphanie G. à Lyon et ses alentours.

Tous mes amis de Lorraine : Alice et Mathieu, Capucine, Carole et Anthony, Charlotte, Damien, Denis, Elodie, Guillaume, Hermance, Jennifer et Pierre, Julie, Marie et Rémi.

Au Laboratoire d'Ecologie Microbienne : Alessandro, Audrey, Babeth, Corentin, Daniel, Danis, David (, Elodie, Guillaume S., Guillaume G., Hasna, Isaline, Jordan VAC (Mr. Le Représentant). Jordan VAL., Juliana, Julien, Laura C., Laura R., Lise, Marie (Little Miss Sunshine), Morgane, Yoann BM., Quentin (Duplay-sir), Marine V., Thiebault, Sébastien, Solène, Salomon, Yvan, William.

Toute l'équipe du bâtiment Nétien qui m'a accueilli les bras ouverts pour mon ATER : Anne, Carine, Caroline, Delphine, Didier, Emily, Florian, Jordan, Laura, Veronica

Bâtiment LWOFF : Barthi, Claire VM., Claire V., Claudia, Batman, Delphine, Guillaume G. (mon négatif), Guillaume M. (El Publicator), Jeanne (dorée comme les champs de blé), Laurent, Martino, Margot, Mylène (tati Goni), Marie-André, Patricia (Chef), Patrick M., Patrick P., Rajiv, Stéphanie, Sudhakar, Van et Yoann S. (Yoyo)



Je remercie ma tutrice Agnès Rodrigue, l'école doctorale, l'association DocE₂M₂ et les membres de mon comité de thèse d'avoir accepté de discuter avec moi de mon projet de thèse et fait en sorte que ma thèse se déroule bien.

Je remercie les membres du jury d'avoir accepté d'évaluer mon travail.

Je remercie en particulier mes encadrants de thèse, pour leur patience, leur aide et leurs conseils, Laurence et Roland, qui m'ont permis d'avoir une excellente expérience dans la recherche et d'être paré à toute éventualité pour les prochaines étapes.

Je remercie aussi les institutions qui ont financé ma thèse directement ou indirectement :

- L'Université Claude Bernard Lyon 1, l'école doctorale E₂M₂ et le ministère de la recherche pour mon salaire pendant 3 ans.
- Le Laboratoire d'Ecologie Microbienne et l'équipe 2 (avec leurs tutelles et financeurs) pour les locaux et le matériel de recherche.
- La Région Rhône-Alpes pour la bourse CMIRA et l'école doctorale E₂M₂ pour les bourses de mobilité.

A toute ma famille : ma sœur, son mari, ma nièce, mon frère, son amie et ma mère, pour m'avoir soutenu et encouragé pendant ces trois années, et permis d'arriver jusque-là, je dis « merci » et je m'excuse aussi de ne pas avoir été très disponible.

A Hélène, qui m'a permis de trouver ma voie, je ne dirai jamais assez « merci ».

Je remercie Jésus, Yahvé, Mouhammad, Bouddha, Odin, Zeus, Shiva, Osiris, Amateratsu, Raël, Tom Cruise, les mânes de mes ancêtres, La Science,... (Juste au cas où).

A tous les petits riens qui font du quotidien un grand tout, je dis « merci » : Les pauses gouter, ma théière en fonte, ma bicyclette Ned, le « BouleChévik »,



les conférences en podcast, le culturisme cultivé, les thés verts coréens, mon « enwoolement » d'hiver, Olympe et Atlas, les diverses expériences artistiques, les films, les séries, mes playlists deezer, l'électricité du verbe de Feu ! Chatterton, le groove de D'Angelo, le swing de Marsalis et la chaleur de Ray Charles.

Et je dédie cette thèse à Silvia Atrian, qui m'a supervisé lors de mon séjour en Espagne ; à Florence Hélène Tran qui m'a égayé chaque jour au laboratoire ; et à mon père qui m'a « supervisé », égayé et aimé la majeure partie de ma vie ; et qui ne pourront plus lire ces pages. Leurs « souvenirs en moi luisent comme un ostensor ».





Liste des figures et tableaux

Introduction Générale

Figure A : Diagrammes conceptuels du système Terre.....	4
---	---

Chapitre 1

Figure I.1 Biodiversité du sol.....	13
Figure I.2 : Classification périodique des éléments de Mendeleïv	17
Figure I.3 : Diagramme représentant l'abondance des éléments métalliques essentiels dans la croûte terrestre en $\log_{10}(\text{ppm})$	19
Figure I.4 : Compartimentation cellulaire de certaines espèces métalliques	23
Figure I.5. Sélectivité des métaux	27
Figure I.6. Schéma simplifié des mécanismes de résistance aux métaux dans le milieu extracellulaire, dans les cellules et dans les endomembranes.....	33
Figure I.7 : Caractéristiques de la structure primaire des MT à partir de séquences protéiques présentes dans Swiss-Prot	41
Figure I.8 : Séquences primaires de métallothionéines illustrant l'organisation ou non des motifs de Cys en modules.....	43
Figure I.9 : Diversité des structures tertiaires de métallothionéines liées aux métaux.....	47
Figure I.10 : Distribution des métallothionéines au sein de l'arbre de vie	51
Figure I.11 : Trois formes de protéines sont trouvées dans les cellules	59
Figure I.12 : Classifications des Métallothionéines	65
Figure I.13. Arbre phylogénétique (PhyML) des séquences de la famille des CRP	77

180

Chapitre 2

Figure II.1 : Schéma de la stratégie expérimentale	81
Figure II.2 : Absorbances des éluats après la dernière étape de chromatographie d'exclusion stérique	97

Chapitre 3

Figures

Figure III.1 : Schéma de la stratégie expérimentale.....	104
Figure III.2 : Diagramme représentant les différentes teneurs en Cd des microcosmes mesurées par ICP-MS en fonction de la teneur théorique en Cd (mg de Cd ajouté/kg de sol sec)	114
Figure III.3 : Spécificité des couples d'amorces conçus pour amplifier les gènes <i>emt</i> des sous familles 1 (amorces EMT1F/R), 3 (EMT3F/R), 4 (EMT4F/R) et 5 (EMT5F/R).....	120
Figure III.4 : Effet positif de la protéine T4 Gene 32 sur la capacité à amplifier un fragment interne du gène <i>emt1</i> (amorces EMT1F/R).....	122
Figure III.5 : Amplification du gène <i>emt1</i> par qPCR	124
Figure III.6 : Amplification du gène <i>emt1</i> par ddPCR	126
Figure III.7 : Coefficients de variation (%) des mesures de qPCR et de ddPCR.....	128

Tableaux

Tableau III.1 : Caractéristiques du sol utilisé dans cette étude prélevé sur le site de Lommel (51° 14' N, 5° 15' E)	104
Tableau III.2 : Amorces et sonde utilisées en qPCR et ddPCR	108

Chapitre 4

Figure IV.1 : Stratégies pour obtenir des régions bordantes à partir d'une séquence nucléotidique connue	134
Figure IV.2 : Protocole expérimental de la capture de gènes	140
Figure IV.3 : Visualisation de la taille des extraits d'ADN par électrophorèse en capillaire	150
Figure IV.4 : Vérification de l'enrichissement en séquences cibles <i>emt</i> après les captures A et B par PCR nichée.....	152
Figure IV.5 : Evaluation de l'enrichissement en séquences <i>emt</i> après captures A (A.) et B (B.) par PCR en utilisant les amorces EMT1F/R spécifiques de la sous-famille 1	154
Figure IV.6 : Séquences <i>emt</i> obtenues après la capture B sur des extraits d'ADN de sol de Lommel	156





Annexes IV.1

Séquences et adaptateurs

Tableau 1 : Séquences des sondes ARN biotinylées utilisées pour la capture ciblée A

Nom	Séquence (5'-3')	Taille (pb)	Dégénérescence	Sous-familles EMT cible
CA.S1	RTACAAACATYATATCAAAAATGACTACTTGTGTGTTGCAAGRTTCYGGAG	50	15	1
CA.S2	ATCTCGTTGCATCAGAGTGCAGRCGGAGGCATGCAAGATATGTGGAT	50	1	1
CA.S3	GAAAGAAGGTGGTGAAGGRAAAGGTTGTACCACATCCWSAGTGCAAATGT	50	7	1
CA.S4	AGTGTGAAAAGTGTGAGTGYAAGCATGAGAAAAGTGAAGTGATCWAGAYKG	50	15	1
CA.S5	GGGGRCCGATCGATCARATCATTTGATTATCAACACTAATTCAAACATTA	50	3	5
CA.S6	TGTTGTCCAAAGTCTTGCTGTGTCAACCGAGTGCAAGACAGARACCTGTCA	50	1	5
CA.S7	TGTGTGTCCCAATCGAAGCATGAAGCAACTGAGGGTGCCGATGTGACTAA	50	0	5
CA.S8	CTTTGGAGGGAAGTGTGCGCATACAAAGTGTGGATGTCCCTTGGTGGTT	50	0	5
CA.S9	AGTGAACGTCAAGRCTCGGGTGCAACTTGTCTAATTTATAATTTAAT	50	1	5
CA.S10	CACATGCTGCCCAARGTCATGTGTCCCACTGAACGCAAGACCAAAATCT	50	1	7
CA.S11	CAGTTGAARGCAGAAGCAGCCACTGAAGATGGGAGCGTCAARGCAGAAGA	50	3	7
CA.S12	CACGAACGTGCATGTGGATGTCCCTTGATAGTCATGTGTAGGTGTGGGACA	50	0	7
CA.S13	CGTGACTAGCGAGWRCCMARTCTGGCGTGGTACAATAATTTAATAAAAA	50	15	7
CA.S14	ACCGRKKGAACAGATCATTTATACATCAACCGACTTTTATAAATCAATCA	50	7	3
CA.S15	CACATGCTGTCCAAAGTCTCGTTGTGTCAACCGAGTGCAATACAGAAATAT	50	0	3
CA.S16	GAACCACAGAGTCTCGCACAAAATGTGGATGTCCCTTGGTAGTGATGTGT	50	0	3
CA.S17	AAGAGCGAAACCAAGACAGAATGTRCCAAGTGTCCCAAGGTGAGTAGAT	50	1	3
CA.S18	CGAGCTCGAGTGTAGGTTAGATTAATATAATAAAGCAAATCAATAATCAC	50	0	3
CA.S19	ACTTTAATTCAAATATCAATCAAAATGGCTACTTGTGTGCAAAATTC	50	0	4
CA.S20	ATGTGAATGTGCATCCACAGTCTTGAGASTTCYGCACWACGTAACRCAT	50	15	4
CA.S21	ATGTCCTTTGATAGTGTGTGTGGATGTGGAGCAGAGTGTGAAAAGTGGA	50	1	4
CA.S22	GAACGCCAAGCTCTAGTACAATAAATTTAACTAAATAAATGTTTTAT	50	0	4
CA.S23	ATTCAAATATCAATGCMAAAATGGCTACTTGTGTGCAAAATTCAGAA	50	1	4
CA.S24	TACATGTTGCCCAAGWCTTGTRYCCCACTGGAAGCAAGRCAGAACTT	50	15	4
CA.S25	TGTGTATCCSAGTCAAAACAAGAAGCAGGCACTGAAGCTAAGGRCAYTAA	50	7	4
CA.S26	ATGTGGATGTCTTTGATAGTGTGTGTGGATGTGGAGCAGARTGTGAAA	50	1	4
CA.S27	GTGATTGAACSCCAARCTCTAGTACAATAYAATTTAACTAAATAARTGG	50	15	4
CA.S28	AAAGTACCAGAAAATGTTAAGAGCGAGGCATGTGTCCACCTACATGCTG	50	0	4
CA.S29	TGTGTATCCCAAGTCAAAACAAGAAGCAGGCACTGAAGCTAAGGGCATTAA	50	0	4
CA.S30	CATTAAGAAGAGAGGGCTTGACCCGATCCACAATGTGAATGTGCATCTA	50	0	4
CA.S31	ACCGACCTGTTGTCCAAAGACTTGTGTCCCACTGGAAGCAAGACAGAAA	50	0	4
CA.S32	TTGTCCTTGCTGCGTATCCCAAGTCAAAACAAGAAGCAGGCACTGAAGCTA	50	0	4
CA.S33	TCAAAACAAGAAGCAGGCACTGAAGCTAAGGACATTAAGAAGAGAAAGGACT	50	0	4
CA.S34	TTAAAAGAGAAGGACTTGACCCGATCCACAATGTGAATGTGCATCCACA	50	0	4
CA.S35	AATGTGAATGTGCATCCACAGTCTTGAAGAGTTCGACTACGTAAGTCA	50	0	4
CA.S36	GGGGGKGCCCATCGAATAGATCATTGCAACAKTGAAATTAATTYAACAT	50	7	2
CA.S37	KCTGTGTTCCCGACTGTAAAGACGGAGRCATGCAAAAAGTGTGRRGTGT	50	15	2
CA.S38	GAAGGACMTRCATCACATCCACAGTGTGAATGTGCAYCCACTGYTTTGGA	50	15	2
CA.S39	ATGTTGATAAACGAACCAAGTGTGGATGTCTCTGATAGTTGTRTGTGGY	50	3	2
CA.S40	ARTGAGTACCGAGTTRGAGTGTGGGAGTARGTAATTTAATAAATGATTT	50	7	2

Tableau 2 : Séquences des adaptateurs utilisés pour les captures ciblées A et B

Nom	Séquence (5'-3')	Taille (pb)
T7	GGATTCTAATACGACTCACTATAGGG (Denonfoux et al. 2013)	26
A	ATCGCACCAGCGTGT (Denonfoux et al. 2013)	15
B	ACTGCGGCTCCTCA (Denonfoux et al. 2013)	14

Tableau 3 : Séquences des sondes ARN biotinylées utilisées pour la capture ciblée B

[illegible]



Annexes IV.2

Amorces et PCR

Tableau 1 : Séquences, température d'hybridation et références des amorces utilisées dans cette étude

PCR	Nom	Séquence (5'-3')	Tm (°C)	Références
PCR 1	T7-A	GGATTCTAATACGACTCACTATAGGGATCGCACCAGCGTGT	58	Denonfoux et al. 2013
	iB	CACTGCGGCTCCTCA	58	Denonfoux et al. 2013
PCR 2	T7NMTF	GGATTCTAATACGACTCACTATAGGGAATGRCTACTTGTGTGTC	56	Cette étude
	NMTR	YTCATGYTTRCASTYRCAC	56	Cette étude
	T7A12A7A1F	GGATTCTAATACGACTCACTATAGGGTGAGCATGTGYCCACCGACCTGT	56	Cette étude
	A12A7A1R	GGARATCCACACTTGGTTCGTTAT	56	Cette étude
	T7A6F	GGATTCTAATACGACTCACTATAGGCGAGTGAGGTTTGTGTCCACCCAC	56	Cette étude
	A6R	TCTATCAGAGGAATCCACACGCTGT	56	Cette étude
	T7Lom6F	GGATTCTAATACGACTCACTATAGGAAGATCCGGAGAATGTCACGAGTG	56	Cette étude
	Lom6R	CGCTGTGTCTGCTGACTCGGTGCCA	53	Cette étude
	T7C4F	GGATTCTAATACGACTCACTATAGGCGAAGTGGAGGTTGCTTGTCCACC	53	Cette étude
	C4R	TCTATCAAGGGACATCCACACGCTG	53	Cette étude
	T7Lom9C5A8B1F	GGATTCTAATACGACTCACTATAGGCTGCCACCTACATGTTGCCCAAAG	56	Cette étude
	Lom9C5A8B1R	TCTTCAAGACTGTGGATGCACATTC	56	Cette étude
	T7SF5F	GGATTCTAATACGACTCACTATAGGTGTACGAGTGAGACGTGCTGCCAC	56	Cette étude
	SF5R	ATATCCACACITTTGTACGTGCGTCA	56	Cette étude
PCR 3	TiA	CCATCTCATCCCTGCGTGTG	58	Denonfoux et al. 2013
	TiB	CCTATCCCCTGTGTGCCTTG	58	Denonfoux et al. 2013
PCR 4	EMTF(NMTF)	AATGRCTACTTGTGTGTC	56	Lehembre et al. 2013
	EMTR(NMTR)	YTCATGYTTRCASTYRCAC	56	Lehembre et al. 2013
	EMTR3(NMTR3)	CYAYCARRGGACATCCACA	56	Lehembre et al. 2013
PCR 5	EMT1F	CGGAGAATGTCACGAGTGAGGTTTGCTG	58	Cette étude
	EMT1R	ACTTGTCTCCACACTCACACTTGGCC	58	Cette étude
PCR 6	M13F (-20)	GTAAAACGACGGCCAG	50-52	Messing, 1983
	M13R	CAGGAAACAGCTATGAC	50-52	Messing, 1984

Tableau 2 : Conditions de PCR réalisées pour cette étude

PCR 1	Kit Platinum® Taq DNA Polymerase High Fidelity							Tampon	
	Volume final	Amorce T7-A	Amorce B	dNTP	BSA	MgSO4	Taq Platinum®		
	50 µL	0,2 µM	0,2 µM	0,2 mM	-	2 mM	0,02 U/µL		
	Etapes	Dénaturation initiale	Dénaturation	Hybridation	Elongation	Nombre de Cycles	Elongation finale		
	Température (°C)	94	94	58	68	35	68		
	Temps	2 min	30 s	30 s	20 s		5 min		
PCR 2	Kit Platinum® Taq DNA Polymerase High Fidelity							Tampon	
	Volume final	Amorces T7- F	Amorces R	dNTP	BSA	MgSO4	Taq Platinum®		
	25 µL	0,2 µM	0,2 µM	0,2 mM	-	2 mM	0,02 U/µL		
	Etapes	Dénaturation initiale	Dénaturation	Hybridation	Elongation	Nombre de Cycles	Elongation finale		
	Température (°C)	94	94	(53,56)	72	30	72		
	Temps	10 min	1 min	1 min	1 min		10 min		
PCR 3	Q5® High-Fidelity DNA Polymerase							Tampon	GC Enhancer
	Volume final	Amorce TiA	Amorce TiB	dNTP	BSA	T4 Gene 32 Portein	Q5 DNA Polymerase		
	50 µL	0,4 µM	0,4 µM	0,6 mM	0,25 mg/mL	0,25 mg/mL	0,02 U/µL		
	Etapes	Dénaturation initiale	Dénaturation	Hybridation	Elongation	Nombre de Cycles	Elongation finale		
	Température (°C)	98	98	58	72	25	72		
	Temps	3 min	30 s	30 s	6 min		10 min		
PCR 4	Taq DNA Polymerase, recombinant, invitrogen							Tampon	
	Volume final	Amorce EMTF	Amorce EMTR/R3	dNTP	BSA	MgCl2	Taq DNA Polymerase		
	25 µL	1 µM	1 µM	0,2 mM	0,25 mg/mL	1,5 mM	0,02 U/µL		
	Etapes	Dénaturation initiale	Dénaturation	Hybridation	Elongation	Nombre de Cycles	Elongation finale		
	Température (°C)	95	95	55	72	35	72		
	Temps	3 min	45 s	45 s	1 min		10 min		
PCR 5	Taq DNA Polymerase, recombinant, invitrogen							Tampon	
	Volume final	Amorce EMT1F	Amorce EMT1R	dNTP	BSA	MgCl2	Taq DNA Polymerase		
	25 µL	0,8 µM	0,8 µM	0,15 mM	0,25 mg/mL	1,5 mM	0,02 U/µL		
	Etapes	Dénaturation initiale	Dénaturation	Hybridation	Elongation	Nombre de Cycles	Elongation finale		
	Température (°C)	94	94	58	72	40	72		
	Temps	10 min	1 min	1 min	30 s		10 min		
PCR 6	Taq DNA Polymerase, recombinant, invitrogen							Tampon	
	Volume final	Amorce M13F	Amorce M13R	dNTP	BSA	MgCl2	Taq DNA Polymerase		
	25 µL	0,2 µM	0,2 µM	0,15 mM	0,25 mg/mL	1,5 mM	0,02 U/µL		
	Etapes	Dénaturation initiale	Dénaturation	Hybridation	Elongation	Nombre de Cycles	Elongation finale		
	Température (°C)	94	94	55	72	30	72		
	Temps	10 min	1 min	1 min	8 min		10 min		





Annexes IV.3

Solutions préparées pour la capture ciblée

- SSC
- SSPE 20X Solution-Sodium Phosphate- EDTA, pH 7.4 ± 0.2 (avec NaOH)
 - $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ 0.2 M (= 0.2 mol/L)
 $M_m = 156.04 \text{ g/mol}$ ($n = m/M$ donc $m = n \times M$)
 $m = 0.2 \text{ mol} \times 156.04 \text{ g/mol} = 31.21 \text{ g pour 1 L}$
 - NaCl 3M (= 3 mol/L)
 $M_m = 58.4 \text{ g/mol}$
 $M = 3 \text{ mol} \times 58.4 \text{ g/mol} = 175.2 \text{ g pour 1L}$
 - $\text{Na}_2 \text{EDTA}$ 20mM (0.02 mol/L)
 $M_m = 372.24 \text{ g/mol}$
 $M = 0.02 \text{ mol} \times 372.24 \text{ g/mol} = 7.44 \text{ g pour 1L}$
- Denhardt's 100X
 - ▶ Filtrer pour stériliser
 - ▶ Aliquoter et stocker à -20°C
 - BSA 2%
2 g pour 100 mL
 - Ficoll 400 2%
2 g pour 100 mL
 - PVP 2%
2 g pour 100 mL
- NaCl 5M
 $M_m = 58.4 \text{ g/mol}$

$$M = 5 \text{ mol} \times 58.4 \text{ g/mol} = 292 \text{ g pour 1L}$$

- Tris HCl 1M (1 mol/L), pH 7.5

$$Mm = 120.14 \text{ g/mol}$$

$$M = 1 \text{ mol} \times 120.14 \text{ g/mol} = 120.14 \text{ g pour 1L}$$

- Na₂ EDTA 0.5 M (0.5 mol/L), pH 8

- Na₂ EDTA 0.5 M (0.5 mol/L), pH 8

$$Mm = 372.24 \text{ g/mol}$$

$$M = 0.5 \text{ mol} \times 372.24 \text{ g/mol} = 186.12 \text{ g pour 1L}$$

Préparation des solutions tampons

Solutions	Réactifs	Concentration	Quantité
Tampon 1 → 50 mL			
	NaCl	1M	10 mL de NaCl 5M
	Tris-HCl	10mM	500 µL de Tris HCl 1M, pH 7,5
	EDTA	1mM	100 µL d'EDTA 0,5M, pH 8
	H ₂ O		39,4 mL
Tampon 2 → 50 mL			
	SSC 1x		2,5 mL de SSC 20x
	SDS 0,1%		500 µL de SDS 10%
	H ₂ O		47 mL
Tampon 3 → 50 mL			
	SSC 0,1x		250 µL de SSC 20x
	SDS 0,1%		500 µL de SDS 10%
Tampon Hybridation → 2mL			
			1000 µL SSPE 20X
			200µL Denhardt's 100X
			40µL EDTA 0.5M, pH8
			40µL SDS 10%
			720 µL H ₂ O





Références Bibliographiques

1. Adhikari, K. & Hartemink, A. E. Linking soils to ecosystem services - A global review. *Geoderma* **262**, 101–111 (2016).
2. Adl, S. M. *et al.* The Revised Classification of Eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–493 (2012).
3. Alloway, B. J. *Heavy Metals in Soils*. (Environmental Pollution, 2013).
4. Almario, J., Moënné-Loccoz, Y. & Muller, D. Monitoring of the relation between 2,4-diacetylphloroglucinol-producing *Pseudomonas* and *Thielaviopsis basicola* populations by real-time PCR in tobacco black root-rot suppressive and conducive soils. *Soil Biol. Biochem.* **57**, 144–155 (2013).
5. Amorim, M. J. B., Pereira, C., Soares, A. M. V. M. & Scott-Fordsmand, J. J. Does long term low impact stress cause population extinction? *Environ. Pollut.* 1–10 (2016).
6. Anderson, I. C. & Cairney, J. W. G. Minireview Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. **6**, 769–779 (2004).
7. Andrews, G. K. Regulation of Metallothionein Gene Expression by Oxidative Stress and Metal Ions. *Biochem. Pharmacol.* **59**, 95–104 (2000).
8. Arndt, N. & Ganino, C. *Metals and Society, an Introduction to Economic Geology*. (Springer Science & Business Media, 2012).
9. Atrian, S. & Capdevila, M. Metallothionein-protein interactions. *Biomol. Concepts* **4**, 143–160 (2013).
10. Bælum, J. *et al.* Direct analysis of *tfdA* gene expression by indigenous bacteria in phenoxy acid amended agricultural soil. *ISME J.* **2**, 677–87 (2008).
11. Bailly, J. *et al.* Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISME J.* **1**, 632–642 (2007).
12. Baker-Austin, C., Wright, M. S., Stepanauskas, R. & McArthur, J. V. Co-selection of antibiotic and metal resistance. *Trends Microbiol.* **14**, 176–182 (2006).
13. Barabási, A. L. & Oltvai, Z. N. Network Biology: Understanding the Cell's Functional Organization. *Nat. Rev.* **5**, 101–113 (2004).
14. Barbi, F. *et al.* Soil Biology & Biochemistry Tree species select diverse soil fungal communities expressing different sets of lignocellulolytic enzyme-encoding genes. *Soil Biol. Biochem.* **100**, 149–159 (2016).
15. Barnosky, A. D. *et al.* Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
16. Barrios, E. Soil biota, ecosystem services and land productivity. *Ecol. Econ.* **64**, 269–285 (2007).
17. Bates, S. T. *et al.* Global biogeography of highly diverse protistan communities in soil. *ISME J.* **7**, 652–659 (2012).
18. Bell, S. G. & Vallee, B. L. The Metallothionein/Thionein System: An Oxidoreductive Metabolic Zinc Link. *ChemBioChem - A Eur. J. Chem. Biol.* **10**, 55–62 (2009).
19. Bellion, M., Courbot, M., Jacob, C., Blaudez, D. & Chalot, M. Extracellular and cellular mechanisms sustaining metal tolerance in ectomycorrhizal fungi. *FEMS Microbiol. Lett.* **254**, 173–181 (2006).
20. Bender, S. F., Wagg, C. & van der Heijden, M. G. A. An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability. *Trends Ecol. Evol.* **31**, 440–452 (2016).
21. Binz, P.-A. & Kägi, J. H. R. in *Metallothionein IV* (ed. Klaassen, C. D.) 7–14 (Springer Basel AG, 1999).
22. Biraben, J. L'évolution du nombre des hommes. *Popul. Sociétés* **394**, 1–4 (2003).
23. Blindauer, C. A. Bacterial metallothioneins: Past, present, and questions for the future. *J. Biol. Inorg. Chem.* **16**, 1011–1024 (2011).
24. Blindauer, C. A. Metallothioneins with unusual residues: Histidines as modulators of zinc affinity and reactivity. *J. Inorg. Biochem.* **102**, 507–521 (2008).

25. Blindauer, C. A. & Leszczyszyn, O. I. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. *Nat. Prod. Rep.* **27**, 720–741 (2010).
26. Bofill, R., Capdevila, M. & Atrian, S. Independent metal-binding features of recombinant metallothioneins convergently draw a step gradation between Zn- and Cu- thioneins. *Metallomics* **229**, 229–234 (2009).
27. Boogard, F. C. *Systems Biology, Philosophical Foundations*. (Elsevier, 2007).
28. Botella, H., Stadthagen, G., Lugo-Villarino, G., de Chastellier, C. & Neyrolles, O. Metallobiology of host-pathogen interactions: An intoxicating new insight. *Trends Microbiol.* **20**, 106–112 (2012).
29. Bragalini, C. *et al.* Solution hybrid selection capture for the recovery of functional full-length eukaryotic cDNAs from complex environmental samples. *DNA Res.* **21**, 685–694 (2014).
30. Brauns, T. C. & Goos, M. Micromanipulation of single cells from tissue imprints is an alternative to laser-assisted microdissection. *J. Cutan. Pathol.* **32**, 461–466 (2005).
31. Brehm-stecher, B. F. & Johnson, E. A. Single-Cell Microbiology: Tools, Technologies, and Applications Single-. *Microbiol. Mol. Biol. Rev.* **68**, 538–559 (2004).
32. Briones, M. J. I. Soil fauna and soil functions: a jigsaw puzzle. *Front. Environ. Sci.* **2**, 1–22 (2014).
33. Bustin, S. A. *et al.* The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.* **55**, 611–622 (2009).
34. Bustin, S. A. & Nolan, T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* **15**, 155–66 (2004).
35. Buttigieg, P. L. & Ramette, A. A guide to statistical analysis in microbial ecology: A community-focused, living review of multivariate data analyses. *FEMS Microbiol. Ecol.* **90**, 543–550 (2014).
36. Calow, P. & Forbes, V. E. Does Ecotoxicology Inform Ecological Risk Assessment? *Environ. Sci. Technol.* **37**, 146A–151A (2003).
37. Capasso, C. *et al.* Phylogenetic Divergence of Fish and Mammalian Metallothionein: Relationships with Structural Diversification and Organismal Temperature. *J. Mol. Evol.* **57**, 250–257 (2003).
38. Capdevila, M., Bofill, R., Palacios, O. & Atrian, S. State-of-the-art of metallothioneins at the beginning of the 21st century. *Coord. Chem. Rev.* **256**, 46–62 (2012).
39. Capdevila, M. & Atrian, S. Metallothionein protein evolution: A miniassay. *J. Biol. Inorg. Chem.* **16**, 977–989 (2011).
40. Cavé, L. *et al.* Efficiency and sensitivity of the digital droplet PCR for the quantification of antibiotic resistance genes in soils and organic residues. *Appl. Microbiol. Biotechnol.* **100**, 10597–10608 (2016).
41. Chan, J., Huang, Z. & Merrifield, M. Studies of metal binding reactions in metallothioneins by spectroscopic, molecular biology, and molecular modeling techniques. *Coord. Chem. Rev.* **233–234**, 319–339 (2002).
42. Chapman, P. M., Fairbrother, A. & Brown, D. A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environ. Toxicol. Chem.* **17**, 99–108 (1998).
43. Cherian, G. & Apostolova, M. Nuclear Localization of Metallothionein during Cell Proliferation and Differentiation. *Cell. Mol. Biol.* **46**, 347–356 (2000).
44. Clemens, S. Molecular mechanisms of plant metal tolerance and homeostasis. *Planta* **212**, 475–486 (2001).
45. Costello, L. C., Guan, Z., Franklin, R. B. & Feng, P. Metallothionein can function as a chaperone for zinc uptake transport into prostate and liver mitochondria. *J. Inorg. Biochem.* **98**, 664–666 (2004).
46. Coûteaux, M. M. & Darbyshire, J. F. Functional diversity amongst soil protozoa. *Appl. Soil Ecol.* **10**, 229–237 (1998).
47. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, (2012).

48. De Groot, R. S., Wilson, M. A. & Boumans, R. M. J. A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecol. Econ.* **41**, 393–408 (2002).
49. Decaria, L., Bertini, I. & Williams, R. J. P. Copper proteomes, phylogenetics and evolution. *Metallomics* **3**, 56–60 (2011).
50. Delmont, T. O. *et al.* Accessing the soil metagenome for studies of microbial diversity. *Appl. Environ. Microbiol.* **77**, 1315–1324 (2011).
51. Denonfoux, J. *et al.* Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Res.* **20**, 185–196 (2013).
52. Ding, C. *et al.* *Cryptococcus neoformans* copper detoxification machinery is critical for fungal virulence. *Cell host Microbe* **13**, 265–276 (2014).
53. Domènech, J. *et al.* Plant metallothionein domains: functional insight into physiological metal binding and protein folding. *Biochimie* **88**, 583–593 (2006).
54. Dondero, F. *et al.* Quantitative PCR analysis of two molluscan metallothionein genes unveils differential expression and regulation. *Gene* **345**, 259–270 (2005).
55. Dong, D., Yan, A., Liu, H., Zhang, X. & Xu, Y. Removal of humic substances from soil DNA using aluminium sulfate. *J. Microbiol. Methods* **66**, 217–222 (2006).
56. Dorak, T. *Real Time PCR*. (Taylors & Francis Group, 2006).
57. Dudev, T. & Lim, C. Metal binding affinity and selectivity in metalloproteins: insights from computational studies. *Annu. Rev. Biophys.* **37**, 97–116 (2008).
58. Duncan, K. E. R. & Stillman, M. J. Metal-dependent protein folding: Metallation of metallothionein. *J. Inorg. Biochem.* **100**, 2101–2107 (2006).
59. Dupont, C. L., Butcher, A., Valas, R. E., Bourne, P. E. & Caetano-Anollés, G. History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proc. Natl. Acad. Sci.* **107**, 10567–10572 (2010).
60. Edwards, K. J. & Bazylinski, D. A. Intracellular minerals and metal deposits in prokaryotes. *Geobiology* **6**, 309–317 (2008).
61. Ehrlich, H. L. in *Ehrlich's Geomicrobiology* 55–68 (Taylor & Francis Group, LLC, 2015).
62. Emoto, T. *et al.* Roles of the conserved serines of metallothionein in cadmium binding. *Biochem. Genet.* **34**, 239–51 (1996).
63. Engel, K., Pinnell, L., Cheng, J., Charles, T. C. & Neufeld, J. D. Nonlinear electrophoresis for purification of soil DNA for metagenomics. *J. Microbiol. Methods* **88**, 35–40 (2012).
64. Espart, A. *et al.* Hints for metal-preference protein sequence determinants: Different metal binding features of the five tetrahymena thermophila metallothioneins. *Int. J. Biol. Sci.* **11**, 456–471 (2015).
65. Fajardo, C., Costa, G., Nande, M. & Martin, M. Three Functional Biomarkers for Monitoring the Nanoscale Zero-Valent Iron (nZVI)-Induced Molecular Signature on Soil Organisms. *Water. Air. Soil Pollut.* **227**, (2016).
66. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* (80-. J.) **320**, 1034–1039 (2008).
67. Filipović, V. & Raspor, B. Metallothionein and metal levels in cytosol of liver, kidney and brain in relation to growth parameters of *Mullus surmuletus* and *Liza aurata* from the Eastern Adriatic Sea. *Water Res.* **37**, 3253–3262 (2003).
68. Fraústo da Silva, J. J. R. & Williams, R. J. P. *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*. (Oxford University Press, 2001).
69. Freisinger, E. Plant MTs-long neglected members of the metallothionein superfamily. *Dalton Trans.* **9226**, 6663–6675 (2008).
70. Gadd, G. M. Heavy metal accumulation by bacteria and other microorganisms. *Experientia* **46**, 834–840 (1990).
71. Galloway, J. N. *et al.* Nitrogen cycles: Past, present, and future. *Biogeochemistry* **70**, (2004).

72. García, S., Prado, M., Dégano, R. & Domínguez, A. A copper-responsive transcription factor, CRF1, mediates copper and cadmium resistance in *Yarrowia lipolytica*. *J. Biol. Chem.* **277**, 37359–37368 (2002).
73. Gasc, C. *et al.* Capturing prokaryotic dark matter genomes. *Res. Microbiol.* **166**, 814–830 (2015).
74. Ghoshal, K. *et al.* and DNA methyltransferase synergistically activate the methylated metallothionein I promoter. *Mol. Cell. Biol.* **22**, 8302–8319 (2002).
75. Giles, N. M. *et al.* Metal and Redox Modulation of Cysteine Protein Function. **10**, 677–693 (2003).
76. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–9 (2009).
77. Gobat, J.-M., Miche, A. & Matthey, W. *Le sol vivant : bases de pédologie, biologie des sols*. (PPUR Presses Polytechniques, 2010).
78. Gold, B. *et al.* Identification of a copper-binding metallothionein in pathogenic mycobacteria. *Nat Chem Biol* **4**, 609–616 (2008).
79. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 1–16 (2016).
80. Guo, W. J., Bundithya, W. & Goldsbrough, P. B. Characterization of the Arabidopsis metallothionein gene family: Tissue-specific expression and induction during senescence and in response to copper. *New Phytol.* **159**, 369–381 (2003).
81. Gutiérrez, J. C. *et al.* Ciliate metallothioneins: Unique microbial eukaryotic heavy-metal-binder molecules. *J. Biol. Inorg. Chem.* **16**, 1025–1034 (2011).
82. Gutiérrez, J. C., Amaro, F. & Martín-González, A. From heavy metal-binders to biosensors: Ciliate metallothioneins discussed. *BioEssays* **31**, 805–816 (2009).
83. Hager, L. J. & Palmiter, R. D. Transcriptional regulation of mouse liver metallothionein-I gene by glucocorticoids. *Nature* **291**, 340–342 (1981).
84. Harnpicharnchai, P. *et al.* An efficient purification and fractionation of genomic DNA from soil by modified troughing method. *Lett. Appl. Microbiol.* **45**, 387–391 (2007).
85. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–94 (1996).
86. Heuchel, R., Radtke, F., Georgiev, O. & Stark, G. The transcription factor MTF-1 is essential for basal and heavy metal-induced metallothionein gene expression. *EMBO J.* **13**, 2870–2875 (1994).
87. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).
88. Hooda, P. S. *Trace Elements in Soils*. (John Wiley & Sons, 2010).
89. Horner, T. J., Lee, R. B. Y., Henderson, G. M. & Rickaby, R. E. M. Nonspecific uptake and homeostasis drive the oceanic cadmium cycle. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2500–5 (2013).
90. Hui, A. & De Boer, H. A. Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **84**, 4762–4766 (1987).
91. Ilie, L., Mohamadi, H., Golding, G. B. & Smyth, W. F. BOND: Basic OligoNucleotide Design. *BMC Bioinformatics* **14**, 69 (2013).
92. Imoto, A. *et al.* Metallothionein-1 isoforms and vimentin are direct PU.1 downstream target genes in leukemia cells. *J. Biol. Chem.* **285**, 10300–10309 (2010).
93. Irving, H. & Williams, R. J. P. Order of Stability of Metal Complexes. *Nature* **161**, 436–437 (1948).
94. Ishii, S., Tago, K. & Senoo, K. Single-cell analysis and isolation for microbiology and biotechnology: Methods and applications. *Appl. Microbiol. Biotechnol.* **86**, 1281–1292 (2010).
95. Iturbe-Espinoza, P. *et al.* The fungus *Tremella mesenterica* encodes the longest metallothionein currently known: Gene, protein and metal binding characterization. *PLoS One* **11**, (2016).

96. Jacobson, M. C., Charlson, R. J., Rodhe, H. & Orians, G. H. *Earth System Science : from Biogeochemical Cycles to Global Change*. (Elsevier, 2000).
97. Jansson, J. K. & Baker, E. S. A multi-omic future for microbiome studies. *Nat. Microbiol.* **1**, 16049 (2016).
98. Jiang, L. J., Maret, W. & Vallee, B. L. The glutathione redox couple modulates zinc transfer from metallothionein to zinc-depleted sorbitol dehydrogenase. *Proc. Natl. Acad. Sci.* **95**, 3483–3488 (1998).
99. Kanagawa, T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* **96**, 317–323 (2003).
100. Kaur, A. *et al.* Survival strategies of an archaeal organism to withstand stress from transition metals. *Genome Res.* 841–854 (2006).
101. Kim, D. J. . *et al.* Discovering the electronic circuit diagram of life : structural relationships among transition metal binding sites in oxidoreductases. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20120257 (2013).
102. Kim, T. G., Jeong, S. Y. & Cho, K. S. Comparison of droplet digital PCR and quantitative real-time PCR for examining population dynamics of bacteria in soil. *Appl. Microbiol. Biotechnol.* **98**, 6105–6113 (2014).
103. Kirchman, D. L. *Processes in Microbial Ecology*. (Oxford University Press, 2012).
104. Kitano, H. Systems Biology: A Brief Overview. *Science (80-.)*. **295**, 1662–1664 (2002).
105. Klaassen, C. D. *et al.* In Vitro and In Vivo Studies on the Degradation of Metallothionein. *Environ. Health Perspect.* **102**, 141–146 (1994).
106. Kotik, M. Novel genes retrieved from environmental DNA by polymerase chain reaction: Current genome-walking techniques for future metagenome applications. *J. Biotechnol.* **144**, 75–82 (2009).
107. Krezel, A. & Maret, W. Different redox states of metallothionein / thionein in biological tissue. *Biochem. J.* **402**, 551–558 (2007).
108. Landrevie, B. La Méditerranée empoisonnée. *Le Monde Diplomatique Mai*, 8 (2015).
109. Lehembre, F. Réponse adaptatives des microorganismes eucaryotes du sol aux pollutions métalliques. (Claude Bernard Lyon 1, 2009).
110. Lehembre, F. *et al.* Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. *Environ. Microbiol.* **15**, 2829–2840 (2013).
111. Lemire, J. a, Harrison, J. J. & Turner, R. J. Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nat. Rev. Microbiol.* **11**, 371–84 (2013).
112. Leoni, C., Volpicella, M., De Leo, F., Gallerani, R. & Ceci, L. R. Genome walking in eukaryotes. *FEBS J.* **278**, 3953–3977 (2011).
113. Lerch, K., Johnson, G. F., Grushoff, P. S. & Sternlieb, I. Canine hepatic lysosomal copper protein: Identification as metallothionein. *Arch. Biochem. Biophys.* **243**, 108–114 (1985).
114. Leszczyszyn, O. I., Imam, H. T. & Blindauer, C. A. Diversity and distribution of plant metallothioneins : a review of structure, properties and functions. *Metallomics* **5**, 1146–69 (2013).
115. Lima-mendez, G. *et al.* Determinants of community structure in the global plankton interactome. *Science (80-.)*. **348**, 1–10 (2015).
116. Liu, S. X. *et al.* Reconstitution of apo-superoxide dismutase by nitric oxide-induced copper transfer from metallothioneins. *Chem. Res. Toxicol.* **13**, 922–931 (2000).
117. Liu, Y. G. & Whittier, R. F. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25**, 674–681 (1995).
118. Liu, Y. *et al.* Two metallothionein genes in *Oxya chinensis*: Molecular characteristics, expression patterns and roles in heavy metal stress. *PLoS One* **9**, (2014).

119. Luis, P., Kellner, H., Martin, F. & Buscot, F. A molecular method to evaluate basidiomycete laccase gene expression in forest soils. *Geoderma* **128**, 18–27 (2005).
120. Lynes, M. A. *et al.* The physiological roles of extracellular metallothionein. *Exp. Biol. Med.* **231**, 1548–1554 (2006).
121. Macfie, S. M. & Welbourn, P. M. The cell wall as a barrier to uptake of metal ions in the unicellular green alga *Chlamydomonas reinhardtii* (Chlorophyceae). *Arch. Environ. Contam. Toxicol.* **39**, 413–419 (2000).
122. Manoharan, L., Kushwaha, S. K., Hedlund, K. & Ahrén, D. Captured metagenomics: Large-scale targeting of genes based on 'sequence capture' reveals functional diversity in soils. *DNA Res.* **22**, 451–460 (2015).
123. Maret, W. Zinc coordination environments in proteins determine zinc functions. *J. Trace Elem. Med. Biol.* **19**, 7–12 (2005).
124. Maret, W. *et al.* The ATP/metallothionein interaction: NMR and STM. *Biochemistry* **41**, 1689–1694 (2002).
125. Martin, F. *Microbial Environmental Genomics*. (Springer, 2016).
126. Martín-González, A., Borniquel, S., Díaz, S., Ortega, R. & Gutiérrez, J. C. Ultrastructural alterations in ciliated protozoa under heavy metal exposure. *Cell Biol. Int.* **29**, 119–126 (2005).
127. Matz, C. & Kjelleberg, S. Off the hook – how bacteria survive protozoan grazing. *Trends Microbiol.* **13**, 302–307 (2005).
128. McKim, J., Choudhuri, S. & Klaassen, C. In Vitro Degradation of Apo-, Zinc-, and Cadmium-Metallothionein by Cathepsins B, C, and D. *Toxicol. Appl. Pharmacol.* **116**, 117–124 (1992).
129. Mehus, A. A. *et al.* Quantitation of Human Metallothionein Isoforms: A Family of Small, Highly Conserved, Cysteine-rich Proteins. *Mol. Cell. proteomics* **13**, 1020–33 (2014).
130. Merchant, S. S. & Helmann, J. D. *Elemental Economy. Microbial Strategies for Optimizing Growth in the Face of Nutrient Limitation. Advances in Microbial Physiology* **60**, (Elsevier Ltd., 2012).
131. Merroun, M. L. *et al.* Complexation of uranium by cells and S-layer sheets of *Bacillus sphaericus* JG-A12. *Appl. Environ. Microbiol.* **71**, 5532–5543 (2005).
132. Miot, J. *et al.* Extracellular iron biomineralization by photoautotrophic iron-oxidizing bacteria. *Appl. Environ. Microbiol.* **75**, 5586–5591 (2009).
133. Miquel, G. *Rapport n° 261 sur les effets des métaux lourds sur l'environnement et la santé*.
134. Montoya, J. M., Pimm, S. L. & Sole, R. V. Ecological networks and their fragility. *Nature* **442**, 259–264 (2006).
135. Mukhopadhyay, R., Rosen, B. P., Phung, L. T. & Silver, S. Microbial arsenic: From geocycles to genes and enzymes. *FEMS Microbiol. Rev.* **26**, 311–325 (2002).
136. Müller, M. S., Scheu, S. & Jousset, A. Protozoa Drive the Dynamics of Culturable Biocontrol Bacterial Communities. *Curr. Sci.* **8**, 1–6 (2013).
137. Nebes, V. L., DeFranco, D. & Morris, S. M. Cyclic AMP induces metallothionein gene expression in rat hepatocytes but not in rat kidney. *Biochem. J.* **255**, 741–743 (1988).
138. Nesme, J. *et al.* Back to the Future of Soil Metagenomics. *Front. Microbiol.* **7**, 1–5 (2016).
139. Nicolaisen, M. H., Bøelum, J., Jacobsen, C. S. & Sørensen, J. Transcription dynamics of the functional *tfdA* gene during MCPA herbicide degradation by *Cupriavidus necator* AEO106 (pRO101) in agricultural soil. *Environ. Microbiol.* **10**, 571–579 (2008).
140. Nies, D. H. Microbial heavy-metal resistance. *Appl. Microbiol. Biotechnol.* **51**, 730–750 (1999).
141. Noguera, D. R., Wright, E. S., Camejo, P. & Yilmaz, L. S. Mathematical tools to optimize the design of oligonucleotide probes and primers. *Appl. Microbiol. Biotechnol.* **98**, 9595–9608 (2014).
142. Nriagu, J. O. Human influence on the global cycling of trace metals. *Palaeogeogr. Palaeoclimatol. Palaeoecol. (Global Planet. Chang. Sect.* **82**, 113–120 (1990).
143. Nuss, P. & Eckelman, M. J. Life cycle assessment of metals: A scientific synthesis. *PLoS One* **9**, 1–12 (2014).

144. O'Halloran, T. V. & Culotta, V. C. Metallochaperones, an intracellular shuttle service for metal ions. *J. Biol. Chem.* **275**, 25057–25060 (2000).
145. Oliver, T. H. *et al.* Biodiversity and Resilience of Ecosystem Functions. *Trends Ecol. Evol.* **30**, 673–684 (2015).
146. O'Malley, M. A. & Dupré, J. Fundamental issues in systems biology. *BioEssays* **27**, 1270–1276 (2005).
147. Orihuela, R. *et al.* Ferritin and metallothionein: dangerous liaisons. *ChemBioChem - A Eur. J. Chem. Biol.* **47**, 12155–12157 (2011).
148. Ozawa, S., Okabe, S. & Ishii, S. Specific Single-Cell Isolation of *Escherichia coli* O157 from Environmental Water Samples by Using Flow Cytometry and Fluorescence-Activated Cell Sorting. *Foodborne Pathog. Dis.* **13**, 456–461 (2016).
149. Palacios, Ò., Atrian, S. & Capdevila, M. Zn- and Cu-thioneins: A functional classification for metallothioneins? *J. Biol. Inorg. Chem.* **16**, 991–1009.
150. Palacios, Ò. *et al.* Full characterization of the Cu-, Zn-, and Cd-binding properties of CnMT1 and CnMT2, two metallothioneins of the pathogenic fungus *Cryptococcus neoformans* acting as virulence factors. *Metallomics* **6**, 279–291 (2014).
151. Palacios, Ò. *et al.* Shaping mechanisms of metal specificity in a family of metazoan metallothioneins: evolutionary differentiation of mollusc metallothioneins. *BMC Biol.* **9**, 4.
152. Palumaa, P. Copper chaperones. the concept of conformational control in the metabolism of copper. *FEBS Lett.* **587**, 1902–1910 (2013).
153. Palumaa, P., Mackay, E. A. & Vasak, M. Nonoxidative Cadmium-Dependent Dimerization of Cd7-Metallothionein from Rabbit Liver. *Biochemistry* **31**, 2181–2186 (1992).
154. Parisot, N., Denonfoux, J., Dugat-Bony, E., Peyret, P. & Peyretailade, E. KASpOD-a web service for highly specific and explorative oligonucleotide design. *Bioinformatics* **28**, 3161–3162 (2012).
155. Peršoh, D., Theuerl, S., Buscot, F. & Rambold, G. Towards a universally adaptable method for quantitative extraction of high-purity nucleic acids from soil. *J. Microbiol. Methods* **75**, 19–24 (2008).
156. Piggott, J. J., Townsend, C. R. & Matthaei, C. D. Reconceptualizing synergism and antagonism among multiple stressors. *Ecol. Evol.* **5**, 1538–1547 (2015).
157. Pinheiro, L. B. *et al.* Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* **84**, 1003–1011 (2012).
158. Pivetal, J. *et al.* Micro-magnet arrays for specific single bacterial cell positioning. *J. Magn. Magn. Mater.* **380**, 72–77 (2015).
159. Pohl, G. & Shih, M. Principle and applications of digital PCR. *Expert Rev Mol Diagn* **4**, 41–47 (2004).
160. Pozhitkov, A. *et al.* Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res.* **34**, (2006).
161. Prosser, J. I. Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. *Nat. Rev. Microbiol.* **13**, 439–46 (2015).
162. Ramade, F. *Précis d'écotoxicologie*. (Masson, 1992).
163. Raudenska, M. *et al.* Metallothionein polymorphisms in pathological processes. *Metallomics* **6**, 55–68 (2014).
164. Rebrikov, D. V. & Trofimov, D. Y. Real-time PCR: A review of approaches to data analysis. *Appl. Biochem. Microbiol.* **42**, 455–463 (2006).
165. Rico, D., Martín-González, A., Díaz, S., de Lucas, P. & Gutiérrez, J.-C. Heavy metals generate reactive oxygen species in terrestrial and aquatic ciliated protozoa. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* **149**, 90–96 (2009).
166. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

167. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
168. Rulíšek, L. & Vondrášek, J. Coordination geometries of selected transition metal ions (Co²⁺, Ni²⁺, Cu²⁺, Zn²⁺, Cd²⁺, and Hg²⁺) in metalloproteins. *J. Inorg. Biochem.* **71**, 115–127 (1998).
169. Ruyters, S., Salaets, P., Oorts, K. & Smolders, E. Copper toxicity in soils under established vineyards in Europe: A survey. *Sci. Total Environ.* **443**, 470–477 (2013).
170. Sadhu, C. & Gedamu, L. Metal-specific posttranscriptional control of human metallothionein genes. *Mol. Cell. Biol.* **9**, 5738–5741 (1989).
171. Schalk, I. J., Hannauer, M. & Braud, A. New roles for bacterial siderophores in metal transport and tolerance. *Environ. Microbiol.* **13**, 2844–2854 (2011).
172. Scherrer, B. *Biostatistique*. (Les Editions de la Chenelière, 2007).
173. Schimel, D. Terrestrial ecosystems and the carbon cycle. *Glob. Chang. Biol.* **1**, 77–91 (1995).
174. Scientific Committee on Health and Environmental Risks (SCHER). *Voluntary Risk Assessment Report on Copper and its compounds, Environmental Part*. (2009).
175. Scott, C. *et al.* Bioavailability of zinc in marine systems through time. *Nat. Geosci.* **6**, 125–128 (2013).
176. Shamsi, T. N. & Fatima, S. Metallothionein: classification, biochemical features and clinical applications. *J. Proteins Proteomics* **5**, 25–33 (2014).
177. Sies, H. Glutathione and its role in cellular functions. *Free Radic Biol Med* **27**, 916–921 (1999).
178. Smolders, E. *Risk assessment of cadmium in soil- the EU perspective. Report FOOD 21 N° 5/2002*. (2002).
179. Sonke, J. E., Hoogewerff, J. A., Van der Laan, S. R. & Vangronsveld, J. A chemical and mineralogical reconstruction of Zn-smelter emissions in the Kempen region (Belgium), based on organic pool sediment cores. *Sci. Total Environ.* **292**, 101–119 (2002).
180. Srivastava, P. & Kowshik, M. Mechanisms of metal resistance and homeostasis in Haloarchaea. *Archaea* **2013**, (2013).
181. Stangegaard, M., Dufva, I. H. & Dufva, M. Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA. *Biotechniques* **40**, 649–657 (2006).
182. Stuart, G. W., Searle, P. F., Chen, H. Y., Brinster, R. L. & Palmiter, R. D. A 12-base-pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 7318–7322 (1984).
183. Sugden, D. & De Winter, P. in *Molecular Biomethods Handbook, 2nd Edition* (2008).
184. Sutherland, D. E. K. & Stillman, M. J. Challenging conventional wisdom: single domain metallothioneins. *Metallomics* **6**, 702–28 (2014).
185. Sutherland, D. E. K., Summers, K. L. & Stillman, M. J. Modeling the Zn(II) and Cd(II) metalation mechanism in mammalian metallothionein 1a. *Biochem. Biophys. Res. Commun.* **426**, 601–607 (2012).
186. Thornton, B. & Basu, C. Real-time PCR (qPCR) primer design using free online software. *Biochem. Mol. Biol. Educ.* **39**, 145–154 (2011).
187. Trinchella, F., Riggio, M., Filosa, S., Parisi, E. & Scudiero, R. Molecular cloning and sequencing of metallothionein in squamates: New insights into the evolution of the metallothionein genes in vertebrates. *Gene* **423**, 48–56 (2008).
188. Tsujikawa, K. *et al.* Localization of metallothionein in nuclei of growing primary cultured adult rat hepatocytes. *FEBS Lett.* **283**, 239–242 (1991).
189. Vasconcelos, M. H., Tam, S. C., Beattie, J. H. & Hesketh, J. E. Evidence for differences in the post-transcriptional regulation of rat metallothionein isoforms. *Biochem. J.* **315**, 665–671 (1996).
190. Vita, N. *et al.* Bacterial cytosolic proteins with a high capacity for Cu(II) that protect against copper toxicity. *Sci. Rep.* **6**, 39065 (2016).
191. Vita, N. *et al.* A four-helix bundle stores copper for methane oxidation. *Nature* **525**, 140–143 (2015).
192. Wall, D. H., Adams, G. & Parsons, A. N. in *Global Biodiversity in a Changing Environment* 47–82 (Springer Science & Business Media, 1995).

193. Williams, R. J. P. The fundamental nature of life as a chemical system: The part played by inorganic elements. *J. Inorg. Biochem.* **88**, 241–250 (2002).
194. Williams, R. J. P. Chemical selection of elements by cells. *Coord. Chem. Rev.* **216–217**, 583–595 (2001).
195. Yadav, R. K. *et al.* Construction of sized eukaryotic cDNA libraries using low input of total environmental metatranscriptomic RNA. *BMC Biotechnol.* **14**, 80 (2014).
196. Yannone, S. M., Hartung, S., Menon, A. L., Adams, M. W. W. & Tainer, J. A. Metals in biology: Defining metalloproteomes. *Curr. Opin. Biotechnol.* **23**, 89–95 (2012).
197. Yepiskoposyan, H. *et al.* Transcriptome response to heavy metal stress in *Drosophila* reveals a new zinc transporter that confers resistance to zinc. *Nucleic Acids Res.* **34**, 4866–4877 (2006).
198. Zalasiewicz, J. *et al.* When did the Anthropocene begin? A mid-twentieth century boundary level is stratigraphically optimal. *Quat. Int.* **383**, 196–203 (2015).
199. Zalewska, M., Trefon, J. & Milnerowicz, H. The role of metallothionein interactions with other proteins. *Proteomics* **14**, 1343–1356 (2014).
200. Zangger, K., Shen, G., Otvos, J. D. & Armitage, I. M. Oxidative dimerization in metallothionein is a result of intermolecular disulphide bonds between cysteins in the alpha domain. *Biochem. J.* **359**, 353–360 (2001).
201. Ziller, A. *et al.* Metagenomics analysis reveals a new metallothionein family: Sequence and metal-binding features of new environmental cysteine-rich proteins. *J. Inorg. Biochem.* **167**, 1–11 (2017).

1–49,49–201





RESUME

Le sol est essentiel à toute société humaine notamment pour la production d'aliments. Son fonctionnement repose sur des réseaux d'interactions entre les éléments qui le composent et toute perturbation modifie ces réseaux. Les microorganismes eucaryotes représentent une composante importante de l'écosystème édaphique car ils sont impliqués dans des processus essentiels comme la régulation de populations de procaryotes. Mais paradoxalement, ils restent peu étudiés comparés aux bactéries notamment lorsqu'on s'intéresse aux cycles biogéochimiques autres que celui du carbone comme par exemple ceux des métaux. Suite à une pollution par des métaux, certains de ces microorganismes eucaryotes développent des mécanismes « de résistance » cellulaires. Dans ce contexte, le laboratoire d'accueil a isolé, directement à partir d'extraits d'ARN de sol, des gènes eucaryotes impliqués dans la résistance cellulaire au Cd. Ces nouveaux gènes forment une famille codant des protéines riches en cystéines dont les positions sont conservées au sein de cette famille.

Mon projet de thèse a eu pour but de caractériser l'origine taxonomique et la fonction de cette famille de gènes. Dans un premier temps, la purification de cinq de ces protéines produites dans *Escherichia coli* et leurs caractérisations biochimiques par des méthodes spectrométriques ont permis de montrer que cette famille génique constitue une nouvelle famille de métallothionéines capables de chélater *in vitro* le Zn, le Cu et le Cd. Dans un second temps, une méthode de quantification par PCR quantitative de l'expression de ces gènes, extraits à partir de sol provenant de microcosmes, a été mise au point. Dans un troisième temps, nous avons tenté d'obtenir les régions génomiques bordant ces gènes environnementaux afin d'affilier les organismes qui les portent à un groupe taxonomique et d'analyser les régions promotrices de ces gènes par capture ciblée de gènes.

Soil is essential to human societies, especially for food production. Its functioning relies on interaction networks sensitive to environmental alterations. Eukaryotic microorganisms are an important component of the soil ecosystem where they are involved in essential processes such as the regulation of prokaryotic populations. However, they remain poorly studied compared to bacteria, especially concerning their roles in biogeochemical cycles other than the carbon one such as metal cycles. In response to soil metal contamination, some of these eukaryotic microorganisms develop cellular "resistance" mechanisms. In this context, the host laboratory has previously isolated, directly from soils, eukaryotic genes able to confer Cd resistance. These genes form a family coding for cysteine-rich proteins whose cysteine positions are conserved within this sequences.

My thesis project aimed at characterizing the function and taxonomic origin of this gene family. First, the purification of five of these proteins produced in *Escherichia coli* and their biochemical characterizations by spectrometric methods demonstrated that this gene family constitutes a new family of metallothioneins capable of chelating *in vitro* Zn, Cu and Cd. Some of these proteins are also able to confer Zn resistance when expressed in a sensitive yeast strain. In a second step, quantitative PCR methods for measuring expression levels of these genes in soil microcosms were developed. This will allow to evaluate the level of expression of these genes as a function of an increasing supply of exogenous metal. In a third step, we tried to obtain the genomic regions flanking these environmental genes in order to be able to associate the organisms from which they originate to a taxonomic group and to analyze the promoter regions of these genes.

