

Université du Québec
INRS-Institut Armand-Frappier

**Étude de la topologie transmembranaire
de la protéine MntH de *Escherichia coli***

par
Pascal Courville

Mémoire présenté

pour l'obtention du grade

Maître ès science (M.Sc.) en microbiologie appliquée

Jury d'évaluation

Président du jury
Examineur externe
Directeur de recherche

François Shareck (INRS-IAF)
Hervé Le Moual (Université McGill)
Mathieu Cellier (INRS-IAF)

©Droits réservés de Pascal Courville, 2003

Table des matières

Liste des abréviations	iv
Liste des figures et tableaux.....	vi
Résumé	vii
Introduction.....	1
Revue de littérature	5
1. Importance du fer (Fe^{2+}) et du manganèse (Mn^{2+})	6
1.1 Le fer	6
1.2 Le manganèse	7
2. Transporteur de fer divalent (Fe^{2+}) et de manganèse divalent (Mn^{2+}).....	8
3. Les transporteurs membranaires.....	11
3.1 Transport passif : la diffusion.....	11
3.2 Transport actif.....	12
3.2.1 Transport actif primaire	12
3.2.2 Transport actif secondaire.....	13
3.3 La classification universelle des systèmes de transport	14
4. La famille Nramp.....	14
4.1 Nramp 1 et 2	14
4.2 Smfp chez la levure.....	16
4.3 Homologues bactériens MntH	18
4.3.1 Évolution des homologues Nramp/MntH.....	18
4.3.2 Caractérisation des homologues bactériens	20
5. Étude de la relation structure/fonction des transporteurs Nramp	22
5.1 Données structurales connues des différents homologues Nramp/MntH	22
5.1.1 Conservation des séquences dans la famille Nramp/MntH	22
5.2 Prédiction de la structure secondaire chez les protéines membranaires	25
5.3 De la prédiction à la vérification expérimentale.....	27
5.3.1 Les fusions et insertions dirigées	27
5.3.2 Balayage à la cystéine	30
5.4 Résultat topologique chez les homologues Nramp.....	31
6. Objectif de mon projet de maîtrise	33

Discussion et conclusion.....	34
1. Prédiction d'un modèle topologique.....	35
2. Analyse des profils d'hydrophobicité.....	35
3. Fusion aléatoire avec le gène rapporteur Blam (β -lactamase).....	35
4. Vérification de la topologie par la fonctionnalité de la protéine étudiée.....	36
5. Utilisation du rapporteur Cat (chloramphénicol acétyl transférase).....	37
6. Caractérisation d'homologues MntH appartenant aux groupes B et C.....	38
7. Problématique et perspective.....	39
Remerciements.....	40
Bibliographie.....	41
Annexe 1 : Contribution de l'étudiant.....	47
Annexe 2 : Article 1.....	48
Annexe 3 : Article 2.....	97

Liste des abréviations

ABC	: «ATP Binding Cassette»
ADN	: acide désoxyribonucléique
ARN	: acide ribonucléique
ATP	: adénosine 5'-triphosphate
<i>B. subtilis</i>	: <i>Bacillus subtilis</i>
Blam	: β -lactamase
Cat	: chloramphénicol acétyl transférase
CCCP	: hydrazone méta-chlorophénylique de cyanure carbonyle
Cd	: cadmium
CMI	: concentration minimale inhibitrice
Co	: cobalt
Cu	: cuivre
Cys	: cystéine
Da	: dalton
<i>E. faecalis</i>	: <i>Enterococcus faecalis</i>
<i>E. coli</i>	: <i>Escherichia coli</i>
EGTA	: «Ethylene glycol-bis (β -aminoethyl ether)-N,N,N',N'-tetraacetic acid»
ExbB, D	: «excretion of colicin»
Fe ²⁺	: fer divalent
FeoB	: «ferrous iron uptake system»
FhuA	: «ferric hydroxamate uptake system»
gfp	: «green fluorescent protein»
GTP	: guanosine 5'-triphosphate
H ₂ O ₂	: peroxyde d'hydrogène
HA	: hémagglutinine
I	: isoleucine
K	: lysine
L	: leucine
LacZ	: β -galactosidase
M	: méthionine
Me ²⁺	: métaux divalents
MEMSAT	: «membrane protein structure and topology»
Mn ²⁺	: manganèse divalent
MntH	: transporteur de manganèse dépendant du proton
MntR	: «Mn ²⁺ transport repressor»
MTSES	: «2-sulfonatoethylmethane thiosulfonate»
NEM	: n-éthyl maleimide
NO	: oxyde nitrique
NOS	: «nitric oxyde synthase»
Nramp	: «Natural resistance associated macrophage protein»
O ₂ ⁻	: anion superoxyde
O ₂	: oxygène
pH	: potentiel d'hydrogène

PhoA	: phosphatase alcaline
R	: arginine
<i>S. typhimurium</i>	: <i>Salmonella typhimurium</i>
SitABCD	: «Salmonella iron transporter»
Smfp	: «suppressor of mitochondrial import function»
SOD	: superoxyde dismutase
STM	: segment transmembranaire
TC	: commission sur le transport
TetAB	: antiporteur de la tétracycline
TMHMM	: «transmembrane hidden markov model»
TonB	: «T-one resistance in <i>E. coli</i> »
V	: valine

Liste des figures et tableaux

Figure 1	article 1	Phylogenetic analysis of the Nramp/MntH family93
Figure 2	article 1	Genomic organization of <i>mntH</i> C β locus in <i>Wigglesworthia</i>94
Figure 3	article 1	Linkage of oral <i>streptococcaceae</i> <i>mntH</i> C β gene with different IS-like elements.....95
Figure 4	article1	Consensus transmembrane topology predicted for Nramp and MntH proteins.....96
Figure 5	article 2	Sequence analysis of selected MntH homologues from group A.....28
Figure 6	article 2	Detection of MntH-mutant proteins by Western Blot Analysis.....29
Figure 7	article 2	EcoliA MntH-dependent intracellular accumulation of protons in presence of cadmium (Cd ²⁺).....30
Figure 8	article 2	EcoliA MntH consensus transmembrane topology.....31
Figure 9		Modèle du transport des sidérophores à l'aide de la protéine FhuA chez <i>E. coli</i>10
Figure 10		Alignement de séquences du segment transmembranaire 9 des membres de la famille Nramp/MntH.....24
Tableau 1	article 1	Disk assay of metal sensitivity conferred by expression of MntH A, B and C in <i>E. coli</i>92
Tableau 2	article 2	Sequence of the oligonucleotide primers used in this study.25
Tableau 3	article 2	Bacterial strains and plasmids used in this study.....26
Tableau 4	article 2	Characterization of EcoliA Blam constructs.....27
Tableau 5	article 2	Analysis of Cat targeted fusions.....27
Supplemental data 1 article 2		Transport activity of EcoliA is impaired by insertion of the dipeptide GS after L ₁₇₉32
Supplemental data 2 article 2		Detection of R91-Cat, G226-Cat, L324-Cat, K387-Cat and L412-Cat by ³⁵ S-Met metabolic labeling33

Résumé

Les métaux de transition fer et manganèse sont deux cations essentiels pour tout organisme cellulaire, servant principalement de cofacteurs à de multiples enzymes. Par contre, ces métaux peuvent engendrer différentes réactions cytotoxiques, d'où l'importance d'une régulation très stricte de leur métabolisme (l'expression des transporteurs de ces cations). L'importance des fonctions dans lesquelles participent ces métaux (résistance aux infections, nutrition) rend nécessaire la caractérisation fonctionnelle des transporteurs impliqués dans le maintien de leur homéostasie. Chez les bactéries, les protéines de type MntH (transporteur de manganèse dépendant du proton) sont des perméases facilitant l'acquisition du fer et du manganèse divalents. On distingue trois groupes de séquences, MntH A, B et C. Ce dernier est subdivisé en Ca, Cβ et Cγ. Les protéines MntH appartiennent à la famille définie par la Natural resistance associated-macrophage protein (Nramp1) commune aux organismes eucaryotes. Chez les mammifères, les deux homologues Nramp 1 et 2 sont impliqués dans la défense contre différents pathogènes et le métabolisme du fer, respectivement. La caractérisation moléculaire de ces perméases représente un défi de taille de par leur nature de protéine intégralement membranaire. Très peu de données sont connues sur leur aspect structural.

Durant mes études de maîtrise, j'ai entrepris d'élucider la topologie de MntH chez *Escherichia coli* (EcoliA) selon une approche génétique, soit en fusionnant ou en insérant un gène rapporteur. Les gènes rapporteurs utilisés codent pour la β-lactamase et pour la chloramphénicol acétyl transférase (Cat), qui sont actives dans le périplasma et le cytoplasme, respectivement. Les résultats obtenus indiquent que EcoliA possède 11 segments transmembranaires avec son extrémité N-terminale du côté cytoplasmique et C-terminale dans l'environnement périplasmique. La topologie établie a été validée en obtenant des protéines hybrides fonctionnelles.

Nous avons également engagé la caractérisation de différents homologues MntH des groupes B et C en démontrant que certains sont fonctionnels lorsque exprimés chez *E. coli*. Les résultats obtenus ouvrent la voie aux études structurales que je désire entreprendre.


Pascal Courville


Mathieu Cellier

Introduction

Le fer et le manganèse, sont deux micro-nutriments essentiels pour la cellule, agissant principalement à titre de cofacteurs pour de nombreuses enzymes. Du fait de leur importance, les bactéries disposent de plusieurs moyens pour acquérir et assimiler ces cations divalents. Parmi les transporteurs découverts dernièrement, figure la protéine bactérienne MntH (Makui *et al.*, 2000 (1-2)) (Proton-dependent Manganese transporter) appartenant à la vaste famille Nramp (Natural resistance-associated macrophage protein) qui est également présente chez les mammifères (Vidal *et al.*, 1993 (2); Gruenheid *et al.*, 1995 (2)), les plantes (Belouchi *et al.*, 1995 (2)) et les levures (West *et al.*, 1992 (2)). *Nramp1*, un des gènes de cette famille les plus couramment étudiés code pour une protéine membranaire qui permet de contrôler la réplication intracellulaire de certains microbes pathogènes qui résident à l'intérieur de vacuoles dans les phagocytes professionnels (par exemple *Salmonella* ou *Mycobacterium spp.* (Gros *et al.*, 1983)). Par contre, le gène *Nramp2* est exprimé dans de nombreux tissus et est impliqué dans l'acquisition des métaux provenant de l'alimentation en étant actif au niveau de la membrane plasmique des cellules épithéliales de l'intestin. *Nramp2* est également fortement exprimé dans les tissus périphériques (précurseur des érythrocytes, rein) et facilite l'acquisition du fer provenant de la transferrine par les cellules de l'organisme (Canonne-Hergaux *et al.*, 2000; Canonne-Hergaux et Gros, 2002). L'absence d'un gène *Nramp2* fonctionnel est associée à une forme d'anémie chez la souris et le rat (anémie microcytique; Fleming *et al.*, 1998 (2)) alors que l'absence du gène *Nramp1* fonctionnel rend les animaux susceptibles à certaines infections. Ces phénotypes différents impliquent des fonctions non redondantes. Chez les autres organismes où des homologues ont été étudiés, les protéines Nramp semblent toujours être impliquées dans l'homéostasie du fer ou du manganèse.

Un grand nombre d'homologues bactériens appartenant à la famille Nramp/MntH a été révélé grâce aux récents projets de séquençage. Les séquences MntH bactériennes sont réparties en 3 groupes d'homologie, MntH A, B et C, définis par alignement multiple des séquences disponibles et analyses phylogénétiques (Cellier, 2001). La majorité des homologues bactériens récemment caractérisés appartient au groupe A. Par exemple, le gène *mntH* de *Escherichia coli* a été caractérisé comme étant un transporteur de cations divalents ayant une nette

.«Les chiffres entre parenthèses après certaines références indiquent dans quel article elles sont citées»

préférence pour le manganèse (Makui *et al.*, 2000 (1-2)), tandis que chez *Bacillus subtilis* (BsubA), le gène *mntH* est nécessaire pour sa croissance en milieu minimum (Que et Helmann, 2000 (1-2)).

Notre laboratoire s'intéresse particulièrement à l'évolution des homologues de la famille Nramp/MntH. Nous avons récemment démontré que les homologues du groupe B et A seraient d'origine plus ancestrale tandis que ceux du groupe C auraient été acquis récemment, probablement par transfert horizontal (Cellier *et al.*, 2001 (1-2); Richer *et al.*, 2003 (2)). Différents transferts pourraient être à l'origine de trois sous groupes distincts (C α , C β et C γ). Les protéines eucaryotes, seraient apparues suite à un évènement endosymbiotique impliquant un homologue du groupe A. Les protéines MntH du groupe C démontrent un niveau de similarité plus élevé avec des protéines Nramp eucaryotes qu'avec les protéines des autres groupes MntH. Puisqu'aucun homologue appartenant aux groupes B et C n'avait encore été étudié, nous avons pris l'initiative de caractériser certaines protéines appartenant à ces deux groupes afin de comparer leur activité une fois exprimée chez *E. coli*.

L'expression fonctionnelle chez *E. coli* d'homologue MntH de différents groupes, implique qu'ils doivent adopter une topologie transmembranaire fonctionnelle dans la membrane d'*E. coli*. La détermination expérimentale du nombre de segments transmembranaires et de leur orientation permettra une étude ciblée de la relation structure/fonction des protéines intégralement transmembranaires. Des études comparées du mécanisme de transport des cations divalents par la protéine MntH A de *E. coli* et les groupe B et C permettront d'évaluer leur évolution structurale et fonctionnelle.

Aucune étude visant à déterminer la topologie des homologues Nramp/MntH n'ayant encore été effectuée, nous avons entrepris de déterminer la topologie transmembranaire de la protéine MntH A d'*E. coli*. Pour ce faire, nous avons combiné plusieurs approches visant à : i) élaborer d'après l'analyse de la séquence en acide aminé, une prédiction de la topologie transmembranaire à l'aide de plusieurs programmes (prédiction consensus); ii) utiliser un rapporteur fonctionnel indiquant le

côté périplasmique de la membrane pour créer des fusions génétiques au hasard dans la séquence codante *mntH* et des insertions dirigées; iii) utiliser un rapporteur fonctionnel indiquant le côté cytoplasmique pour générer des fusions dirigées; iv) démontrer l'expression des constructions et l'activité des rapporteurs; v) établir l'activité résiduelle de transport des métaux des constructions, comprenant l'insertion ciblée d'un marqueur permettant de s'assurer que la topologie établie correspond à une protéine fonctionnelle.

Les résultats obtenus au cours de ma maîtrise sont présentés en détail dans les articles situés en annexe de ce mémoire : «Horizontal gene transfer of 'prototype' Nramp in bacteria» publié dans *Journal of Molecular Evolution* (2003) 57, 363-376 et «Determination of the transmembrane topology of *Escherichia coli* Nramp ortholog» dans *Journal of Biological Chemistry* in press.

Revue de littérature

1 Importance biologique du fer (Fe^{2+}) et du manganèse (Mn^{2+})

1.1 Le fer

Le fer est un micro-élément essentiel pour toute forme de vie, autant eucaryote que procaryote. Certaines espèces bactériennes comme *Lactobacillus plantarum* (Byers et Arceneaux, 1998) et *Borrelia burgdorferi* (Posey et Gherardini, 2000) font exception à cette règle en utilisant le manganèse (Mn) à la place du fer (Fe). Le fer est impliqué dans des réactions d'oxydo-réduction et peut libérer un électron en passant du fer ferreux (Fe^{2+}) au fer ferrique (Fe^{3+}); le Fe^{3+} peut ainsi agir en tant qu'accepteur final d'électron durant la respiration anaérobie chez certaines bactéries. Le fer est également utilisé comme cofacteur de nombreuses réactions enzymatiques incluant celle pour la synthèse d'ADN (Braun et Hantke, 2001). Malgré ces effets positifs, le fer libre peut catalyser en présence d'oxygène, des réactions potentiellement dangereuses, comme la réaction de Fenton, qui augmente la concentration de radicaux oxygénés dans la cellule (réaction de Fenton : $\text{O}_2^- + \text{H}_2\text{O}_2 \rightarrow \text{O}_2 + \text{OH}^- + \text{OH}$). Les radicaux oxygénés peuvent causer des dommages en attaquant l'ADN. Cependant, le fer peut également agir comme cofacteur d'enzymes protectrices telle que la superoxyde dismutase ($2\text{O}_2^- + 2\text{H}^+ \rightarrow \text{H}_2\text{O}_2 + \text{O}_2$) et la catalase ($2\text{H}_2\text{O}_2 \rightarrow 2\text{H}_2\text{O} + \text{O}_2$), qui permettent l'élimination des radicaux oxygénés ainsi créés (Fridovich, 1978). Inversement, le fer peut servir de cofacteurs à des enzymes qui génèrent de grandes quantités de radicaux libres comme l'anion superoxyde (O_2^-) et l'oxyde nitrique (NO) (oxydase des phagocytes, phox et synthase d'oxyde nitrique NOS) (Braun *et al.*, 1998). Lors d'une infection, l'hôte se défend notamment en séquestrant tout le fer disponible; cependant les pathogènes les plus virulents expriment différents systèmes de capture du fer de haute efficacité.

Par exemple, il a été démontré que le fer est impliqué dans la virulence de la bactérie *Pseudomonas aeruginosa*. Un mutant *tonB* dépourvu de système d'acquisition de Fe^{3+} via le sidérophore pyoverdine, est capable de croître dans les poumons et les muscles de souris immunosupprimées par contre il est incapable de causer la mort de l'animal (Takase *et al.*, 2000).

1.2 Le manganèse

Ce métal est un micro-élément important non seulement pour les espèces *Lactobacillus* et *Borrelia* mais aussi pour la plupart des organismes cellulaires. Le rôle physiologique du manganèse a cependant été relativement peu étudié. Ce cation joue un rôle crucial dans la vie aérobique sur terre en agissant comme co-facteur dans le photosystème II pour l'oxydation de l'eau lors de la photosynthèse (Debus, 2000). Le manganèse peut également favoriser la survie des cellules en facilitant la détoxification des radicaux oxygénés en permettant la conversion de l'ion O_2^- en H_2O_2 et O_2 , soit de manière catalytique (Mn SOD, Mn peroxydase) soit selon des mécanismes encore peu connus (Archibald, 1986; Archibald et Fridovich, 1982a; Archibald et Fridovich, 1982b). La plupart des enzymes nécessaires pour convertir les intermédiaires comme le pyruvate, l'oxaloacetate et le phosphoglycérate sont soit dépendants ou bien stimulés en présence de ce métal, ce qui pourrait signifier que le Mn jouerait un rôle dans le métabolisme du carbone (Kehres et Maguire, 2003). Des systèmes de transport du manganèse ont été récemment impliqués dans la virulence de plusieurs pathogènes bactériens (Cellier, 2001). Il semble également clair que l'hôte restreint l'accès des pathogènes à cette ressource lors de l'infection (Jabado *et al.*, 2000 (1)).

Malgré leurs effets bénéfiques pour les cellules vivantes, ces métaux peuvent être nocifs en trop haute concentration et causer la mort cellulaire. Il est donc primordial de réguler strictement leur entrée et sortie à travers les membranes ainsi que leur stockage sous une forme non dangereuse. Ainsi pour limiter ses effets toxiques, uniquement le fer est entreposé dans la cellule sous forme ferrique, couplé à la ferritine. (Chasteen, 2003). Le fer est la plupart du temps retrouvé dans l'hôte couplé à des protéines possédant un groupement hème, telle que l'hémoglobine, la phagocyte oxydase; il est lié avec une forte affinité par les protéines de transport tel que la transferrine et à la lactoferrine.

2. Transporteur de fer divalent (Fe^{2+}) et de manganèse divalent (Mn^{2+})

Dans un environnement aérobique et au pH physiologique, le fer se retrouve sous la forme Fe^{3+} et a tendance à précipiter et à créer un polymère insoluble ($\text{Fe}(\text{OH})_3$) (van der Helm et Chakraborty, 2001). Pour contourner le problème, les bactéries utilisent différents moyens pour solubiliser et assimiler le fer ferrique. Un des moyens pour contrer la faible solubilité du fer est de synthétiser des sidérophores qui sont des petites molécules capables de se lier au Fe^{3+} avec une affinité exceptionnelle. Le complexe Fe^{3+} -sidérophore, ayant une masse variant entre 500 et 1,500 Dalton (Da) au dessus de la taille d'exclusion des porines de la membrane externe, sera véhiculé activement à travers la membrane externe via un récepteur de très haute affinité, par exemple FhuA. L'énergie nécessaire pour transloquer le complexe, provient du complexe constitué des protéines TonB, ExbB et ExbD, situé dans la membrane cytoplasmique (Fig. 9) (Braun et Hantke, 2001). Alternativement le fer réduit en Fe^{2+} est soluble et diffuse librement. Un grand nombre d'espèces bactériennes utilisent le système codé par les gènes *feoABC* pour transporter le Fe^{2+} à travers la membrane interne (Kammler *et al.*, 1993). L'opéron SitABCD et la protéine MntH peuvent également faciliter le transport de manganèse et de fer divalent à travers la membrane interne.

Le transport du manganèse se fait par trois grandes classes de transporteurs. Tout d'abord la famille MntABC(D) fait partie de la super famille des ATPases et constitue un sous groupe répandu chez de nombreuses espèces (Dintilhac *et al.*, 1997). D'autre part, les homologues appartenant à la famille Nramp définissent une autre famille très conservée dans l'évolution (Cellier *et al.*, 1995 (1-2)). Enfin, les lactobacilles possèdent également un troisième type de transporteur, MntA, une ATPase de type P (Hao *et al.*, 1999). *Lactobacillus plantarum* possède 5 gènes potentiellement impliqués dans le transport du manganèse (1 *mntA*, 1 *mntABCD* et 3 *mntH* C β 1, 2, 3). La redondance en transporteur de manganèse reflète bien la dépendance de cet organisme envers ce métal. D'autres transporteurs doivent exister car *Helicobacter pylori* et *Vibrio cholerae* possèdent des enzymes dépendantes du Mn, mais aucun homologue des transporteurs précédemment mentionnés.

Figure 9. Modèle du transport des sidérophores à l'aide de la protéine FhuA chez *E. coli*. La protéine FhuA permet le transport du complexe Fe^{3+} -sidérophore au travers de la membrane externe. Le complexe formé des protéines ExbB, ExbD et TonB est impliqué dans le transfert de l'énergie provenant de la membrane cytoplasmique vers la membrane externe.

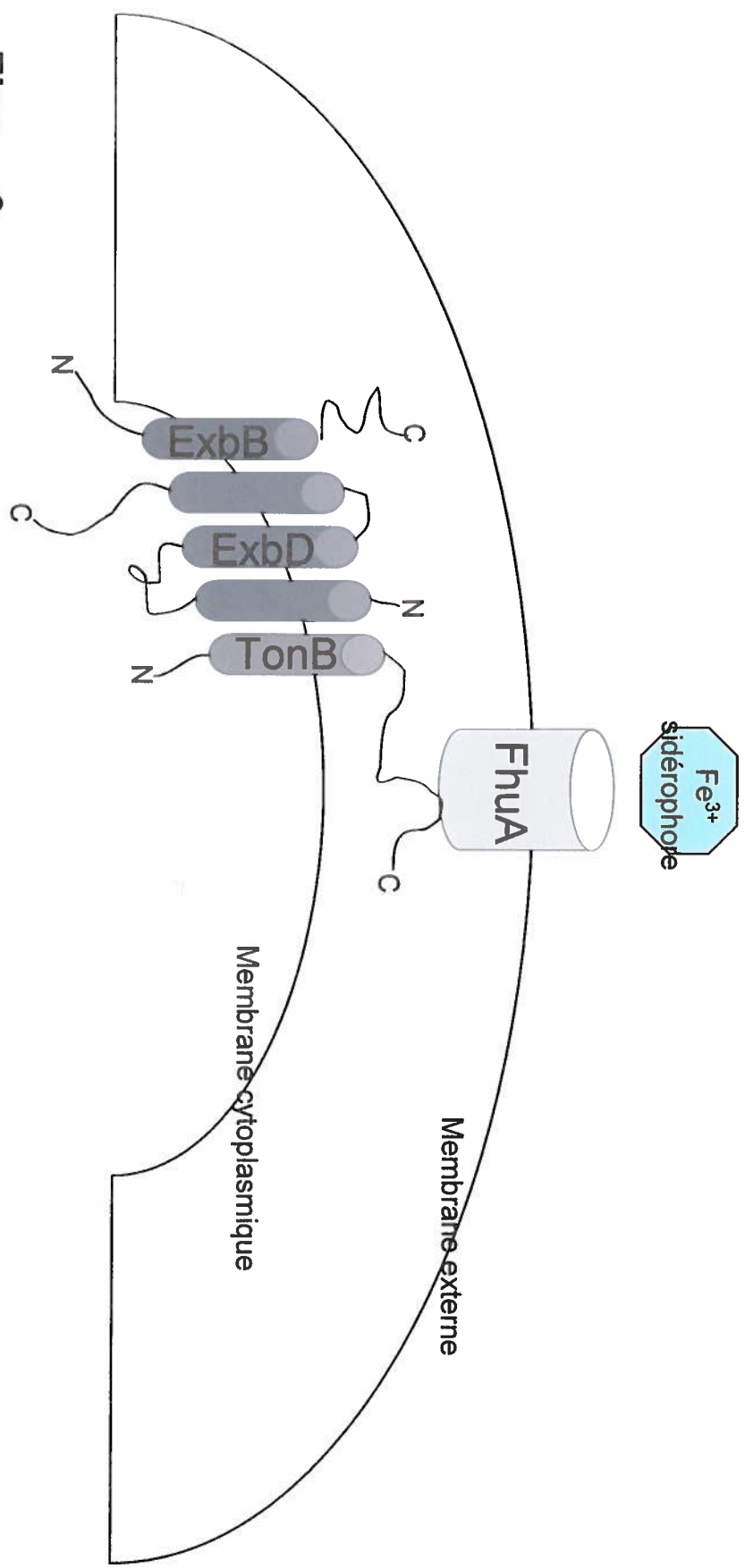


Figure 9

3 Les transporteurs membranaires

La membrane plasmique sert de frontière semi-perméable entre l'intérieur de la cellule et son environnement. Étant donné la très grande diversité des éléments nutritifs, les bactéries maintiennent différents moyens pour accumuler ces substances à travers la bicouche lipidique. Le transport peut se faire soit de façon passive, sans apport énergétique de la cellule, ou bien de façon active en utilisant l'énergie générée par la cellule généralement sous forme d'ATP ou par le gradient électrochimique du proton (Lodish *et al.*, 1997). De par leur rôle crucial pour l'économie cellulaire, les transporteurs membranaires représentent près de 30% du protéome bactérien. Ils constituent donc des cibles pour le développement de médicaments. (Booth *et al.*, 2001).

3.1 Transport passif : la diffusion

La diffusion est un phénomène physique utilisé par la cellule pour accumuler différents types de molécules selon leur gradient de concentration, c'est à dire dans le sens allant du côté de la membrane où le soluté est le plus concentré vers le moins concentré.

La diffusion passive permet de transporter les substances non polaires et relativement liposolubles comme par exemple l'eau, l'oxygène, le gaz carbonique, les graisses et l'alcool. Les molécules à caractère hydrosoluble sont repoussées par la partie interne de la membrane qui est constituée de chaînes hydrocarbonées non polaires. Seules les molécules polaires de taille assez petites peuvent passer aux travers des pores membranaires remplis d'eau (porines).

La diffusion facilitée permet de transporter de plus grosses molécules comme différents sucres simples (glucose) de façon efficace et assez sélective. Ces substances sont transportées dans le cytoplasme par des transporteurs protéiques membranaires qui utilisent le gradient de concentration du soluté et changent de conformation durant le transport (Eberstadt *et al.*, 1996).

3.2 Transport actif

Le transport actif ressemble en certains points à la diffusion facilitée, car ce type de transport nécessite des transporteurs protéiques spécifiques au substrat à véhiculer. Par contre, le transport actif permet de transporter contre le gradient de concentration en utilisant l'énergie générée durant le métabolisme cellulaire. Ce type de transport peut être soit primaire ou secondaire selon la source d'énergie qu'il utilise, et concerne majoritairement les molécules chargées comme les ions et les acides aminés.

3.2.1 Transport actif primaire

Ce type de transport est un processus endergonique, c'est à dire qu'il utilise une source d'énergie primaire pour véhiculer activement un soluté contre son gradient de concentration. La source d'énergie primaire provient souvent de l'hydrolyse des liaisons anhydrides de l'ATP ou d'un lien pyrophosphate inorganique.

L'ATP est généralement utilisé par les protéines appartenant à la super-famille ABC (ATP binding cassette) comme les protéines FeoB ou MntABC(D) retrouvées chez différentes espèces bactériennes et décrits comme des transporteurs de fer ferreux et de manganèse respectivement (Greie *et al.*, 2001). Cependant le GTP au lieu de l'ATP pourrait être la source d'énergie utilisé par FeoB (Hantke, 2003). Les ATPases se divisent en trois sous-groupes selon ce qu'ils transportent et leur structure.

L'ATPase de type P se trouve majoritairement à la membrane plasmique et est phosphorylée directement durant le processus du transport. Elle se distingue des autres ATPases car elle est inhibée par les ions vanadate. MntA, identifié chez *Lactobacillus plantarum*, appartient à cette classe; c'est un transporteur de manganèse exprimé lorsque la bactérie est privée du cation divalent (Hao *et al.*, 1999).

Les ATPases de type F sont composées de deux sous unités F_0 et F_1 et sont exprimées dans la membrane plasmique des bactéries et dans la membrane interne des

mitochondries. Elles sont inhibées par l'oligomycine. Un analogue à cette classe, l'ATPase de type V se retrouve dans les membranes des vacuoles des cellules végétales et dans les vésicules acides comme les endosomes chez les animaux (Lodish *et al.*, 1997).

L'ATP est la source d'énergie la plus couramment utilisée. Une autre source d'énergie primaire est la décarboxylation d'un substrat situé du côté cytoplasmique. Les cellules peuvent également utiliser l'énergie de la lumière comme les rhodopsines microbiennes retrouvées chez les archaebactéries, les bactéries et les moisissures (Saier, 2001 (2)).

3.2.2 Transport actif secondaire

Le transport actif actionné par dissipation d'un gradient d'ions, est communément appelé transport actif secondaire, car il transporte le soluté en utilisant un gradient de potentiel électrochimique préétabli (Voet et Voet, 1998). Plusieurs catégories de transporteurs sont divisés et identifiées par leur mécanisme d'action.

Le premier est de type uniporteur, c'est à dire qu'il véhicule une seule substance à la fois, soit par diffusion facilitée ou bien en utilisant l'énergie du potentiel membranaire. Les deux autres mécanismes utilisent l'énergie générée par le transport d'un premier soluté selon son gradient de concentration pour véhiculer le substrat. On parle de symporteur si les deux solutés sont véhiculés dans la même direction et antiporteur s'ils sont transportés dans une direction opposée.

MntH (Transporteur de Manganèse dépendant du Proton) est un transporteur de type symporteur qui véhicule différents cations divalents à l'intérieur de la cellule en utilisant le gradient électrochimique du proton comme source d'énergie. La dépendance du transport vis à vis du gradient de proton peut être vérifiée en utilisant différents protonophores, comme le "carbonyl cyanide *m*-chlorophenyl hydrazone" (CCCP) qui dissipe le gradient des protons et rend ce transport impossible (Makui *et al.*, 2000 (1-2)).

3.3 La classification universelle des systèmes de transport

Un laboratoire a développé une méthode pour classer tous les types de transporteurs membranaires (commission sur le transport; TC transport commission). (Saier, 1999). Cette nouvelle classification est accessible via Internet (<http://www-biology.ucsd.edu/~msaier/transport/>). Ce système classe les protéines selon la phylogénie et la fonction des transporteurs. La valeur de la phylogénie pour la classification indiquent que deux protéines homologues qui proviendraient d'un ancêtre commun et qui possèdent des séquences en acides aminés similaires, devraient partager des structures secondaires et tertiaires similaires. Certaines exceptions ont cependant pu être décelées (Saier, 2001 (2)). Les membres d'une même famille d'homologues doivent également avoir un mécanisme de transport similaire. À la lumière de ces principes, la nouvelle classification regroupe 4 critères hiérarchiques. Premièrement, les transporteurs sont classés selon le type d'énergie utilisé. Ensuite, ils sont divisés en familles reconnaissables, selon les critères qui sont propres à chacune d'elles, en troisième lieu en groupes phylogénétiquement proches et finalement, selon le substrat spécifique à chacune de ces familles (Saier, 2001 (2)).

4 La famille Nramp

4.1 Nramp 1 et 2

Le gène *Nramp* (Natural Resistance Associated Macrophage protein), a été découvert par clonage positionnel (Vidal *et al.*, 1993 (2)). Il a été caractérisé initialement comme un facteur permettant de limiter la réplication intracellulaire de certains microbes pathogènes qui peuvent résider dans des vacuoles (par exemple *Salmonella* ou *Mycobacterium spp.*) (Gros *et al.*, 1983). Le rôle de ce gène est démontré par l'observation d'une réplication microbienne intracellulaire lors de la phase initiale de l'infection qui est incontrôlée quand le gène est inactif mais qui est contrôlée lorsque le gène fonctionnel est présent. L'expression de gènes codant pour des homologues Nramp telles que les protéines de rat DMT1 (Nramp2) (Fleming *et al.*, 1998 (2)) et de levure Smf1p (West *et al.*, 1992 (2)) dans des œufs de xénopes a

permis de mesurer des activités fonctionnelles similaires, indiquant que les protéines Nramp sont des transporteurs secondaires de type symporteur facilitant le co-transport d'un ion métallique divalent (Me^{2+}) avec celui d'un proton, dans la même direction (Canonne-Hergeaux *et al.*, 2001). Par contre, une controverse existe sur le sens du transport des métaux par Nramp1 à travers la membrane phagosomale. Certaines études indiquent que Nramp1 évacue les cations du phagosome privant ainsi les bactéries de ces éléments nutritifs (Jabado *et al.*, 2000 (1)). D'autres études suggèrent que Nramp1 pourrait favoriser l'accumulation des métaux dans le phagosome et ainsi faciliter la réaction de Fenton dans le but d'éliminer les bactéries pathogènes, en proposant des mécanismes d'action très différents : (symporteur) à topologie inversée (Kuhn, Lafuse et Zwillings 2001 (2)) ou bien antiporteur (Goswami *et al.*, 2001). À ce jour, de nouvelles études tendent à valider la première hypothèse. Ce mécanisme de symport Me^{2+}/H^+ est également impliqué dans le rôle joué par Nramp2 dans l'homéostasie du Fer (Fe^{2+}) chez les mammifères (Fleming *et al.*, 1998 (2)) ainsi que les rôles démontrés des protéines Smf1-3p dans le contrôle de l'homéostasie du Fe^{2+} et Mn^{2+} chez la levure (Portnoy, Liu et Culotta, 2000 (2)).

Nramp2, le second gène appartenant à la famille Nramp découvert chez les mammifères est, contrairement à Nramp1 qui a été détecté spécifiquement dans les membranes du phagosome des phagocytes (Gruenheid *et al.*, 97 (1)), actif dans la membrane apicale des cellules de l'intestin grêle (duodénum), et permet l'absorption des métaux provenant de l'alimentation. Cette fonction est indépendante de la transferrine (Canonne-Hergaux *et al.*, 1999); Nramp2 est également impliquée dans le cycle de la transferrine qui permet l'absorption du fer par les tissus périphériques. Nramp2 est alors actif au niveau de la membrane plasmique et de l'endosome en recyclage vers la membrane plasmique, facilitant ainsi l'acquisition du fer dépendant de la transferrine. La présence de Nramp2 au niveau de l'endosome des cellules épithéliales de rein a aussi été rapportée (Canonne-Hergaux et Gros, 2002). L'absence d'une protéine fonctionnelle Nramp2 est associée à une forme d'anémie (anémie microcytique) chez la souris et chez le rat (Fleming *et al.*, 1998).

4.2 Smfp chez la levure.

Des membres de la famille Nramp ont également été caractérisés chez d'autres mammifères (Cellier *et al.*, 1994 (1)), des invertébrés (Rodrigues *et al.*, 1995) et chez les plantes (Belouchi *et al.*, 1995 (2)), où ils participent au métabolisme des cations divalents métalliques. Chez la levure, on retrouve 3 homologues appartenant à la famille Nramp. Leur dénomination Smf1p, Smf2p et Smf3p, indique suppressor of mif1; *mif1* est une peptidase mitochondriale impliquée dans l'importation des protéines à l'intérieur de la mitochondrie. La surexpression de ces protéines a permis d'éliminer le défaut associé à *mif1*. Horwich *et al.*, pensaient donc que Smf1p et Smf2p étaient des facteurs impliqués dans la translocation des protéines dans la mitochondrie (West *et al.*, 1992 (2)). Cependant, le groupe de N. Nelson travaillant à caractériser les gènes de levure permettant de résister à l'appauvrissement du milieu à l'aide de chélateur, a démontré que Smf1p et Smf2p sont des transporteurs de manganèse de haute affinité (Nelson, 1999). Le troisième homologue Smf3p, qui possède un taux d'identité de 50% avec Smf1p et Smf2p, est plutôt impliqué dans l'homéostasie du fer chez la levure.

Smf1p, est un transporteur de cations divalents dépendant du proton ayant comme substrat préférentiel le manganèse mais il peut également véhiculer le fer, le zinc, le cadmium et le cuivre (Liu *et al.*, 1997). Cette protéine agit à la surface cellulaire. L'élimination du gène chromosomique a un effet marginal; ce qui signifie que Smf1p n'est pas le seul transporteur de manganèse dans les conditions normales. Les mutants de *Smf1p* sont très sensibles à la présence du chélateur de métaux EGTA, principalement en période de stress oxydatif (Supek *et al.*, 1996 (2)). Ce défaut peut être éliminé en supplémentant le milieu avec du manganèse, qui peut agir comme anti-oxydant.

Contrairement à Smf1p, Smf2p a été plutôt localisé dans la membrane de vésicules intracellulaires permettant ainsi l'approvisionnement en manganèse et sa distribution dans la cellule. Le manganèse agit par exemple à titre de cofacteur de l'enzyme superoxyde dismutase 2 (SOD2) retrouvée dans la mitochondrie et de

l'invertase, enzyme impliqué dans la glycosylation à l'intérieur de l'appareil de Golgi (Rudolph *et al.*, 1989). Les protéines Smf1p et Smf2p sont exprimées lorsque la cellule est privée de manganèse. Dans ces conditions, elles ne sont pas reconnues par la protéine Bsd2p et peuvent ainsi se rendre à leur compartiment respectif. Par contre, lorsque la cellule trouve suffisamment de manganèse, Smf1p et Smf2p sont reconnues par Bsd2p, qui les dirige vers la vacuole pour être dégradées (Liu et Culotta, 1999; Portnoy *et al.*, 2000). Le double mutant *smf1p* et *smf2p* est utilisé comme souche permettant de tester la fonction de différents homologues, tant eucaryotes que procaryotes, afin de déterminer s'ils peuvent supprimer la sensibilité à l'EGTA. *Nramp2* a permis de compléter ce double mutant mais pas *Nramp1* (Pinner *et al.*, 97). L'homologue Nramp chez la bactérie *Mycobacterium leprae*, est également capable de compléter l'absence du gène *smf3p* (Reeve *et al.*, 2002 (2)). La complémentation des doubles mutants *smf1-2* a été utilisée pour isoler des homologues de plantes (Curie *et al.*, 2000 (1); Thomine *et al.*, 2000 (1)). Ceci démontre la grande conservation de Nramp à travers l'évolution.

Smf3p, malgré son homologie avec les autres Smfp, semble être impliquée dans l'homéostasie du fer qui est emmagasiné dans la vacuole. Smf3p est exprimée dans la membrane de la vacuole chez la levure. Sa localisation ne change pas lorsque l'environnement est pauvre en métaux et cette protéine n'est pas régulée par Bsd2p. Elle est régulée distinctement par trois éléments promoteurs différents. Premièrement, Smf3p est sous contrôle du fer grâce au facteur de transcription Aft1 (Rutherford *et al.*, 2001), et par l'intermédiaire d'une autre régulation, indépendante de Aft1 (Portnoy *et al.*, 2002). Troisièmement, l'expression de Smf3p est régulée par l'oxygène. Smf3p est exprimé, lorsque la levure se trouve en condition anaérobie; cette régulation implique le répresseur Rox1p (Kastaniotis et Zitomer, 2000). La double régulation de Smf3p par le fer et l'oxygène a l'avantage de limiter au maximum la mobilisation du fer intracellulaire stocké dans la vacuole, ce qui limite les effets néfastes de ce métal en condition aérobie (Jensen et Culotta, 2002).

4.3 Homologues bactériens MntH

4.3.1 Évolution des homologues Nramp/MntH

Depuis la réalisation des différents projets de séquençage génomiques, un nombre important d'homologues bactériens de la famille Nramp a été découvert; on les dénomme MntH (Transporteur de Manganèse dépendant du Proton (H⁺)). Cette famille protéique est présente autant chez les bactéries à Gram positif que négatif. À ce jour, aucune protéine MntH n'a été identifiée chez les Archaeobactéries. Les protéines MntH sont classées en 3 groupes d'homologie, MntH A, B et C, définis par alignement multiple des séquences disponibles et analyses phylogénétiques (Cellier *et al.*, 2001 (1-2)). Les protéines MntH du groupe B ont jusqu'à présent été identifiées seulement chez des bactéries préférant des conditions anaérobiques. La plupart des protéines MntH caractérisées fonctionnellement jusqu'à maintenant font partie du groupe A; quelques membres appartenant aux groupes B et C, démontrent une sensibilité aux cations divalents lorsque surexprimés chez *E. coli* (Richer *et al.*, 2003 (2)). Le fait qu'une analyse phylogénique basée sur les séquences MntH A soit consistante avec la phylogénie déduite de l'analyse des séquences d'ARN 16S, ou de l'analyse des génomes complets, suggère fortement qu'un gène *mntH* A soit apparu avant la séparation des espèces à Gram positif et des espèces à Gram négatif. Puisque les protéines MntH B sont les plus éloignées du reste de la famille Nramp, leur origine pourrait être ancestrale. Un gène *mntH* A représente donc un candidat plus probable à l'origine du gène Nramp eucaryote (Richer *et al.*, 2003 (2)).

Plus récemment, une analyse phylogénétique approfondie incluant des gènes issus du séquençage génomique d'espèces eucaryotes diverses, indique que les protéines eucaryotiques Nramp se séparent en deux sous groupes soit les Nramp "prototype" et "archétype". Dans les "prototypes", on retrouve des séquences qui sont proches les une des autres et des homologues bactériens, comme par exemple Smf1-3p chez la levure. Les Nramp "archétypes" se retrouvent majoritairement chez

les plantes et les mammifères; les vertébrés n'ont conservé que les Nramp "archétype" (Richer *et al.*, 2003 (2)).

Les protéines MntH appartenant au groupe C, démontrent une plus grande similarité aux protéines Nramp eucaryotes qu'à des protéines bactériennes MntH B et A. De plus, des analyses phylogéniques basées sur les séquences du groupe MntH C sont en désaccord avec la phylogénie connue des espèces dont elles dérivent, ce qui indique la possibilité de transferts horizontaux de gènes *mntH C* entre espèces éloignées. Trois sous-groupes de séquences sont distingués (C α , C β et C γ) (Cellier *et al.*, 2001 (1-2); Richer *et al.*, 2003 (2)). Plusieurs séquences se retrouvant dans le sous-groupe C α démontrent une très forte conservation de leurs séquences avec des protéines eucaryotes, suggérant la possibilité d'un transfert horizontal de gène *Nramp* eucaryote vers un génome bactérien (Richer *et al.*, 2003 (2)). À l'intérieur du sous-groupe C β , on retrouve majoritairement des bactéries à Gram positif (avec un taux de nucléotide GC faible) comme les Staphylococcaceae et des bactéries lactiques. Ce sous groupe inclus cependant la séquence d'une entérobactérie qui est un endosymbiote obligatoire, possédant également un génome riche en nucléotides AT. Le dernier sous-groupe, (C γ) est hétérogène, présentant des similarités avec des homologues du groupe C α et A.

À l'intérieur d'une même espèce bactérienne, on peut retrouver deux gènes *mntH* appartenant à des groupes ou sous-groupes différents comme par exemple un gène du groupe A et un du groupe C γ . Ce qui indique que, au cours de l'évolution, une espèce bactérienne pourrait avoir acquis ces gènes soit par duplication soit en possédant déjà un gène et en acquérant le second par transfert horizontal. Au cours de cette évolution complexe, il est possible que leur activité et/ou leur mode de régulation ait changé.

Plusieurs hypothèses sur l'évolution des homologues Nramp/MntH sont proposées. Tout d'abord, MntH B serait apparu dans des bactéries anaérobiques, lorsque l'atmosphère contenait très peu d'oxygène. Par la suite, MntH A serait apparu dans des bactéries aérobiques avant l'émergence des Gram positif et Gram négatif, dans un environnement adéquat pour leur développement. Puis les eucaryotes auraient

acquis un gène du groupe A lors d'un évènement endosymbiotique, donnant naissance au "prototype" Nramp. Chez les eucaryotes, il y a eu duplication du gène ancestral générant les Nramp "prototype" et "archetype" et ce avant la séparation des plantes, animaux et Mycetozoa (Dictyostelium). Dernièrement, un Nramp "prototype" pourrait avoir donné naissance à un gène du groupe MntH C par transfert horizontal, et les différents sous groupes de MntH C pourraient découler de différents transferts (Richer *et al.*, 2003 (2)).

4.3.2 Caractérisation des homologues bactériens

Des homologues bactériens ont été caractérisés chez plusieurs espèces à Gram positif ou à Gram négatif (Cellier *et al.*, 2001 (1-2)). La plupart des homologues caractérisés appartient au groupe A telle que la protéine MntH de *E. coli* (EcoliA). EcoliA a été démontré comme étant un transporteur, de type symporteur, dépendant du proton, facilitant l'acquisition des cations divalents avec une meilleure affinité pour le manganèse. Il permet également le passage du fer, du cadmium, du cobalt, du zinc et du cuivre (Makui *et al.*, 2000). EcoliA est réprimé par la protéine Fur (ferric uptake regulator) en présence de fer et par MntR (Mn²⁺ transport repressor) en présence de manganèse (Patzner et Hantke, 2001).

L'homologue MntH retrouvé chez *Bacillus subtilis* (BsubA), est également un transporteur de manganèse dépendant du proton (H⁺) qui est nécessaire pour la croissance en milieu minimum. La perte du gène peut être remplacée par l'addition de manganèse dans le milieu minimum. BsubA confère également une sensibilité au cadmium et est nécessaire pour la résistance au peroxyde d'hydrogène (H₂O₂). Tout comme son homologue EcoliA, il est réprimé en présence de manganèse par MntR (Que et Helmann, 2000 (1-2)). La protéine de *Mycobacterium tuberculosis* facilite également le transport de certains cations (Agranoff *et al.*, 1999 (1-2)), mais ne semble pas jouer un rôle important dans la virulence de cette espèce (Domenech *et al.*, 2002).

Chez *Salmonella typhimurium*, MntHA joue également le rôle de transporteur de manganèse dépendant du proton et est nécessaire pour la résistance au peroxyde d'hydrogène. Il est régulé par les métaux de la même manière que EcoliA; le rôle d'un troisième régulateur, l'activateur OxyR répondant au H₂O₂, a également été démontré (Kehres *et al.*, 2000 (1); Kehres *et al.*, 2002). Le rôle physiologique des transporteurs de fer et de manganèse n'est pas encore totalement compris. Un rôle majeur de *mntH A* dans la virulence de *S. typhimurium in vivo* n'a pas encore été démontré, bien que cela soit déjà le cas pour SitABCD et que le fer et le manganèse soient essentiels à la pathogénèse de *S. typhimurium* (Boyer *et al.*, 2002 (2)).

L'homologue MntH de *Staphylococcus aureus* fait partie du sous-groupe Cβ (SaurCβ). Contrairement aux homologues précédents, SaurCβ serait régulé positivement par MntR. Cette différence au niveau de la régulation des gènes pourrait s'expliquer par le fait que SaurCβ soit éloigné de BsubA et qu'il aie été acquis par transfert horizontal. SaurCβ est également impliqué dans la virulence de *S. aureus* conjointement avec un autre transporteur de manganèse MntABC(D), homologue de SitABCD (Horsburgh *et al.*, 2002 (2)).

L'hypothèse du transfert horizontal d'un gène *Nramp* eucaryote suppose un gain fonctionnel par la bactérie receveuse lui conférant un avantage dans certaines conditions. Si les protéines MntH A, B et C ont apparemment conservé des modes d'action et des substrats similaires, il serait logique que les différences de séquences entre les groupes correspondent à des variations de structure (ex : arrangement/contact des segments transmembranaires) et/ou de mécanisme d'action et de régulation. L'adaptation de la protéine Nramp à la cellule eucaryote pourrait avoir permis son utilisation dans la lutte contre les pathogènes intracellulaires, en facilitant l'acquisition des métaux divalents par la cellule hôte au détriment des bactéries phagocytées. Les protéines MntH C pourraient présenter certaines propriétés avantageuses d'origine eucaryote, permettant aux bactéries qui les exprimeraient d'être plus performantes lors de l'infection de l'hôte ou vis à vis d'autres bactéries en compétition pour coloniser la même niche (Richer *et al.*, 2003 (2)).

5 Etude de la relation structure/fonction des transporteurs Nramp

Pour comprendre le fonctionnement d'une protéine, soit globulaire ou membranaire, il est nécessaire d'avoir une bonne appréhension de sa structure et de son mécanisme d'action. Pour les protéines de type intégralement membranaire, il s'agit d'un défi important de recherche en biologie moléculaire. Cette difficulté est attribuée à leur caractère hautement hydrophobe car il est difficile de recréer leur environnement natif, la bicouche lipidique. À ce jour, très peu de protéines membranaires ont été cristallisées (Booth *et al.*, 2001). L'importance des transporteurs appartenant à la famille Nramp/MntH pour le métabolisme de différents cations divalents et notamment l'homéostasie du fer et du manganèse chez l'homme, justifie l'effort nécessaire pour les caractériser. Dans le but à long terme d'envisager une approche thérapeutique pour les personnes ayant une carence au niveau de l'absorption des métaux, il est primordial d'avoir une bonne idée de la structure secondaire et tertiaire de ces transporteurs et de connaître les divergences qui ont pu apparaître au cours de l'évolution concernant leur affinité aux métaux et à leur mode de régulation (Forbes et Gros, 2003), notamment dans le cas des pathogènes bactériens exprimant des homologues MntH C.

5.1 Données structurales connues des différents homologues Nramp/MntH

5.1.1 Conservation des séquences dans la famille Nramp/MntH

La discrimination des différents membres appartenant à la famille Nramp/MntH se base principalement sur l'alignement des séquences et sur les données phylogénétiques qui en découlent (Cellier *et al.*, 2001 (1-2)). Ceci est illustré par le fait qu'un motif de séquence situé dans le 9^e segment transmembranaire prédit permet de différencier les protéines des groupes A, B, C des protéines Nramp eucaryotes. Les homologues appartenant aux groupes A possèdent un espacement de 5 résidus dans leur séquence aminée comparativement aux homologues du groupe B. Par contre, les séquences protéiques des membres du groupe C ne possèdent qu'un seul espacement, tout comme les séquences eucaryotes (Fig. 10).

Figure 10. Alignement de séquences des membres de la famille Nramp/MntH. La portion de l'alignement correspond au segment transmembranaire 9. La barre hachurée indique la région où est prédit le segment 9 et la zone rouge indique une région permettant la discrimination des séquences des trois groupes MntH A, B et C. Les noms des séquences sont constitués de l'initiale du nom d'espèce, et de deux à trois lettres du nom de genre, suivi de A, B, ou C pour les séquences des groupes MntH A, B ou C, et de NR pour les séquences Nramp eucaryotes. Des chiffres sont ajoutés dans le cas de plusieurs gènes présents dans un même organisme.

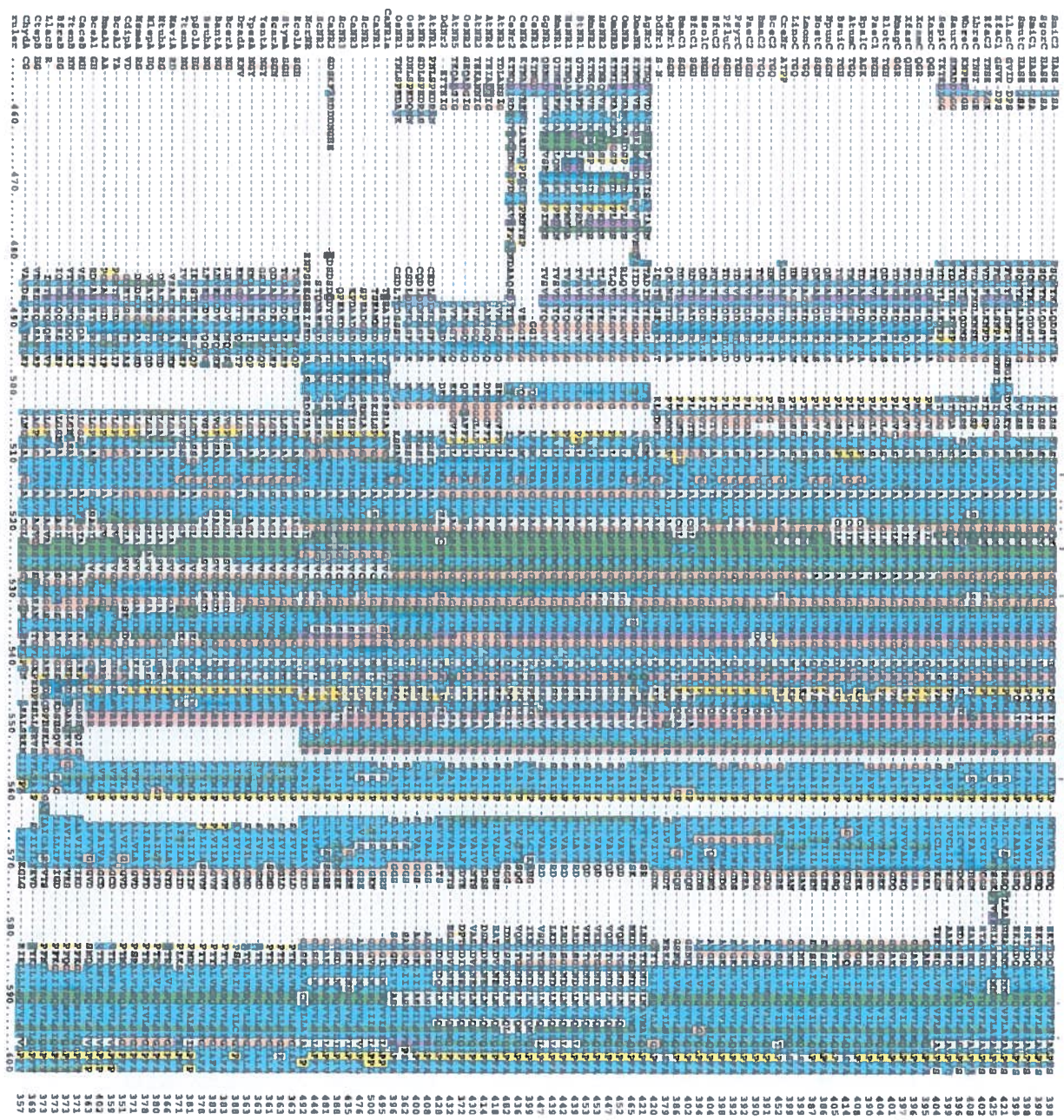


Figure 10

Le pourcentage d'identité à l'intérieur des différents groupes est remarquablement élevé. La protéine de levure Smf1p possède un taux d'identité de 52% avec Smf2p. Les homologues se retrouvant chez les plantes et les animaux démontrent une identité de 40% (Cellier *et al.*, 1995 (1-2)). Chez les bactéries par contre, les taux d'identité sont moins élevés, comme par exemple, le taux minimum de 36% d'identité trouvé dans le groupe B.

Une autre caractéristique de la famille Nramp/MntH est le caractère hautement hydrophobe démontré par les profils d'hydrophobicité très similaires entre les protéines des bactéries et des vertébrés (Richer *et al.*, 2003 (2)). Cette conservation se limite aux dix premiers segments hydrophobes qui constituent le cœur hydrophobe de la famille Nramp. Les protéines appartenant aux règnes eucaryotes possèdent apparemment 12 segments hydrophobes comparativement aux procaryotes qui en possèderaient seulement que 11. Cependant certains homologues comme *Enterococcus faecalis* (EfeaC β) et *Lactococcus lactis* (LlaC β) montrent un douzième segment transmembranaire possible sur leur profil d'hydrophobicité (Richer *et al.*, 2003 (2)). Le taux d'identité à l'intérieur de ce coeur hydrophobe est très élevé, exemple : 33 à 75% chez les eucaryotes (Cellier *et al.*, 1995 (1-2)). Par contre, la région couvrant les segments hydrophobes 8 à 10 a clairement subi au cours de l'évolution des changements de séquence susceptible d'affecter sa structure et sa fonction (Richer *et al.*, 2003 (2)).

5.2 Prédiction de la structure secondaire chez les protéines membranaires

Avant d'élucider certains aspects de la relation structure/fonction d'une protéine, il est important d'avoir une idée générale de sa structure. Un aspect important pour la structure des protéines de type intégralement membranaire est de connaître sa topologie; c'est-à-dire le nombre de segments transmembranaires ainsi que leur orientation par rapport à la membrane.

La première analyse consiste à représenter le profil d'hydrophobicité, en utilisant une fenêtre contenant entre 19 et 21 résidus. Cette largeur de fenêtre correspond au nombre d'acides aminés moyens pour que la taille du segment

hydrophobe permette de traverser la bicouche lipidique (Kyte et Doolittle, 1982 (2)). La présence de pics hautement hydrophobes, indique le nombre probable de segments membranaires. Il est également utile pour prédire la structure secondaire d'analyser la ressemblance de la séquence en acides aminés de la protéine avec des protéines de structure connue. Chez les protéines de type intégralement membranaire, le caractère physico-chimique des acides aminés joue un rôle crucial pour le bon arrangement des hélices transmembranaires. Par exemple, les résidus glycine, isoleucine et valine qui sont reconnus comme étant des résidus briseurs d'hélices α dans un environnement aqueux, sont favorables pour constituer des hélices transmembranaires dans un environnement lipidique (Ott et Lingappa, 2002). La séquence en acides aminés des protéines membranaires peut donc renseigner sur leur structure d'après la préférence des résidus pour un environnement donné. Les résidus non polaires ou hydrophobes, comme la leucine (L), la valine (V) l'isoleucine (I) et la méthionine (M) ont une préférence pour l'environnement lipidique d'une membrane. Par contre, les résidus chargés positivement, comme l'arginine (R) et la lysine (K) se retrouvent majoritairement dans les boucles situées du côté cytoplasmique de la membrane et peuvent servir de signal topogénique pour qu'un segment transmembranaire s'insère dans la bonne orientation par rapport à la membrane (von Heijne et Gavel, 1988). Le remplacement de ces résidus par des acides aminés non chargés positivement pourra conduire à l'inversion de l'orientation du segment membranaire (Gafvelin et von Heijne, 1994 (2)). La proline est considérée comme un résidu conférant une certaine flexibilité à l'hélice transmembranaire surtout lorsqu'elle se trouve à côté d'une glycine. Cette liberté est causée par une perte des liaisons hydrogène intra hélice (Cordes *et al.*, 2002). La présence du résidu proline au centre d'une longue hélice transmembranaire, permet de générer deux courtes hélices (Nilsson *et al.*, 1998). Ceci favorise les contacts inter-hélices nécessaires pour la mobilité de la protéine transmembranaire.

Après avoir vérifié le nombre de segments transmembranaires prédits à l'aide du profil d'hydrophobicité et étudié leur composition en acides aminés, il est important de connaître l'orientation de chacun des segments transmembranaires prédits et des résidus les délimitant. Pour ce faire, on utilise différents algorithmes pour prédire la topologie transmembranaire de la protéine incluant le nombre et l'orientation des

segments. Certains programmes ont une efficacité d'environ 85% comme TMHMM. Ce programme utilise la probabilité de chacun des 20 acides aminés à se retrouver dans un environnement donné, en utilisant une configuration composée de 7 états possibles (cœur de l'hélice, les extrémités de l'hélice, petite boucle du côté cytoplasmique, les boucles du côté périplasmique courtes et longues ainsi qu'un domaine globulaire) (Krogh *et al.*, 2001 (2)). Un programme différent et tout aussi efficace comme MEMSAT, est lui basé sur 5 états possibles (boucle intérieure, le début, le milieu et la fin de l'hélice ainsi que la boucle extérieure) et détermine la fréquence des 20 acides aminés à un endroit en utilisant une table de substitution possible basée sur différentes protéines intégralement membranaires de structures connues. Il permet également de générer une série de prédictions incluant tous les segments membranaires possibles (Jones, Taylor et Thornton, 1994 (1-2)). Le programme PHD a l'avantage d'effectuer une analyse utilisant un alignement de séquences homologues, permettant d'inclure dans la prédiction l'information déduite du remplacement en acides aminés observés entre protéines appartenant au même groupe phylogénétique (Rost, Fariselli et Casadion, 1996). Pour améliorer encore la prédiction, on peut déduire un consensus à partir des prédictions générées par différents algorithmes (Melen, Krogh et von Heijne, 2003 (2); Lao, Okuno et Shimizu, 2001 (2); Moller, Croning et Apweiler, 2001 (2)). La totalité de ces analyses basées sur différents principes aura comme avantage de renforcer le modèle prédit.

5.3 De la prédiction à la vérification expérimentale

5.3.1 Les fusions et insertions dirigées

Pour valider la topologie des protéines membranaires, la localisation de chacun des segments hydrophobes ainsi que les boucles servant à lier les différentes hélices doivent être vérifiées. Une première approche est de cibler des insertions à différents endroits de la protéine pour vérifier leur accessibilité de part et d'autre de la membrane en utilisant des réactifs imperméables à la membrane. Ces sites d'insertions peuvent correspondre à des sites de glycosylation, un épitope reconnu par un anticorps ou bien un site reconnu par une protéase (van Geest et Lolkema, 2000 (2)). Malgré leurs effets peu invasifs sur la protéine, ces méthodes possèdent comme inconvénient

qu'il faut posséder une idée préalable de la topologie globale pour bien choisir les sites d'insertions.

Pour nous donner une idée générale de l'endroit des segments hydrophobes, l'approche des gènes rapporteurs couramment utilisés pour générer un modèle topologique des protéines bactériennes, a été employée. Cette approche utilise des rapporteurs (enzymes) dont l'activité dépend de la localisation. Le gène rapporteur peut être fusionné en différents endroits du gène étudié (*mntH*), ayant comme effet de substituer la portion C-terminale du gène par le rapporteur. Ceci a comme inconvénient de travailler avec une protéine tronquée et non fonctionnelle. En revanche, l'insertion du rapporteur à un endroit ciblé de la protéine, permet de travailler sur une protéine entière et peut-être fonctionnelle.

L'enzyme rapporteur la plus couramment utilisée est la phosphatase alcaline codée par le gène *phoA* chez *E. coli* (Boyd *et al.*, 1993; Manoil et Beckwith, 1986; Traxler *et al.*, 1993). Cette enzyme est active seulement du côté périplasmique où elle peut se reposer efficacement et être totalement fonctionnelle. Lorsque PhoA mature (sans son peptide signal) est fusionnée dans une boucle de l'environnement cytoplasmique, elle est inactive. Par contre, lorsque la fusion se trouve du côté périplasmique l'activité phosphatase peut être facilement détectée. Cependant à proximité de l'extrémité N-terminale de la protéine, une activité élevée peut être détectée même si la fusion est située du côté cytoplasmique (Manoil et Beckwith, 1986).

Le deuxième gène rapporteur largement utilisé pour des études topologiques, est le gène codant pour la β -lactamase (Blam) dépourvu de son peptide signal. (Broome-Smith *et al.*, 1990). Ce dernier permet à la bactérie de croître à de forte concentration d'ampicilline (>45 μ g/mL) lorsque la fusion est située dans une boucle de l'environnement périplasmique. Le degré d'exportation à travers la membrane plasmique peut être vérifié en regardant la concentration minimale inhibitrice (CMI) d'ampicilline (Traxler *et al.*, 1993). Ce rapporteur a l'avantage d'être facile à cribler et est actif même lorsque les points de jonctions sont situés près de l'extrémité N-terminale de la protéine d'intérêt. De plus, il est utilisable en insertion ciblée.

Ces deux rapporteurs sont actifs seulement lorsqu'ils se retrouvent du côté périplasmique. Le criblage des boucles prédites du côté périplasmique par la méthode des fusions peut générer un certain nombre de faux positifs (boucle prédite cytoplasmique démontrant une activité avec les rapporteurs (Blam, PhoA). Si la fusion se situe en amont d'un signal topologique, composé principalement de résidus chargés positivement, la boucle hydrophile cytoplasmique sera moins stable et pourra être exposée plus facilement du côté du périplasme, résultant en une activité du gène rapporteur pouvant être contraire à la prédiction (Traxler *et al.*, 1993; van Geest et Lolkema, 2000).

Pour générer un modèle le plus valable possible, il faut compléter les données générées du côté périplasmique en utilisant un gène rapporteur actif lorsque fusionné dans une boucle exposée du côté cytoplasmique. Deux rapporteurs, la β -galactosidase (*lacZ*) (Froshauer *et al.*, 1988) et la chloramphénicol acétyl transférase (*cat*) (Zelazny et Bibi, 1996 (2)) sont actifs uniquement lorsqu'ils se retrouvent dans un environnement cytoplasmique. Chacun de ces gènes rapporteurs cytoplasmiques a été utilisés pour compléter les données obtenues avec les rapporteurs périplasmiques, PhoA et Blam (Zelazny et Bibi, 1996 (2); Adler et Bibi, 2002 (2), Hirata *et al.*, 1998 (2); Blank et Donnenberg, 2001; Ethier et Boyd, 2000).

Une alternative pour établir avec plus de confiance la topologie consiste à insérer le rapporteur tout en préservant l'activité fonctionnelle de la protéine étudiée. Ainsi, une nouvelle méthode rapide a été proposée pour vérifier la topologie d'une protéine membranaire chez *E. coli*. Il s'agit de prédire la structure de la protéine avec plusieurs algorithmes et de fusionner l'extrémité C-terminale, avec un rapporteur périplasmique et cytoplasmique, par exemple les gènes *phoA* et la *gfp* (green fluorescent protein) respectivement. Cependant, selon Drew *et al.*, (Drew *et al.*, 2002 (2)) il est possible d'obtenir un modèle juste que pour environ 10% des quelques 800 protéines membranaires situées dans la membrane interne de *E. coli*.

Les approches décrites utilisées de concert permettent d'établir un modèle global pour le nombre de segments transmembranaires et vérifier leur orientation. Pour obtenir une idée plus précise sur les résidus délimitant chacune des hélices et

ceux impliqués d'un point de vue structural et/ou fonctionnel, il faut utiliser une approche dite chimique.

5.3.2 Balayage à la cystéine

La cristallisation des protéines intégralement membranaire est très difficile (Booth *et al.*, 2001). L'approche de prédilection pour obtenir une idée précise de la structure tertiaire des protéines membranaires est de remplacer chacun des résidus par l'acide aminé cystéine (balayage à la cystéine) (Sorgen *et al.*, 2002). Ce résidu a la caractéristique d'être très réactif et peut être introduit à plusieurs endroits sans avoir d'effet néfaste sur la fonction (Traxler *et al.*, 1993; van Geest et Lolkema, 2000 (2)). Un pré requis pour effectuer un balayage à cystéine est de remplacer tous les résidus cystéines endogènes par des résidus peu dommageables comme l'alanine ou la sérine. Une fois que la protéine ne contient qu'une seule cystéine dans la position désirée, la méthode consiste à faire réagir cette cystéine avec différents réactifs spécifiques des groupements sulfhydryles. Ces produits possèdent des tailles différentes et diffèrent par leur degré de perméabilité à la membrane. Un grand nombre de réactifs sulfhydryles existent, comprenant des groupements soit fluorescents, radioactifs ou encore, biotinylé. Chacun de ces marqueurs peuvent être détectés et ainsi démontrer la localisation de la cystéine dans la position étudiée. Le réactif le plus couramment utilisé est le NEM (N-éthylmaleimide), qui est de petite taille et suffisamment hydrophobe pour être perméable à la membrane. Il permet de localiser les résidus dans leur environnement local et spécifique au sein de la structure tertiaire. La réactivité du NEM peut être bloquée en pré-traitant la protéine par le MTSES (2-sulfonatoethylmethane thiosulfonate), un petit réactif hydrophile et imperméable, ce qui permet d'établir si l'accessibilité de la cystéine étudiée correspond plutôt à un environnement aqueux (Riordan et Vallee, 1972). Cette approche serait avantageuse chez la protéine EcolIA car elle ne possède aucune cystéine intrinsèque.

Tous les résidus des protéines TetAB (Tamura *et al.*, 2001) et la lactose perméase (Frillingos *et al.*, 1998), ont été remplacés par une cystéine et leur réactivité a été analysée en utilisant le N-éthylmaleimide (NEM; Riordan et Vallee, 1972). Ces

protéines, assez différentes par leur séquence d'acide aminé, leur spécificité de substrat et leur mécanisme d'action (H⁺-symporteur ou -antiporteur), possèdent néanmoins une structure tridimensionnelle fortement similaire. C'est à dire que quatre hélices sont présumées se retrouver dans un environnement totalement hydrophobe et forment la partie extérieure du canal transmembranaire (réactivité minimale au NEM sur toute leur longueur); quatre autres sont situés dans la région intérieure du canal et démontrent un caractère amphiphile (une face réagissant fortement au NEM sur toute leur longueur) et les quatre derniers segments, montrant qu'une moitié de leur hélice est située dans l'environnement hydrophobe et l'autre moitié dans l'environnement hydrophile (réactivité au NEM intermédiaire, une moitié inaccessible au NEM, et une moitié présentant une face accessible au NEM). Ce patron de réactivité peut s'expliquer par le fait que les hélices forment un angle dans la membrane. Chez les deux protéines, les groupes ont découvert que très peu de résidus sont essentiels pour le bon fonctionnement; 6 chez la lactose perméase (Frillingos *et al.*, 1998) et 17 pour l'antiporteur de la tétracycline (Tamura *et al.*, 2001).

5.4 Résultats topologiques chez les homologues Nramp

À ce jour, aucune étude exhaustive de la structure en général n'a été effectuée chez un homologue appartenant à la grande famille Nramp; les données connues sont très fragmentaires. Chez la levure, un épitope HA (Hémagglutinine) a été introduit en amont et en aval du 11^e segment transmembranaire de la protéine Smf3p qui est localisée dans la membrane de la vacuole. La protéolyse de l'épitope C-terminal seulement démontre que l'extrémité C-terminale de la protéine est située du côté externe de la membrane et que le onzième et dernier segment est orienté du cytoplasme vers le côté extracellulaire (intravacuolaire). Ces constructions, n'altèrent ni la fonction ni la localisation de Smf3p à l'intérieur de la cellule (Portnoy, Liu et Culotta, 2000 (2)).

En ce qui concerne l'homologue Nramp2, un groupe a démontré que la quatrième boucle périplasmique située entre les segments membranaires 7 et 8 est située du côté extracellulaire (Picard *et al.*, 2000 (2)). Ce même groupe a introduit dans la boucle correspondante de l'homologue Nramp1 un épitope HA et a démontré

que cette insertion était située du côté extracellulaire, par contre, la protéine se retrouvait exprimée dans la membrane plasmique au lieu des membranes lyso-et phagosomales (Forbes et Gros, 2003). Les deux extrémités de Nramp1 ont été démontrées comme étant situées du côté cytoplasmique en insérant un épitope c-myc à l'extrémité C-terminale et en faisant réagir un sérum polyclonal contre l'extrémité N-terminale de la protéine (Vidal *et al.*, 1996 (2)). Dans les deux cas, la perméabilisation préalable de la membrane plasmique était nécessaire pour obtenir un signal *in situ*. Récemment, une étude a démontré que le quatrième segment hydrophobe de Nramp2, précisément entre les résidus V₁₈₇ à K₂₀₁ adoptait une conformation hélicoïdale dans des micelles en présence de détergent (Li *et al.*, 2003 (2)).

Une autre étude effectuée sur Nramp1 a montré qu'un sérum polyclonal dirigé contre la 4^e boucle extracellulaire inhibait l'association de fer radioactif avec des phagosomes purifiés, tandis qu'un sérum polyclonal dirigé contre l'extrémité C-terminale ne l'inhibait pas. Les auteurs ont interprété leurs résultats en proposant que la boucle entre les segments 7-8 est située du côté cytoplasmique (Kuhn, Lafuse et Zwilling, 2001 (2)). Cette étude suggérant que Nramp1 pourrait adopter une topologie inversée par rapport à Nramp2 est en contradiction avec les autres études effectuées avec des protéines de levure et Nramp1 (Forbes et Gros, 2003).

Concernant les homologues bactériens, une étude a proposé un modèle topologique pour les homologues du groupe A. La topologie est basée sur l'étude de prédictions et de la réactivité d'une cystéine unique (Cys299) de la protéine de *Mycobacterium leprae* avec le marqueur méthaneithiosulfonate. L'analyse par spectroscopie de résonance paramagnétique de l'interaction dipolaire entre le groupement nitroxyde-Cys299 et un métal paramagnétique (Cu), a montré que cette cystéine est située dans un environnement hydrophobe, et correspond au segment transmembranaire 10 selon ces auteurs (Reeve *et al.*, 2002 (2)).

Cette conclusion semble en contradiction avec les modèles déduits des analyses effectuées sur les homologues Nramp1, 2 et chez Smf3p et des prédictions basées sur l'analyse des séquences eucaryotes et procaryotes (Cellier *et al.*, 1995 (1-2); Richer *et al.*, 2003 (2)). Comme aucune autre étude ne traite de la topologie des

homologues MntH, nous avons décidé d'établir un modèle topologique en utilisant la protéine d'*E. coli*.

6. Objectifs de mon projet

Dans ce mémoire, sont présentés les résultats de recherche visant deux objectifs principaux.

- 1) Déterminer la topologie transmembranaire de la protéine MntH A de *E. coli*, en utilisant deux rapporteurs topologiques complémentaires, soit la β -lactamase (Blam) et la chloramphénicol acétyl transférase (Cat).
- 2) Effectuer la caractérisation préliminaire d'homologues appartenant aux groupes MntH B et C par clonage et expression dans le système hétérologue *E. coli*.

Discussion et conclusion

1. Prédiction d'un modèle topologique

L'utilisation de plusieurs algorithmes de prédiction basés sur différentes approches a permis d'élaborer une prédiction consensus de la topologie de la protéine MntH d'*E. coli* (EcoliA) (traité plus en détail dans l'article 2). Cette prédiction indique 11 segments transmembranaires avec les extrémités N- et C-terminales du côté cytoplasmique et périplasmique respectivement (Fig. 4 article 2). Ces données concordent avec le profil d'hydrophobicité qui démontre 11 segments hydrophobes. Cette prédiction est en accord avec le model proposé pour l'homologue Smf3p qui possède également 11 segments membranaires.

2. Analyse des profils d'hydrophobicité

Étant donné la conservation des profils d'hydrophobicité observés avec les protéines eucaryotes Nramp, il nous a paru intéressant d'étudier les profils représentant de nombreux groupes phylogénétiques identifiés MntH B, C et "prototype" Nramp. La plupart des profils analysés ont démontré un nombre de segments hydrophobes attendus pour les homologues procaryotes et eucaryotes, 11 et 12, respectivement. Par contre, plusieurs homologues du sous groupe C β présenteraient un profil d'hydrophobicité à 12 segments membranaires. Cette analyse a permis d'observer que des protéines appartenant au sous groupe MntH C α (ex: *Nostoc punctiforme* (NpunC α)) démontrent un profil ressemblant d'avantage aux protéines eucaryotes Nramp "prototype" (*Dictyostelium discoideum* (DdNR1) et *Chlamydomonas reinhardtii* (CreNR)) qu'aux protéines MntH A et B (Fig. 2 article 1). Ce résultat est en accord avec les distances phylogénétiques mesurées entre les membres des groupes C α et "prototype" Nramp d'une part, et celles mesurées entre ces groupes et les groupes A et B d'autre part.

3. Fusion aléatoire avec le gène rapporteur Blam (β -lactamase)

Puisque le modèle prédit n'acquiert sa valeur qu'une fois après avoir été validé expérimentalement, j'ai en premier lieu utilisé une méthode permettant de générer des données sans hypothèse préalable, par fusion aléatoire. Le choix de séquences des

clones démontrant une forte résistance à l'ampicilline, correspond à la plus grande chance d'être dans le bon cadre de lecture; la sélection par la taille des clones a permis de couvrir toute la molécule, soit d'indiquer les boucles intracytoplasmiques 2/3, 6/7, 8/9, 10/11 et les boucles extrapériplasmiques 1/2, 5/6, 7/8, 9/10 et l'extrémité C-terminale qui sont en accord avec la prédiction. Par contre, 3 constructions démontraient une forte résistance à l'ampicilline malgré leur localisation prédite cytoplasmique (Tableau 4, article 2). L'observation de la séquence après les points de fusions a révélé la présence de 3 résidus basiques. La série d'acide aminé, K R v K, pourrait correspondre à un signal topologique et son absence résulterait en une insertion incorrecte du 10^e segment membranaire dans la membrane plasmique.

Afin de déterminer si la résistance à l'ampicilline était le résultat des protéines fusionnées, j'ai vérifié le niveau d'expression des différentes constructions par immunobuvardage de type Western, en utilisant un anticorps dirigé contre la β -lactamase. Toutes les constructions, à l'exception de V350 démontrent une bande à la taille attendue. Par contre, tous les mutants sauf (A165, F248) démontraient également la présence d'une forme de Blam sécrétée. Ceci démontre donc que les régions proches des points de jonction sont sujettes à protéolyse après avoir atteint le périplasme (Fig. 6 2A, article 2). Uniquement pour A165, le niveau d'expression était corrélé avec la résistance à l'ampicilline (45 μ g/mL).

4. Vérification de la topologie par la fonctionnalité de la protéine étudiée

L'obtention des fusions ne démontre pas la topologie de la protéine fonctionnelle, mais les données recueillies ont permis de cibler des régions pour l'insertion de la Blam, dans le but de préserver une protéine MntH fonctionnelle. Trois sites ont été choisis d'après les prédictions et l'alignement des séquences qui correspondent aux deux plus grosses boucles périplasmiques (entre les segments 5/6 et 7/8) ainsi qu'à l'extrémité C-terminale. Cette extrémité est considérée comme essentielle pour toute étude topologique pour discriminer entre deux prédictions démontrant une localisation différente pour l'extrémité C-terminale (Drew *et al.*, 2002).

Chacune de ces protéines dites hybrides, conférait une forte résistance à l'ampicilline, confirmant leur localisation du côté périplasmique. Par contre, ces protéines chimériques se distinguaient par leur niveau de fonctionnalité résiduelle. L'insertion après le dernier résidu (L412) a démontré une activité de niveau comparable à celui du contrôle positif et d'un mutant obtenu par l'ajout d'un épitope c-myc après le résidu L412. (Tableau 4, article 2).

La protéine contenant une Blam insérée entre les segments membranaires 6 et 7 a démontré une sensibilité aux métaux divalents similaire à EcoliA par contre, le gène était incapable de compléter totalement la souche SL93 *HflB1ΔmntH*, dépendante des cations divalents pour croître à la température non permissive, 42°C (Herman *et al.*, 1995) (Tableau 4, article 2). Cependant, l'observation d'un transport de protons facilité par cette protéine hybride et dépendant d'un cation substrat pour MntH implique un degré de fonctionnalité. Un épitope HA inséré entre les mêmes segments chez Nramp2, démontre également une localisation extracellulaire (Picard *et al.*, 2000). Ce résultat montre que cette boucle est localisée dans un environnement analogue correspondant aux protéines procaryotes (périplasmique) et eucaryote (extracellulaire).

En revanche, l'insertion dans la boucle périplasmique entre les segments transmembranaires 5/6 a profondément altéré l'expression et la fonction de MntH, évaluées par immunobuvardage et test de sensibilité au Mn^{2+} et au Fe^{2+} (Tableau 4, article 2). L'élimination de la Blam, générant un di-peptide GS au site d'insertion a démontré une baisse de sensibilité aux cations divalents, ce qui suggère que cette portion est intolérante aux insertions. On peut donc déduire que cette région a un rôle structurel ou fonctionnel dans la protéine EcoliA (Tableau 4, article 2).

5. Utilisation du rapporteur Cat (chloramphénicol acétyl transférase)

Étant donné l'obtention de fusion Blam contredisant le modèle, nous avons également voulu vérifier des boucles prédites cytoplasmiques en utilisant le gène rapporteur *cat*. Les cinq fusions dirigées correspondant aux boucles entre les segments hydrophobes 2/3, 6/7, 8/9, 10/11 et à l'extrémité C-terminale de la protéine étaient

toutes positives, sauf la dernière, pour la résistance au chloramphénicol (Tableau 5, article 2). Ces résultats ont complété ceux obtenus avec la Blam et a permis d'éliminer les faux positifs.

Le niveau d'expression des fusions Cat ont été détectés immunobuvardage de type western sur des préparations membranaires en utilisant un anticorps dirigés contre ce rapporteur. Le taux d'expression observé ne correspond pas avec le niveau de résistance à l'antibiotique concerné (R91-Cat); ceci a été formellement observé en effectuant une expérience de marquage métabolique au ^{35}S -Met. La fusion est instable, ne s'accumule pas et est difficilement détectée par immunobuvardage; par contre, elle confère une forte résistance au chloramphénicol (R91-Cat) (Tableau 5 et supplemental data 2 de l' article 2).

En conclusion, les résultats concernant la topologie de la protéine fonctionnelle EcoliA démontrent 11 segments membranaires avec les extrémités N et C terminale localisées du côté cytoplasmique et périplasmique respectivement.

6. Caractérisation d'homologues MntH appartenant aux groupes B et C

La famille Nramp/MntH démontre une évolution complexe. Nous avons donc commencé à caractériser certains homologues bactériens appartenant à différents groupes phylogénétiques en les surexprimant dans un système hétérologue (*E. coli*) et en les testant pour leur sensibilité aux cations divalents. On a observé que plusieurs homologues démontraient un certain niveau de sensibilité suggérant une conservation fonctionnelle.

Un résultat inattendu a été obtenu et mérite une investigation plus poussée concernant EfaC β 1. Nous avons exprimé deux versions de la protéine, avec une extrémité N-terminale tronquée et une version longue. Seulement la version longue a démontré une sensibilité au fer, cobalt et cadmium; indiquant ainsi que la queue N-terminale de EfaC β 1 est importante pour son fonctionnement. Un alignement de séquence avec différents homologues du groupe C β démontre une région conservée de 14 résidus en amont du 1^{er} segment membranaire. Cette région, conservée

seulement parmi les homologues du groupe C β , suggère une spécificité structurale ou fonctionnelle possible.

Une caractérisation préliminaire des différents homologues MntH démontre que certaines protéines sont exprimées et adoptent une topologie fonctionnelle au niveau de la membrane interne chez *E. coli*.

7. Problématique et perspective

La compréhension du mécanisme d'action d'une protéine intégralement transmembranaire requiert une connaissance de sa structure. La détermination de la topologie de EcolIA est la première étape requise pour étudier la relation structure/fonction de ces protéines. Les données fonctionnelles des protéines hybrides nous permettent d'établir de façon solide la topologie de la protéine dans la membrane. L'insertion de rapporteurs à l'extrémité C-terminale a permis de démontrer qu'il est exposé dans le périplasma et l'insertion dans la boucle périplasmique 6/7 a permis de préserver un excellent degré de fonctionnalité, démontré par le transport des métaux et le transport des protons.

La caractérisation fonctionnelle de différents homologues MntH nous indique qu'au cours de l'évolution, les protéines ont conservé une fonction similaire. En se basant sur la très grande conservation des profils d'hydrophobicité et de l'alignement des séquences, l'étude sur la topologie de EcolIA pourra servir comme modèle pour les différents membres appartenant à la vaste famille Nramp. Une étude préalable devra être effectuée chez des homologues du groupe B et C pour s'assurer que cette topologie est conservée.

Ceci ouvre des perspectives intéressantes visant à identifier les segments qui font face à l'intérieur de la protéine et/ou qui sont impliqués dans les interactions inter-hélices. Pour arriver à cette fin, on pourra planifier un balayage à la cystéine sur des segments hydrophobes définis dans la présente étude. Les résidus mutés devront être soigneusement choisis en se basant sur l'analyse de la conservation des résidus à l'intérieur d'un même segment membranaire.

Remerciements

Je tiens premièrement à remercier mon directeur de recherche, le docteur Mathieu Cellier, pour m'avoir permis d'effectuer ma maîtrise et de m'avoir bien dirigé tout au long de cette longue et enrichissante maîtrise. Je voudrais également remercier tout les membres du laboratoire, Isabelle Bergevin, Étienne Richer, Élise Boyer, Evelyn Castillo, Frédéric Veyrier ainsi que Roman Chaloupka, pour m'avoir aidé et supporté tout au long de cette aventure.

Je voudrais remercier les correcteurs : François Shareck, Mathieu Cellier et Hervé Le Moual, pour avoir pris le temps de corriger mon mémoire et d'émettre des commentaires constructif.

J'adresse un très gros merci au technicien du laboratoire, Jacques Beaubien, qui m'a permis de me surpasser autant du point de vue académique que personnel et pour ses précieux conseils.

Je remercie également ma copine Chantal pour m'avoir encouragé et supporté durant les deux dernières années. Je remercie également mes amis Steve et Karl ainsi que mon frère Stéphane pour leur aide.

En dernier lieu, et non le moindre, je tiens à remercier sincèrement mes parents, Luc et Nicole Courville, pour m'avoir encouragé et aidé financièrement tout au long de mes longues études.

Merci à tous ceux qui ont pris le temps de lire ce mémoire.

Bibliographie

Archibald, F. (1986) Manganese: its acquisition by and function in the lactic acid bacteria. *Crit Rev Microbiol* **13**: 63-109.

Archibald, F. S. et Fridovich, I. (1982a) Investigations of the state of the manganese in *Lactobacillus plantarum*. *Arch Biochem Biophys* **215**: 589-596.

Archibald, F. S. et Fridovich, I. (1982b) The scavenging of superoxide radical by manganous complexes: in vitro. *Arch Biochem Biophys* **214**: 452-463.

Booth, P. J., Templer, R. H., Meijberg, W., Allen, S. J., Curran, A. R., et Lorch, M. (2001) In vitro studies of membrane protein folding. *Crit Rev Biochem Mol Biol* **36**: 501-603.

Boyd, D., Traxler, B., et Beckwith, J. (1993) Analysis of the topology of a membrane protein by using a minimum number of alkaline phosphatase fusions. *J Bacteriol* **175**: 553-556.

Braun, V. et Hantke, K. (2001) Mechanisms of bacterial transport. In *Microbial transport systems*. Winkelmann G (ed.) Weinheim: Wiley-VCH, pp. 289-311.

Braun, V., Hantke, K., et Koster, W. (1998) Bacterial iron transport: mechanisms, genetics and regulation. In *Iron transport and storage in microorganisms, plants, and animals*. Sigel, A. et Sigel, H. (eds.) New York: Marcel Dekker, Inc., pp. 67-131.

Broome-Smith, J. K., Tadayyon, M., et Zhang, Y. (1990) Beta-lactamase as a probe of membrane protein assembly and protein export. *Mol Microbiol* **4**: 1637-1644.

Byers, B. R. et Arceneaux, J. E. (1998) Microbial iron transport: iron acquisition by pathogenic microorganisms. *Met Ions Biol Syst* **35**: 37-66.

Canonne-Hergaux, F., Fleming, M. D., Levy, J. E., Gauthier, S., Ralph, T., Picard, V., Andrews, N. C., et Gros, P. (2000) The Nramp2/DMT1 iron transporter is induced in the duodenum of microcytic anemia mk mice but is not properly targeted to the intestinal brush border. *Blood* **96**: 3964-3970.

Canonne-Hergaux, F. et Gros, P. (2002) Expression of the iron transporter DMT1 in kidney from normal and anemic mk mice. *Kidney Int* **62**: 147-156.

Canonne-Hergaux, F., Gruenheid, S., Ponka, P., et Gros, P. (1999) Cellular et subcellular localization of the Nramp2 iron transporter in the intestinal brush border and regulation by dietary iron. *Blood* **93**: 4406-4417.

Cellier, M. (2001) Bacterial genes controlling manganese accumulation. In *Microbial transport system*. Winkelmann G (ed.) Weinheim: Wiley-VCH, pp. 325-345.

Chasteen, N. (2003) Ferritin. Uptake, storage and release of iron. In *Iron transport and storage in microorganisms, plants and animals*. Sigel, A. et Sigel, H. (eds.) New York: Marcel Dekker Inc..

Cordes, F. S., Bright, J. N., et Sansom, M. S. (2002) Proline-induced distortions of transmembrane helices. *J Mol Biol* **323**: 951-960.

- Debus, R. J. (2000) The polypeptides of photosystem II and their influence on manganotyrosyl-based oxygen evolution. *Met Ions Biol Syst* **37**: 657-711.
- Dintilhac, A., Alloing, G., Granadel, C., et Claverys, J. P. (1997) Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol* **25**: 727-739.
- Domenech, P., Pym, A. S., Cellier, M., Barry, C. E., III, et Cole, S. T. (2002) Inactivation of the *Mycobacterium tuberculosis* Nramp orthologue (mntH) does not affect virulence in a mouse model of tuberculosis. *FEMS Microbiol Lett* **207**: 81-86.
- Eberstadt, M., Grdadolnik, S. G., Gemmecker, G., Kessler, H., Buhr, A., et Erni, B. (1996) Solution structure of the IIB domain of the glucose transporter of *Escherichia coli*. *Biochemistry* **35**: 11286-11292.
- Ethier, J. et Boyd, J. M. (2000) Topological analysis and role of the transmembrane domain in polar targeting of PilS, a *Pseudomonas aeruginosa* sensor kinase. *Mol Microbiol* **38**: 891-903.
- Forbes, J. R. et Gros, P. (2003) Iron, manganese, and cobalt transport by Nramp1 (Slc11a1) and Nramp2 (Slc11a2) expressed at the plasma membrane. *Blood* **102**: 1884-1892.
- Fridovich, I. (1978) The biology of oxygen radicals. *Science* **201**: 875-880.
- Frillingos, S., Sahin-Toth, M., Wu, J., et Kaback, H. R. (1998) Cys-scanning mutagenesis: a novel approach to structure function relationships in polytopic membrane proteins. *FASEB J* **12**: 1281-1299.
- Froshauer, S., Green, G. N., Boyd, D., McGovern, K., et Beckwith, J. (5-4-1988) Genetic analysis of the membrane insertion and topology of MalF, a cytoplasmic membrane protein of *Escherichia coli*. *J Mol Biol* **200**: 501-511.
- Goswami, T., Bhattacharjee, A., Babal, P., Searle, S., Moore, E., Li, M., et Blackwell, J. M. (2001) Natural-resistance-associated macrophage protein 1 is an H⁺/bivalent cation antiporter. *Biochem J* **354**: 511-519.
- Greie, J., Deckers-Hebestreit, G., et Altendorf, K. (2001) Energy-transducing ion pumps in bacteria: structure and fonction of ATP synthase. In *Microbial transport system*. Winkelmann G (ed.) Weinheim: Wiley-VCH, pp. 23-45.
- Gros, P., Skamene, E., et Forget, A. (1983) Cellular mechanisms of genetically controlled host resistance to *Mycobacterium bovis* (BCG). *J Immunol* **131**: 1966-1972.
- Hantke, K. (2003) Is the bacterial ferrous iron transporter FeoB a living fossil? *Trends Microbiol* **11**: 192-195.
- Hao, Z., Reiske, H. R., et Wilson, D. B. (1999) Characterization of cadmium uptake in *Lactobacillus plantarum* and isolation of cadmium and manganese uptake mutants. *Appl Environ Microbiol* **65**: 4741-4745.

- Jensen, L. T. et Culotta, V. C. (2002) Regulation of *Saccharomyces cerevisiae* FET4 by oxygen and iron. *J Mol Biol* **318**: 251-260.
- Kammler, M., Schon, C., et Hantke, K. (1993) Characterization of the ferrous iron uptake system of *Escherichia coli*. *J Bacteriol* **175**: 6212-6219.
- Kastaniotis, A. J. et Zitomer, R. S. (2000) Rox1 mediated repression. Oxygen dependent repression in yeast. *Adv Exp Med Biol* **475**: 185-195.
- Kehres, D. G., Janakiraman, A., Slauch, J. M., et Maguire, M. E. (2002) Regulation of *Salmonella enterica* serovar *Typhimurium* mntH transcription by H₂O₂, Fe²⁺, and Mn²⁺. *J Bacteriol* **184**: 3151-3158.
- Kehres, D. G. et Maguire, M. E. (2003) Emerging themes in manganese transport, biochemistry and pathogenesis in bacteria. *FEMS Microbiol Rev* **27**: 263-290.
- Liu, X. F., Supek, F., Nelson, N., et Culotta, V. C. (1997) Negative control of heavy metal uptake by the *Saccharomyces cerevisiae* BSD2 gene. *J Biol Chem* **272**: 11763-11769.
- Lodish, Baltimore, Berk, Zipursky, Matsudaira, et darnell (1997) *Biologie moléculaire de la cellule*. DeBoeck Université.
- Manoil, C. et Beckwith, J. (1986) A genetic approach to analyzing membrane protein topology. *Science* **233**: 1403-1408.
- Nelson, N. (16-8-1999) Metal ion transporters and homeostasis. *EMBO J* **18**: 4361-4371.
- Nilsson, I., Saaf, A., Whitley, P., Gafvelin, G., Waller, C., et von Heijne, G. (1998) Proline-induced disruption of a transmembrane alpha-helix in its natural environment. *J Mol Biol* **284**: 1165-1175.
- Ott, C. M. and Lingappa, V. R. (2002) Integral membrane protein biosynthesis: why topology is hard to predict. *J Cell Sci* **115**: 2003-2009.
- Patzer, S. I. et Hantke, K. (2001) Dual repression by Fe²⁺-Fur and Mn²⁺-MntR of the mntH gene, encoding an NRAMP-like Mn²⁺ transporter in *Escherichia coli*. *J Bacteriol* **183**: 4806-4813.
- Portnoy, M. E., Jensen, L. T., and Culotta, V. C. (2002) The distinct methods by which manganese and iron regulate the Nramp transporters in yeast. *Biochem J* **362**: 119-124.
- Posey, J. E. et Gherardini, F. C. (2000) Lack of a role for iron in the Lyme disease pathogen. *Science* **288**: 1651-1653.
- Riordan, JF et Vallee, BL (1972) reactions with N-ethylmaleimide and p-Mercuribenzoate. *Methods in enzymology* **25**: 449-456.

Rodrigues, V., Cheah, P. Y., Ray, K., et Chia, W. (1995) malvolio, the Drosophila homologue of mouse NRAMP-1 (Bcg), is expressed in macrophages and in the nervous system and is required for normal taste behaviour. *EMBO J* **14**: 3007-3020.

Rudolph, H. K., Antebi, A., Fink, G. R., Buckley, C. M., Dorman, T. E., LeVitre, J., Davidow, L. S., Mao, J. I., et Moir, D. T. (1989) The yeast secretory pathway is perturbed by mutations in PMR1, a member of a Ca²⁺ ATPase family. *Cell* **58**: 133-145.

Rutherford, J. C., Jaron, S., Ray, E., Brown, P. O., et Winge, D. R. (2001) A second iron-regulatory system in yeast independent of Aft1p. *Proc Natl Acad Sci U S A* **98**: 14322-14327.

Saier, M. H., (1999) A functional-phylogenetic system for the classification of transport proteins. *J Cell Biochem Suppl* **32-33**: 84-94.

Sorgen, P. L., Hu, Y., Guan, L., Kaback, H. R., et Girvin, M. E. (2002) An approach to membrane protein structure without crystals. *Proc Natl Acad Sci U S A* **99**: 14037-14040.

Takase, H., Nitantai, H., Hoshino, K., et Otani, T. (2000) Requirement of the *Pseudomonas aeruginosa* tonB gene for high-affinity iron acquisition and infection. *Infect Immun* **68**: 4498-4504.

Tamura, N., Konishi, S., Iwaki, S., Kimura-Someya, T., Nada, S., et Yamaguchi, A. (2001) Complete cysteine-scanning mutagenesis and site-directed chemical modification of the Tn10-encoded metal-tetracycline/H⁺ antiporter. *J Biol Chem* **276**: 20330-20339.

Traxler, B., Boyd, D., et Beckwith, J. (1993) The topological analysis of integral cytoplasmic membrane proteins. *J Membr Biol* **132**: 1-11.

van der Helm, D. et Chakraborty, R. (2001) Structures of siderophore receptors. In *Microbial transport systems*. Winkelmann G (ed.) Weinheim: Wiley-VCH, pp. 261-287.

Voet, D. et Voet, J. (1998) *Biochimie*. Paris: Freeman et Company.

von Heijne, G. et Gavel, Y. (1988) Topogenic signals in integral membrane proteins. *Eur J Biochem* **174**: 671-678.

Annexes

Annexe 1

Lettre concernant la contribution scientifique de Pascal Courville dans les articles :

- 1- Horizontal gene transfer of 'prototype' Nramp in bacteria (2003) *J. Mol. Evol.* 57: 1-14
- 2- Determination of the transmembrane topology of *Escherichia coli* Nramp ortholog (2003) soumis.

1. Je suis premier auteur à contribution égale avec Étienne Richer. J'ai préparé l'ADN génomique des souches *Enterococcus faecalis*, *Lactococcus lactis*, amplifié par PCR, cloné les gènes *mntH* correspondants et vérifié leur séquence avec les données des banques de données. J'ai déposé dans GenBank les séquences des gènes EfaC β 1 et EfaC β 2. Les numéros d'accès sont AF545506 (EfaC β 1) et AF545507 (EfaC β 2).

J'ai sous cloné tous les gènes *mntH* pour étudier leur activité. J'ai réalisé tous les disques essais permettant d'évaluer la fonction de tous les homologues testés dans cet article (tableau 1). J'ai effectué l'analyse des profils d'hydrophobicité des différentes protéines testées (Figure 2A, B). J'ai également contribué à la rédaction de ce manuscrit.



Pascal Courville




Étienne Richer



Mathieu Cellier

2- J'ai compilé et établi une prédiction consensus pour EcoliA. J'ai produit toutes les constructions que j'ai utilisées pour établir la topologie de MntH d'*E. coli* (tableaux 1 à 4). J'ai réalisé tous les disques essais pour évaluer l'activité des protéines hybrides. J'ai effectué toutes les préparations de membrane et les immunobuvardages de type western pour révéler les constructions exprimées. J'ai également rédigé plusieurs parties du manuscrit.




Pascal Courville



Roman Chaloupka



Frédéric Veyrier



Mathieu Cellier

Annexe 2

Article 1: Horizontal gene transfer of «prototype» Nramp in bacteria. (2003) Journal of molecular evolution. 57: 363-376.

Horizontal gene transfer of 'prototype' Nramp in Bacteria

Richer Etienne*, Courville Pascal*, Bergevin Isabelle and Cellier Mathieu FM§

INRS-Institut Armand-Frappier, 531, Bd des Prairies, Laval, PQ, CANADA H7V 1B7

* Equally contributing authors

§ Corresponding author :

531, boul des Prairies

Laval (Québec)

Canada
H7V 1B7

Phone : (450) 687-5010 ext. 4681; Fax : (450) 686-5501

E-mail : mathieu.cellier@inrs-iaf.quebec.ca

Running Title: Evolution of the Nramp family in Bacteria

Keywords:

Phylogenetic analyses; bacterial-eukaryotic protein similarity; metal permeases;

bacterial infection

ABSTRACT

Eukaryotic *Nramp* genes encode divalent metal ion permeases important for nutrition and resistance to microbial infection. Bacterial homologs encode proton-dependent transporters of manganese (MntH), and other divalent metal ions. Bacterial MntH were classified in three homology groups (A, B, C) and MntH C further subdivided in C α , C β , C γ . The proteins from *C. tepidum* (MntH B) and *E. faecalis* (MntH C β 1, 2), divergent in sequence and hydropathy profile, conferred increased metal sensitivity when expressed in *E. coli*, suggesting conservation of divalent metal transport function in MntH B and C. Several genomic evidence suggest horizontal gene transfer (HGT) of *mntHC* genes: i) The enterobacteria *Wigglesworthia mntH C β* gene is linked to an *Asn t-RNA*, and its sequence most conserved with Gram positive bacteria homologs; ii) All the C β genes identified in oral streptococcaceae are associated with different potentially mobile DNA elements; iii) *Lactococcus lactis* and *Burkholderia mallei* genomes contain an *mntH* gene prematurely terminated and a novel full-length *mntHC* gene; iv) Remarkable sequence relatedness between the unicellular alga *C. reinhardtii* 'prototype' *Nramp* and some MntH C α (e.g. *Nostoc* spp., *Listeria* spp.) suggests HGT between *Eukarya* and *Bacteria*. Other 'prototype' *Nramp* genes (intronless, encoding proteins strongly conserved with MntH A and B proteins) identified in invertebrates represent a possible source for transfer of *Nramp* genes toward opportunistic bacteria. This study demonstrates complex evolution of MntH in *Bacteria*. It is proposed that 'prototype' *Nramp* are ancestors of bacterial MntH C proteins, which could facilitate bacterial infection.

INTRODUCTION

A functional Nramp1 (Natural resistance-associated macrophage protein 1) is required in animals for natural resistance toward various intracellular pathogens. Nramp1 is active at the level of the phagosome in professional phagocytes (Govoni et al. 1999; Gruenheid et al. 1997; Hackam et al. 1998), where it is believed to restrict metal availability for ingested microbes (Jabado et al. 2000). Nramp2 is a similar protein, active in the plasmalemma and endocytic compartments to facilitate uptake of Fe(II) and Mn(II) in the brush border of the intestine, and as part of the Tf/TfR-dependent pathway in most other cells (Forbes et al. 2001). Absence or low level expression of mutant protein Nramp2 induces microcytic anemia (Fleming et al. 1997).

Nramp proteins define a family of permeases conserved in yeast and plants (Cellier et al. 1995) that use the transmembrane proton electrochemical gradient to facilitate cellular uptake of divalent cations (Gunshin et al. 1997; Chen et al. 1999). Functional complementation of yeast and *Drosophila* mutants was observed with mammalian Nramp 2 and Nramp1, respectively (D'Souza et al. 1999; Pinner et al. 1997). Some plant Nramp are also capable of complementing metal deficiency phenotypes in yeast (Curie et al. 2000; Thomine et al. 2000). The conservation of sequence and physico-chemical properties of Nramp proteins therefore indicate the preservation of a functional hydrophobic core, facilitating transmembrane co-transport of metal ion and proton.

Three divergent groups of bacterial Nramp homologs (MntH) were defined based on sequence similarity (Identity levels, presence of indels, Blast scores; Cellier et al.

2001). MntH group B contains few sequences, identified in organisms that are known as anaerobic: the Green sulfur bacterium *C. tepidum* ; from a related group, *Bacteroides fragilis*, and two species of low GC Gram positive (Firmicutes) Clostridiaceae, *Clostridium acetobutylicum* and *Thermoanaerobacter tengcongensis*. An *mntH* B sequence was also identified in the lactic acid bacterium *Lactococcus lactis*. Group B sequences show relatively low level conservation with each other and are also distantly related to other MntH and Nramp groups. For these reasons it was proposed that group B sequences were probably closer to the ancestor of the MntH/Nramp family (Cellier et al. 2001).

Group A MntH sequences are shared by Gram positive and negative organisms; they are more conserved than group B sequences, and pairwise relationships between MntH A sequences reflects well the phylogeny of the species from which they derive (based on analyses of 16 S RNA sequences, or whole genome content; Korbelt et al. 2002), with the exception of *Burkholderia* spp. MntH A proteins from Gram negative (*E. coli*, *S. typhimurium*) and Gram positive (*B. subtilis*, *M. tuberculosis*) were demonstrated to transport divalent metals, and Mn in particular (Agranoff et al. 1999; Kehres et al. 2000; Makui et al. 2000; Que and Helmann 2000). Similarly to Nramp, MntH A proteins are permeases dependent on the transmembrane proton electrochemical gradient and ΔpH , facilitating bacterial metal uptake. It is thus assumed that an *mntH* group A gene could have been transferred to the eukarotic genome consequently to the endosymbiotic process at the origin of mitochondria, and then evolved into *Nramp*.

MntH group C is more heterogenous, with protein sequences resembling more eukaryotic Nramp than MntH A or B sequences. Some MntH C sequences display similarity levels in disagreement with the phylogeny of the species from which sequences are derived. In addition, there is a higher proportion of infectious agents either symbiotic or pathogenic in species possessing *mntH C* genes vs *mntH A* or *B* genes. The most parsimonious explanation to the particularities of MntH C sequences is to invoke DNA transfer from eukaryote to prokaryote. In the context of bacterial infection there could be an advantage for bacteria to express Nramp-derived (= MntH group C) proteins that could be more efficient to compete with, or resist to host Nramp defences. Bacteria living in close association with eukaryotes could thus have acquired *Nramp* DNA by horizontal transfer. An alternative explanation would involve several gene duplications, subsequent gene divergence and/or deletions; it should also include horizontal DNA transfer to explain some sequence relationships observed within MntH group C, and seems less likely (Cellier et al. 2001).

In this study the possible pathways of MntH/Nramp molecular evolution were re-evaluated using more than thirty novel *mntH* gene sequences identified in *Bacteria*. The presence of three groups of bacterial MntH A, B, C was confirmed, and genes from the *Bacteria Enterococcus faecalis* (EfaC β 1, EfaC β 2), *Lactococcus lactis* (LlacB, LlacC β) and *Chlorobium tepidum* (CtepB) were cloned and expressed in *E. coli*, and tested for conferring increased sensitivity to divalent metal cations. A novel type of eukaryotic *Nramp* homolog was found in protists and simple eukaryotic models, denominated 'prototype' Nramp. Sequence and comparative genomic analyses of *mntH* genes provided several examples suggesting horizontal gene transfer of *mntH* group C genes.

METHODS

Sequence analyses.

Nramp/MntH homologs were identified by routine analyses of the Taxonomy Reports of TBLASTN analyses performed in available databases (NCBI, complete and unfinished genomes) with concatenated MntH sequences from *D. radiodurans* (MntH A) and *C. tepidum* (MntH B). The sequence of novel genes or corresponding proteins were retrieved from the databases and analyzed using Vector NTI Suite 7.0 (Informax, Bethesda MD USA). Full length Orf were aligned using the program ClustalX and multiple sequences alignment were analyzed with the package Molecular Evolution Genetic Analyses (MEGA version 2.1, Kumar et al. 2001). Pairwise deletion setting was chosen for the analyses since specific indels are useful to discriminate MntH groups. A Poisson correction was used for distance matrix calculations, and trees were generated by either the Neighbour-Joining or Minimal Evolution methods (Kumar et al. 2001); the statistical significance of each of the tree topologies obtained was tested by the Interior branch and Bootstrap tests. Trees were also generated using the Maximum Parsimony algorithm and bootstrap test. When genomic information was required to further characterize *mntH* loci, the nucleotide sequence flanking the gene were used in BlastN analyses to retrieve more sequence to be analyzed by BlastX and BlastP and TblastN analyses with the deduced translated Orf. Additional protein sequence analyses included the generation of hydropathy profile and transmembrane topology prediction using different softwares : PHD (Rost B, 1996), accessible through the PredictProtein interface [<http://cubic.bioc.columbia.edu/predictprotein/>], and Memsat (Jones et al. 1994) via the protein structure prediction server Psipred v2.4

[<http://bioinf.cs.ucl.ac.uk/psipred/>] and others available through ExPASy server (Appel et al. 1994) [<http://ca.expasy.org/>].

Functional expression of *mntH* genes.

Genomic DNA from *Chlorobium tepidum* (ATCC #49652D), and DNA extracted (De Lamballerie et al. 1992) from food grade isolate of *Lactococcus lactis* subsp. *lactis* (ATCC #11454) and *E. faecalis* V583 strains was used to PCR amplify *mntH* genes with *Pfu* DNA polymerase (Stratagene, La Jolla, CA, USA). DNA fragments were purified using appropriate Qiagen kit and digested using restriction endonucleases from NEB. *mntH* genes were cloned for expression in *E. coli* after double digestion with restriction endonucleases and ligation with T₄ DNA ligase (Gibco, Burlington, ON) to similarly restricted vector, that had been dephosphorylated with Calf Intestine phosphatase (NEB, Missisauga, ON). *Nco* I and *Kpn* I restriction sites were used to clone in pBS5'-*PmntH* vector (Bergevin and Cellier, unpub.) and *Nco* I and *Xba* I sites to clone in pBAD24 (Guzman et al. 1995). These manipulations were performed following the manufacturers' protocol instructions. The DNA sequence of *mntH* genes was determined using two clones for each gene using both strands and sequencing systems (CEQ 2000 XL DNA Analysis System, Beckman Coulter, Fullerton, CA, USA and ABI3700 DNA Analyzer, Applied Biosystems, Foster City, CA, USA). DNA sequences obtained for genes from *L. lactis* and *E. faecalis* were submitted to GenBank (*L. lactis mntH B* (AF544405) and *mntH C β* (AF544404); *E. faecalis mntH C β 1* (AF545506) and *mntH C β 2* (AF545507)). Disk-assay of metal sensitivity was performed in DH11S *mntH* and SL93 *mntH* strains, as previously described (Makui et al. 2000).

An internal *Nco* I restriction site was removed from *C. tepidum mntH* B gene by silent mutation using bubble PCR method (Ho et al. 1989). Two DNA fragments (1216 bp and 290 bp) were amplified with 4mM MgSO₄, the oligonucleotide primer pairs (mutated nucleotide underlined) CtepF (5'- TCG ACC ATG GCA GTT TCA TTT TCG -3') and CtepIR (5'- CGA CCA CCA TTG TAA TCA GCA CC -3') ; Ctep IF (5'- GGT GCT GAT TAC AAT GGT GGT CG -3') and CtepR (5'- GGT ACC CTG TTG TTG CAT CGC G -3'), and cycling parameters : 95°C 45s, (95°C 45s, 52°C 45s, 72°C 2m30s) four times, (95°C 45s, 60°C 45s, 72°C 2m30s) 22 times. The DNA fragments were gel purified, mixed and PCR amplified using the oligonucleotide primer pair CtepF and CtepR and the cycling parameters above and purified.

Lactococcus lactis subsp. *lactis* group B *mntH* gene was amplified with 4mM MgSO₄ and the oligonucleotide primer pair LlactisBF (5'- TGA CCA TGG AAA GGA TAA AAA TGA GG -3') and LlactisBR (5'- CAA TAC TAT GGT ACC AAA TTT TTC GG -3') and cycling parameters: 95°C 45s (95°C 45s, 50°C 45s, 72°C 2m30s) four times, (95°C 45s, 60°C 45s, 72°C 2m30s) 22 times. *L. lactis* group C *mntH* gene was amplified using 2mM MgS O₄ and the oligonucleotide primer pair LlactCF (5'- AAC CAT GGA TGA TAA AGA ACA TCA AC -3') and LlactCR (5'- AAA TAT TTA TCT GGT ACC CCA AAA AG -3'), and cycling parameters: 95°C 45s (95°C 45s, 48°C 45s, 72°C 2m30s) four times, (95°C 45s, 56°C 45s, 72°C 2m30s) 22 times. The resulting 1300 bp (B) and 1700 bp (C) fragments were purified. Truncation mutants were obtained by PCR mutagenesis with 4 mM MgS O₄ and the oligonucleotide primer pair LlactisBF (above) and LBΔ3' (5' ATG TCT AGA TTA TGC ATC ATA A) and cycling parameters: 95°C 45s (95°C 45s, 35°C 45s, 72°C 2m30s) four times, (95°C 45s, 49°C 45s, 72°C 2m30s) 22 times for LlacB Δ-COOH, and for . LlacC Δ-COOH

LCΔ3' with 6mM MgSO₄, the oligonucleotide primer pair LlactisCF (above) and LCΔ3' (5' CT TCT AGA ACA TTA TGG CAT TCT C) and cycling parameters: 95°C 45s (95°C 45s, 45°C 45s, 72°C 2m30s) four times, (95°C 45s, 53°C 45s, 72°C 2m30s) 22 times. The resulting DNA fragments had a stop codon followed by an *Xba* I and were purified for subcloning in pBAD24.

Enterococcus faecalis group Cβ1 and Cβ2 *mntH* genes were amplified with 10mM MgSO₄ and the oligonucleotide primer pair Efeac1542 A ATT GCC ATG GGG AAA AAT TCA GAA GAA CC and EfeacR AAT TTT TCT AGA TTT CAT TAT TTA ACC and cycling parameters: (95°C 45s, 52°C 45s, 72°C 3m) four times, (95°C 45s, 60°C 45s, 72°C 3m) 22 times for full-length EfaCβ1, and with 8mM MgSO₄ and the oligonucleotide primer pair EfeacF AAA CCA TGG GTT TTT GGA AAA CGT and EfeacR (above) and cycling parameters: (95°C 45s, 45°C 45s, 72°C 3m) four times, (95°C 45s, 58°C 45s, 72°C 3m) 22 times for Efa Cβ1^{*}Δ-NH₂. *E. faecalis mntH* Cβ2 gene was amplified with 8mM MgSO₄ and the oligonucleotide primer pair Efeac25 5' ATGT TCC ATG GGC AAA GAA AAG AAA ACG GAA and Efeac23 5' AG CCT TTG TCT AGA TTA ATT CGT TAA ACC A and cycling parameters: (95°C 45s, 45°C 45s, 72°C 3m) four times, (95°C 45s, 58°C 45s, 72°C 3m) 22 times. The resulting DNA fragments were purified.

RESULTS

Subdivision of *mntH* group C in subgroups α , β , γ

Prokaryotic Nramp/MntH homologs were identified only in eubacterial domain *Bacteria*. One sequence from *Archaea*, showing the highest level of similarity to MntH/Nramp homologs was used as an outgroup to root the tree presented (Fig. 1). Up to 94 MntH/Nramp sequences were used to produce multiple sequence alignments which were used for phylogenetic and molecular evolutionary analyses (Kumar et al. 2001). Neighbour-joining and minimal evolution phylogeny calculations converged in both the interior branch and bootstrap statistical tests (Kumar et al. 2001) to confirm the existence of three groups of MntH A, B and C, and further define three subgroups C, identified $C\alpha$, $C\beta$, $C\gamma$ (Fig. 1).

Bootstrap tests of maximum parsimony calculations were also in agreement to differentiate the MntH groups A, B, C and subgroups $C\alpha$, $C\beta$, $C\gamma$. There were local differences in the tree topology for eukaryotic and MntH C sequences except the branching of subgroup MntH $C\beta$ which remained unchanged. Parsimony analyses yielded a topology with a single node from which diverged all eukaryotic Nramp and MntH group C sequences. Statistical support for this alternate local topology was lower than for the topology presented. This circle representation (Fig. 1) displays the successive divergence of bacterial i) MntH B and ii) MntH A, then iii) eukaryotic Nramp separated in iv) 'prototype' Nramp and MntH $C\alpha$, β , γ and yeast Nramp, and v) metazoa and plant Nramp.

Metazoa (e.g. Vertebrata, Insecta) and plant (e.g. Monocot, Eudicot) *Nramp* genes include homologs from i) two types of plant sequences (type I: AtNR1, 6, OsNR1, 3 and type II: AtNR2-5, OsNR2), ii) one type of invertebrate (e.g. Mycetozoa, Insecta) sequences (type II : DdNR2; AgNR2, DmeNR, CeNR2-4), and iii) vertebrate (fish, bird, mammals) sequences (OmNR1, 2, GgNR1, HsNR1,2).

The 'prototype' *Nramp* include sequences from the protist unicellular alga *Chlamydomonas reinhardtii* (CreNR, Chlorophyta), and another type of invertebrate sequences (type I : DdNR1 and AgNR1) identified in the amoeba *Dictyostelium discoideum* and mosquito *Anopheles gambiae* (Fig. 1). The statistical support for node iv) (above) increases to 71 vs. 50 (Fig. 1) if yeast *Nramp* diverge before this node as a separate eukaryotic cluster (statistical support 63, minimal evolution, and interior branch test, (Kumar et al. 2001).

MntH C α proteins were found in species that belong to distant phyla based on phylogenetic analyses using 16S RNA sequence and full genome content (Korbel et al. 2002): the *Cyanobacteria* (*Nostoc* spp.), Gram positive (*Listeria* spp.) and several Gram negative bacteria (α - and γ -proteobacteriaceae; Fig. 1). The sequences from *Nostoc*, *Listeria* spp., *Agrobacterium tumefaciens* showed remarkably high BLAST scores with the 'prototype' *Nramp* sequences found in protists and invertebrates (CreNR, DdNR1, AgNR1; Fig. 1). Other *mntH C α* genes found in γ - and α -proteobacteriaceae (Xanthomonales et Rhizobacteria, respectively) encoded closely related proteins (XaxoC α , XcamC α , XfasC α , and MlotC α , MmagC α Fig. 1) and shared in addition a common genomic organization with an *mntR* gene (Mn-dependent

regulator;) in a putative operon divergently transcribed (I. Bergevin M. Cellier, unpub.).

MntH subgroup C β contains sequences derived from low GC Gram positive bacteria (Lactic acid bacteria, including oral streptococcaceae, and staphylococcaceae) which are somewhat related Gram positive species. This group contains also a sequence derived from a distant bacterial species, the Gram negative bacteria *Wigglesworthia glossinidia*. Another particularity of the sequences from subgroup MntH C β is their consistently lower levels of similarity to known eukaryotic Nramp sequences than those displayed by MntH C α and C γ sequences. However some sequences from this subgroup (e.g. LlacC β) show an hydropathy profile that is more resembling those of eukaryotic sequences (e.g. DdNR1) than MntH A or B profiles (shown in Fig. 4B).

MntH C γ contains sequences from β - and γ -proteobacteriaceae, including several *Burkholderia* spp. and Pseudomonads, respectively. *Burkholderia* spp. possess multiple *mntH* genes. *B. cepacia* genome contains two *mntH* A genes (BceA1 and A2) and one *mntH* C gene (BceC γ 2), *B. fungorum* possesses two *mntH* C γ genes (BfuC γ 1 and C γ 2) and *B. pseudomallei* has one gene of each group (BpsA2 and BpsC γ 1) whereas *B. mallei* has one group A gene, one group C pseudogene, and one group C gene (BmaA2, BmaC γ 1 and C γ 2, respectively). Also, *P. aeruginosa* possesses two *mntH* C genes from distinct subgroups (PaeC α 1 and PaeC γ 2; Fig. 1). MntH C γ 2 sequences are highly related to PaeC α 1 (in the range of 70 %-52 % identity with the β - proteobacteria *Ralstonia* homolog, Rso1C γ 2, and with the second *P. aeruginosa* predicted protein Pae C γ 2, respectively). MntH C γ 1 sequences are

distinct and seem related to the *Burkholderia* proteins BceA2 and BmaA2 (Blast Scores >284, mean 294).

Identification of 'prototype' eukaryotic Nramp

We report the existence of at least a pair of genes in distant invertebrates species, the slime mold *D. discoideum* and the Diptera *A. gambiae*. One sequence from each of these organisms (referred to as Invertebrate type I Nramp or 'prototype' Nramp) shows high level sequence identity with its invertebrate ortholog (DdNR1, AgNR1) and with bacterial sequences, from MntH C α in particular. Invertebrate type II sequence present in *D. discoideum* and *A. gambiae* shows less conservation with its invertebrate ortholog (DdNR2, AgNR2) and is either more closely related to plant II sequences (DdNR2), or to animal invertebrates sequences (AgNR2).

The 'prototype' *DdNR1* orf is intronless (GenBank accession Number: AC117080, 1644 bp Orf), and *AgNR1* albeit lacking the 3' end in its current version (GenBank accession Number: CRA_x9P1GAV53BU) does not seem to contain an intron in 5' . An intron was however identified in the *D. discoideum* entry corresponding to *DdNR2* gene (The Preliminary Directory Of Dictyostelium Genes - Version 2.1 <http://dicty.sdsc.edu/annotationdicty.html>; Contig10319, unspliced DNA, 1816 bp). The position of this intron corresponds in the protein to the sequence immediately preceding the transmembrane domain 1 of Nramp proteins. This intron position is conserved in *A. gambiae* gene *AgNR2*, and in genes of *Drosophila* (Berkeley *Drosophila* Genome Project port <http://www.fruitfly.org/annot/index.html>, and the Gadfly database accession: FBgn0011672, unspliced DNA, 9239 bp), *C. elegans* (Wormbase, <http://www.wormbase.org/> clone Y69A2AR.4, unspliced DNA, 6506 bp

and clone K11G12.4, unspliced DNA, 37936 bp respectively), *A. thaliana Nramp1* and *Nramp6* genes (GenBank accession Number: AY074865, AY074865 for the respective mRNA, and NC_003070 for chromosome 1 sequence; unspliced DNA, 2578 and 2865 bp, respectively) and *H. sapiens NRAMP1* gene (Cellier et al. 1994). Current data indicate that *AgNR2* gene is composed of at least eight exons, generally conserved in position with those identified in human *NRAMP1* gene and in *Drosophila Nramp* gene.

The 'prototype' Nramp proteins DdNR1 and AgNR1 are more closely related to bacterial MntH A and B groups than any other eukaryotic sequences, and would thus appear logical descendants from a gene of a bacterial endosymbiont close to the ancestor of mitochondria. Interestingly, the protist *C. reinhardtii*'s Nramp sequence (CreNR, GenBank accession number AF515631) is about 50 % identical with the MntH C α sequences of *Nostoc* and *Listeria* spp., and was consistently found to group with MntH C α cluster in phylogeny analyses (Fig. 1). CreNR is less related than DdNR1 and AgNR1 to the other groups of bacterial sequences MntH A and B, and is slightly more distant from both DdNR1 and AgNR1 than these are from each other. The clustering of CreNR and MntH C α protein sequences strongly suggest they derive from a common ancestor.

C. reinhardtii Nramp sequence is predicted to encode 11 transmembrane domains (TMD; Fig. 4A), a topology similar to those predicted with sequences from other unicellular organisms (fungi and bacteria). The other 'prototype' Nramp sequence DdNR1 displayed however, like other metazoa and plant sequences a profile with 12 hydrophobic segments (Fig. 4B). MntH C α sequences displayed hydrophobic profiles

more similar to both 'prototype' Nramp sequences CreNR and DdNR1 than to MntH A or B hydrophathy profiles (Fig. 4B).

Divalent metal transport by MntH B and MntH C β proteins in *E. coli*

The level of sequence similarity between MntH groups suggested MntH B and C proteins should function as divalent metal ion transporters. To test this possibility several Orfs from group B (*C. tepidum*, *C. acetobutylicum* and *L. lactis*) and group C β (*E. faecalis* and *L. lactis*) were cloned and expressed in *E. coli*. Protein expression was induced with arabinose, at a concentration depending on the toxicity due to expression of each MntH protein. Bacterial sensitivity to certain divalent metals was measured by disk-diffusion assay. The results obtained with *E. coli* MntH A protein showing increased sensitivity to Mn and Fe in particular (Table 2) were previously correlated to increased cellular uptake of radio-labelled Mn(II) and Fe(II) due to *E. coli* MntH A expression (Makui et al. 2000). This activity was abrogated by insertion of a mature β -lactamase polypeptide in the loop connecting the predicted transmembrane domains 5 and 6 (EcolA 179Blam, Table 2, Courville and Cellier, unpub.).

Bacteria transformed with the plasmid alone show that Cd and Co are the most toxic metals. Increased sensitivity to cobalt, cadmium, iron and manganese was observed after inducing the expression of *E. coli* MntH A protein with 0.06% arabinose (EcolA, Table 2). Expression of *C. tepidum* MntH B was induced with 0.06% arabinose and conferred increased sensitivity to Co and Fe (Table 2). Similar results were obtained with *C. acetobutylicum* MntH B which was also more toxic in *E. coli* (data not shown).

E. faecalis MntH C β 1 and C β 2 (EfaC β 1 and EfaC β 2) proteins were toxic in *E. coli*. The expression of EfaC β 1 increased sensitivity to Cd, Co and Fe. An N-terminus truncated form of this protein was as toxic as the full length protein but conferred only little increase in Co sensitivity. Similar residual sensitivity to Co was observed with EcolA 179 Blam mutant. This suggested that metal sensitivity conferred by expression of the full-length MntH C β 1 protein was not a mere consequence of decreased cell viability due to protein toxicity. The expression of EfaC β 2 conferred increased sensitivity to Co and Cd well above the levels observed with *E. faecalis* MntH C β 1 Δ -NH₂ and EcolA 179 Blam.

Attempts at functional expression of *L. lactis* MntH C β 1 protein in *E. coli* were not successful. This protein was not toxic and conferred increased sensitivity to Cd upon expression. However, a C-terminal truncation mutant lacking the predicted transmembrane domains 9-12 (see Fig. 4) conferred similar Cd sensitivity. Similar results were obtained with *L. lactis* *mntH* B pseudogene (Table 2, and below), indicating that *L. lactis* MntH proteins were non functional in *E. coli*.

The data show that MntH B and C β proteins from different species facilitate intracellular accumulation of divalent metal ions in *E. coli*, suggesting functional homology between MntH A, B and MntH C proteins. Differences in sequence motifs and deduced hydropathy profile between these groups should therefore reflect distinct evolutionary pathways. Sequence relatedness of MntH C with prototype Nramp suggest the bacterial proteins were acquired by horizontal transfer.

Horizontal transfer of an *mntH* C β gene in *Wigglesworthia*

Recent horizontal transfer of genes can be indicated by their presence within a DNA fragment with defined properties (e.g. presence of a sequence strongly conserved across species such as a *tRNA* sequence, and sequence relationships in disagreement with species phylogeny). An *mntH* C β gene was found in the genome of a species from a distinct *Enterobacteria* lineage, *Wigglesworthia glossinidia*, that is an obligatory endosymbiont of the tsetse fly (Akman et al. 2002; Akman et al. 2001; Aksoy 1995; Chen et al. 1999). The genome of *Wigglesworthia* displays hallmarks of endosymbiont genome: low GC content and size (<0.8 Mb; Akman et al. 2002). *Wigglesworthia mntH* gene is linked to an adjacent *asn-tRNA* gene in 5' (Fig. 2; Akman et al. 2002). Segments of the nucleotide sequence of *Wigglesworthia mntH* Orf show conservation with homologs from the Gram positive *E. faecalis* and *Staphylococcus* spp. (e.g. sequence encoding the first transmembrane domain of the protein). *Wigglesworthia* MntH protein is more closely related to MntH C β proteins from Gram positive bacteria than to MntH proteins from any Gram negative species proteins (from either MntH A, MntH C α , or C γ groups, Fig. 1, 2).

Proteins encoded by the genes surrounding *Wigglesworthia asn-tRNA-mntH* locus appeared similar to proteins of closely related *Enterobacteriaceae* species. Portions of *Wigglesworthia* DNA segment were similar to *E. coli* chromosome gene content and order (Fig. 2). The *asn-tRNA* linked to *mntH* C β gene, and the *glt-tRNA 16S RNA* locus were most conserved with *Buchnera* species (Silva et al. 2001). These observations suggest distinct origins for *mntH* C β (Gram positive) and surrounding genes (*Enterobacteriaceae*). Simultaneous comparison of *Wigglesworthia* genome

proteins with those of *E. coli* and *S. aureus* or *S. mutans* reveals clearly WbreC β as the protein most related to Gram positive species vs. *E. coli*. Thus the ~ 1.7 kb DNA fragment carrying the *asn-tRNA* linked to *mntH* gene is likely the result of the recent integration of a mobile DNA element.

Linkage of *mntH* C β genes with different IS-like genes in oral streptococcaceae

S. mutans genome contains a single *mntH* C β gene which is adjacent to a presumed insertion sequence, member of IS861 family (Fig. 3). A different arrangement was found in *S. mitis* with a member of IS861 family in opposite orientation, adjacent in 3' to *mntH* C β 1 gene (Fig. 3). This *S. mitis* IS861-like sequence is most conserved between Mitis and Mutans groups and also in *S. equii*.

A second *S. mitis* gene (*mntH* C β 2) was found linked to another putative insertion sequence (IS1381 family), that is located in 5' upstream of *mntH* C β 2 (Fig. 3). It is not known whether this IS1381-like element is the sole exemplary of this family found in the genome of *S. mitis* (biovar 2), as previously reported in strain NCTC 12261 (Sanchez-Beato et al. 1997). Sequence alignment of the IS1381-like elements from these two strains revealed 20 nucleotide substitutions (3 % divergence), located in the portion of the IS distal of *mntH*, corresponding to OrfA of the putative transposase. The significance of this variation is not yet known. The nucleotide sequence of *S. mitis* *mntH* C β 2 gene (contig 390) was more conserved with *S. gordonii* *mntH* C β gene (88%) than with *S. mitis* *mntH* C β 1 gene (80%). This closer

relatedness between *S. mitis* *mntH* C β 2 and *S. gordonii* homologs was also observed when comparing the deduced protein sequences (Fig. 1).

No IS related to *IS1381* was identified in the genome of *S. gordonii* including in the vicinity of *mntH* C β gene. *S. gordonii* gene is placed within a gene arrangement different from *S. mitis* *mntH* C β 2 gene, juxtaposed to another type of insertion sequence, *IS1562*-like (Fig. 3). Copies of this IS type were identified in regions encoding virulence factors of *S. pyogenes* (Berge et al. 1998) and other GAS (Smoot et al. 2002). Further studies will be required to investigate the role of *mntH* C β gene in the virulence of *S. gordonii*.

Co-existence of prematurely terminated *mntH* orf and a novel *mntH* C gene

Lactococcus lactis genome (subsp. *lactis*) contains two hypothetical *mntH* genes: the gene *ytjB*, similar to *mntH* B, was found prematurely terminated at position 654, (GenBank Accession number: AE006421), and the gene *ykjB* similar to *mntH* C β seemed truncated (GenBank Accession number: AE006340 and AE006341). Both genes were PCR amplified from a different isolate of *L. lactis* and cloned for expression in *E. coli* and sequenced. The presence of a stop codon in *ytjB* was confirmed (but at position 981), and expression in *E. coli* of the natural pseudogene or a truncation mutant (1-981; indicated in Fig. 4) both resulted in sensitivity to Cd (Table 2). *ykjB* gene encodes a predicted full-length MntH C β protein of 510 amino acid residues, that displays an hydropathy profile with 12 predicted TMD. This characteristics is shared with two closely related MntH group C β sequences, from *E.*

faecalis and *E. faecium* (Figure 1), and with all the eukaryotic Nramp homologs from multicellular organisms (e.g. *D. discoideum*, Fig. 4). However, no clear evidence of transport by this MntH group C β protein could be obtained after expression in *E. coli* (Table 2).

The co-existence of a natural mutation interrupting *Lactococcus mntH B* gene and of novel full-length *mntH C β* gene in the same genome suggested it could represent an intermediate step in a process that would lead ultimately to the elimination of the *mntH* group B gene. Indeed, pseudogenes are rarely observed in bacterial genomes (Lawrence et al. 2001). Examination of *mntH* gene sequences in other multigenic species revealed another example of co-existence of a pseudogene and a novel *mntH C γ* gene in the same genome. *B. pseudomallei* and *B. mallei* are very closely related bacterial pathogens (Heine et al. 2001; Ulett et al. 2001; Utaisincharoen et al. 2001). *B. pseudomallei* genome contains two genes (*mntH A*, *C γ 1*), whereas *B. mallei* possesses three sequences (*mntH A*, *C γ 1*, *C γ 2*; Table 1), *C γ 1* being a pseudo gene interrupted by a termination codon at position 93. The findings of *mntH* genes prematurely terminated in genomes containing an additional full-length *mntH C* gene suggest a plausible causal link that will lead to the pseudogene replacement by a novel *mntH C* gene.

DISCUSSION

Several novel findings presented in this work confirm the polyphyletic origins of bacterial Nramp (MntH groups A, B, and C). The sequence from the low GC Gram positive *Carboxydotherrmus hydrogenoformans* appeared as an intermediate between MntH groups B and A, which is not inconsistent with the proposition that MntH B would be more ancestral than MntH A, and of progressive evolution from one type to another (Cellier et al. 2001). The biases previously reported in species living in anaerobiosis (MntH group B), or that are capable of infection (symbiotic or pathogenic, MntH group C) are still observed, as well as the absence of MntH in Archaea out of 17 genomes sequenced. In addition, preliminary evidence of divalent metal transport by MntH homologs from either group B (*C. tepidum*) or C β (*E. faecalis* 1 and 2) expressed in *E. coli* are presented. The results suggest conservation of MntH transport function in divalent metal uptake. Differences in sequence motifs and deduced hydropathy profile between MntH groups could reflect their distinct origins and/or correspond to differences in their structure/function relationship. Further detailed studies of MntH proteins appear thus warranted.

Several additional results support of the hypothesis of horizontal transfer *Nramp* gene from eukaryote to prokaryote at the origin of bacterial MntH group C.

First, it appears that the phylogeny based on MntH C sequence is not very congruent with the phylogenies based of the 16S RNA and whole genome content (Korbel et al. 2002). This indicates horizontal transfer of *mntH* group C genes subsequent to the divergence of these bacterial species.

In subgroup MntH C β , the *t-RNA-mntH* C β gene block of apparent foreign origin found on the chromosome of *Wigglesworthia*, and the oral *Streptococcus S. mitis* IS1381-*mntH* C β gene block constitute two convincing examples of likely horizontal gene transfer. The putative IS transposase is conserved with *S. pneumoniae* IS1381 element, which was previously shown to modulate the expression of gene adjacent to its insertion point in *S. gordonii* genome (Demuth et al. 1997). IS and *t-RNA* genes are important elements in the development of new bacterial variants through recombination events (Hou 1999; Mahillon et al. 1999). The data suggest a direct role for these elements in horizontal transfer of *mntH* C β genes.

In MntH C α , the high level of identity between sequences derived from distant species, and the precise conservation of an *mntR-mntH* C α putative operon in some α -proteobacteriaceae (but not in *R. palustris*, *A. tumefaciens* or *B. suis*; Fig.1), and in some γ -proteobacteriaceae (but not in *P. aeruginosa*; Fig. 1), also suggests horizontal DNA transfer of this composite element subsequent to the divergence of these bacterial species. The net advantage could be to import both the gene and its regulator of expression, enabling the use of *mntH* C α in a regulated fashion (Doolittle RF 2002.).

Concerning MntH C γ genes, the divergent position of the cluster C γ seems to result from specific sequence relationships between some *Burkholderia* spp. group A sequences (BmaA2 and BceA2) and the MntH C γ 1 sequences (BmaC γ 1 and BfuC γ 1) on one hand, and on the other hand between *P. aeruginosa* MntH C α 1 and MntH C γ 2

sequences. *Burkholderia* sequence relationships MntH A2/C γ 1 seem consistent with the horizontal acquisition of an *mntH* C gene by an organism already possessing a group A gene, and further conservation of both groups A and C genes. The close relatedness of *P. aeruginosa* MntH C α 1 and the sequences BfuC γ 2 and RsoC γ 2 also suggests horizontal transfer of a sequence similar to PaeC α 1.

In addition, co-existence in the genomes of Gram positive and Gram negative bacteria of an *mntH* pseudogene together with a novel full-length *mntH* group C β and C γ gene, respectively, suggest relatively recent horizontal acquisition of group C *mntH* genes. Further work is required to determine if *L. lactis* MntH C β protein is active as are the MntH C C β proteins from *E. faecalis*.

Second, the current topology of MntH C subtree suggests independent gene transfers at the origin of subgroups C α , β and γ . The presence of facultative or obligate intracellular organisms in each subgroup C indicate close association with eukaryotic cells is a property common to species of each subgroup C (e.g. C α , *N. punctiforme*; C β , *W. glossinidia*; C γ , *Burkholderia* spp.).

The *Cyanobacteria Nostoc* spp. are multicellular bacteria capable of differentiation, that perform independent and/or incompatible functions such as oxygenic photosynthesis and symbiotic N₂ fixation (Meeks et al. 2002). *W. glossinidia* is an obligate intracellular bacterium having established mutualistic interaction with cells of the tsetse fly to create the bacteriome tissue (Chen et al. 1999; Dale et al. 2001).

Burkholderia spp. strains STM 678 & 815 are closely related to *B. cepacia* and were identified as novel β -proteobacteria rhizobia (present in the symbiotic nodules of leguminous plants (Minerdi et al. 2001; Moulin et al. 2001)). The presence of distinct MntH C proteins in these unrelated bacteria associated with various host cells suggest many opportunities for direct transfer of an eukaryotic *Nramp* gene toward bacteria.

Third, the discovery of ‘prototype’ *Nramp* in the social amoeba *D. discoideum* and the mosquito *A. gambiae* indicate potential eukaryotic sources of *Nramp* genes that might have been easily acquired and expressed by opportunistic bacteria because they are intronless.

The ‘prototype’ *D. discoideum DdNR1* gene is intronless (Glockner et al. 2002) and the intervening sequence found in *DdNR2* exhibits typical properties of *Dictyostelium* introns (small size, frequency of 1.2 per gene, >95.5 % AT-rich, intronic only splicing signals: 5’ GTA, 3’ AAG; (Glockner et al. 2002), and data not shown). Although one cannot yet conclude that *A. gambiae Nramp1* orf (*AgNR1*) is intronless, it is clear that *AgNR2* is interrupted by at least 8 introns. In addition, the precise conservation of the position of the splicing sites joining exons 2 and 3 of human *NRAMP1* gene (Cellier et al. 1994) in both *Dictyostelium* and *Anopheles* genes would support the view that mammalian *Nramp* derive from the common ancestor of *DdNR2* and *AgNR2*.

It appears possible that an early ancestral eukaryotic *Nramp* gene was duplicated before the divergence of Mycetozoa, and that ‘prototype’ *Nramp* genes have been since maintained in the genomes of some simple organisms. Opportunistic intracellular bacteria could have exploited close association with such organisms to

capture an intronless copy of *Nramp*. *D. discoideum* is a professional phagocyte that feeds on bacteria and may use Nramp proteins at the level of the phagosome as part of a bacterio-static or -toxic functions, which include also an NADPH oxidase (GenBank accession number AF123275). Capture of an intronless gene encoding a 'prototype' Nramp could thus be advantageous for a microbe that may have to compete with host Nramp after phagocytosis.

Fourth, the remarkable sequence conservation between the 'prototype' Nramp identified in the unicellular alga *C. reinhardtii* and MntH C α proteins supports the feasibility of direct gene exchange between eukaryotes and bacteria.

This relationship contrasts with the divergence of *Chlamydomonas* sequence from others 'prototype' Nramp and MntH A and B protein sequences. This divergence is displayed in a distinct hydropathy profile (Fig. 4B). In a multiple sequence alignment including the three known 'prototype' Nramp and the closest bacterial (*N. punctiforme*) sequence, 29 % of positions were found invariant; 9 % were identical between *D. discoideum* and *A. gambiae* and another species (4 and 5 %) and 3% were unique to these species, whereas 15 % were identical between *Chlamydomonas* and *N. punctiforme* and another species (7 and 8 %) and 5% were unique to these species. Thus the level of ~ 50 % sequence identity between *Chlamydomonas* and *N. punctiforme* sequences is predominantly due to residues conserved with other 'prototype' Nramp sequences, highlighting the close relatedness of MntH C α and 'prototype' Nramp sequences in general.

Fifth, the relatively short distances measured in MntH C α clade compared to MntH A, support the suggestion of a relatively recent origin for MntH C α .

Distance calculations were based on full-length sequences omitting pairwise deletions and using a Poisson correction. Branch length obtained in the tree topology after interior branch testing were measured for each clade, between the last common group-specific node and the tip of the branch of constituting sequences, giving values in the range: 0.28-0.32 a.a. replacement/site (NostC α - BsuiC α), 0.28-0.39 a.a. repl./site (SaurC β - EfaC β 1), 0.36-0.40 a.a. repl./site (PaeC γ 2 - BpsC γ 1) ; 0.39-0.52 a.a. repl./site (BcerA - MlepA, and ChydA, 0.73 a.a. repl./site) and 0.41-0.53 a.a. repl./site (CtepB - LlacB). Similar results showing significantly shorter distances for MntH C groups were obtained using only 272 informative sites of the multiple alignment (see supplementary Figure 5), indicating that the distance differences observed between clades are not due to the more variable parts of the sequences. Phylogenetic inference based on a maximum likelihood approach using the 272 sites data set confirmed these topology and distance estimates (data not shown). Smaller distances observed within MntH C α compared to MntH A, which both contains sequences from distant species thus suggest a more recent origin for the group MntH C α .

Lastly, analyses of the a.a. substitution rate within MntH groups, either among sites (following a gamma distribution, estimated with the shape parameter alpha; Gu and Zhang, 1997) or between groups of homologs (independence of evolutionary rate, estimated with the coefficient of divergence theta, and likelihood ratio test, LRT; Gaucher *et al.*, 2002) underscore the differences between older groups, MntH B and A on the one hand and more recent MntH C groups on the other hand.

A small value of alpha parameter obtained for MntH C α (0.64) indicates heterogeneity in the substitution rate among sites in this group (e.g. juxtaposition of conserved and not conserved sites). Higher values corresponding to a more homogeneous substitution pattern were obtained for MntH A (alpha = 1.1, and 1.22 when subgroup MntH A' (BceA1-2 and BpsA2) is omitted ; see Supplementary Figure 6). This is consistent with accumulation of sequence differences due to speciation of Gram positive and negative orthologs, reducing overall rate variation among sites in MntH A group (Gaucher *et al.*, 2002). Preservation of the pattern of a.a. substitution in MntH C α group may thus be explained by its more recent origin. Similarly, the alpha parameter values obtained for MntH groups B (1.31) and C β (0.61) seem consistent with more ancient and recent origins, respectively.

MntH C γ showed a high alpha value (1.1) that dropped when subgroup MntH C γ 1 (BfuC γ 1, BpsC γ 1) was omitted (0.65). This indicates different a.a. distributions in subgroups C γ 1 and C γ 2, which are more conserved in sequence with groups MntH A' (BpsA2, BceA2) and MntH C α , respectively.

Estimation of the coefficient of functional divergence theta by the subtree likelihood method can reveal differences in evolutionary rate between homologous genes (Gaucher *et al.*, 2002) and may therefore provide indications concerning *mntH* gene duplication and functional divergence. The coefficient of divergence theta were measured between MntH groups and a LRT calculated to test the null hypothesis (similar evolutionary rate between the groups compared).

Minimal LRT values were obtained comparing MntH A' and MntH A or MntH B suggesting rather similar evolutionary rates for these groups. The higher LRT values were obtained comparing 'archetype' Nramp and MntH C α or C β , suggesting potential divergence between these groups. MntH A and C α exhibited great amplitude in potential divergence from other groups, with higher LRT values obtained for MntH A compared to MntH C α , archetype Nramp, MntH C β , or yeast Nramp, and between MntH C α and 'archetype' Nramp, MntH A or B, or yeast Nramp (supplementary Figure 6).

MntH C γ and C γ 2 subgroups showed low LRT values when compared to MntH A and C α , suggesting less divergence for *mntH* C γ genes. Also, MntH A' subgroup (including BceA1-2 and BpsA2) showed reduced LRT values with the groups that were most divergent from MntH A. These observations of limited sequence divergence together with distinct a.a. distribution patterns suggest a different pathway of evolution for subgroups MntH A' and MntH C γ 2. These results may be explained by horizontal transfer of prototype *Nramp* or *mntH* C-like genes into species already possessing an *mntH* A gene, which resulted in the current subgroups MntH A' and MntH C γ .

We conclude that the remarkable sequence conservation between bacterial MntH C proteins and eukaryotic Nramp proteins does not mean MntH C proteins are the bacterial ancestor of eukaryotic Nramp but rather indicate horizontal gene transfer from eukaryotes to bacteria. This interpretation is consistent with conclusions of other

studies (Canback et al. 2002; Koonin et al. 2001). Opportunistic bacteria living in close contact with primitive eukaryotic cells may have captured an intronless *Nramp* prototype gene that would improve bacterial metal acquisition and survival at host contact.

ACKNOWLEDGMENTS

We would like to thank two reviewers for their constructive comments, Dr. S. Aksoy (Yale University, USA) for sharing *Wigglesworthia* genomic data and generously providing purified *Wigglesworthia* genomic DNA, and Dr. P. Courvalin (Institut Pasteur, Paris, France) for kindly providing *E. faecalis* V583 strain. We acknowledge gratefully the NCBI online facilities (National Library of Medicine, Washington, DC), including BLAST search and Taxonomy servers and also in particular, the contributing Microbial Genome Sequencing Projects. This work is supported by a research grant from the Canadian Institutes of Health Research. I. B. and E. R. are supported by the NSERC and CIHR, respectively. Dr. Cellier is a scholar from the Fonds pour la Recherche en Santé du Québec.

REFERENCES

Agranoff D, Monahan IM, Mangan JA, Butcher PD, Krishna S (1999) *Mycobacterium tuberculosis* expresses a novel pH-dependent divalent cation transporter belonging to the Nramp family. *J Exp Med* 190:717-724

Akman L, Aksoy S (2001) A novel application of gene arrays: *Escherichia coli* array provides insight into the biology of the obligate endosymbiont of tsetse flies. *Proc Natl Acad Sci U S A* 98:7546-7551

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossidina*. *Nat Genet* [epub ahead of print]

Aksoy S (1995) *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *Int J Syst Bacteriol* 45:848-851

Appel RD, Bairoch A, Hochstrasser DF (1994) A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem Sci* 19:258-260

Berge A, Rasmussen M, Bjorck L (1998) Identification of an insertion sequence located in a region encoding virulence factors of *Streptococcus pyogenes*. *Infect Immun* 66:3449-3453

Canback B, Andersson SG, Kurland CG (2002) The global phylogeny of glycolytic enzymes. *Proc Natl Acad Sci U S A* 99:6097-6102

Cellier M, Govoni G, Vidal S, Kwan T, Groulx N, Liu J, Sanchez F, Skamene E, Schurr E, Gros P (1994) Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. *J Exp Med* 180:1741-1752

Cellier M, Prive G, Belouchi A, Kwan T, Rodrigues V, Chia W, Gros P (1995) Nramp defines a family of membrane proteins. *Proc Natl Acad Sci* 92:10089-10093

Cellier MF, Bergevin I, Boyer E, Richer E (2001) Polyphyletic origins of bacterial Nramp transporters. *Trends Genet* 17:365-370

Chandler M, Fayet O (1993) Translational frameshifting in the control of transposition in bacteria. *Mol Microbiol* 7:497-503

Chen X, Li S, Aksoy S (1999) Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *J Mol Evol* 48:49-58

Chen XZ, Peng JB, Cohen A, Nelson H, Nelson N, Hediger MA (1999) Yeast SMF1 mediates H(+)-coupled iron uptake with concomitant uncoupled cation currents. *J Biol Chem* 274:35089-35094

- Curie C, Alonso JM, Le Jean M, Ecker JR, Briat JF (2000) Involvement of NRAMP1 from *Arabidopsis thaliana* in iron transport. *Biochem J* 347 Pt 3:749-755
- Dale C, Welburn SC (2001) The endosymbionts of tsetse flies: manipulating host-parasite interactions. *Int J Parasitol* 31:628-631
- De Lamballerie, X, Zandotti C, Vignoli C, Bollet C, De Micco P (1992) A one-step microbial DNA extraction method using "Chelex 100" suitable for gene amplification. *Res Microbiol* 143:785-790
- Demuth DR, Duan Y, Jenkinson HF, McNab R, Gil S, Lamont RJ (1997) Interruption of the *Streptococcus gordonii* M5 *sspA/sspB* intergenic region by an insertion sequence related to IS1167 of *Streptococcus pneumoniae*. *Microbiology* 143 (Pt 6):2047-2055
- Doolittle RF (2002) Biodiversity: microbial genomes multiply. *Nature* 416:697-700
- D'Souza J, Cheah PY, Gros P, Chia W, Rodrigues V (1999) Functional complementation of the malvolio mutation in the taste pathway of *Drosophila melanogaster* by the human natural resistance-associated macrophage protein 1 (Nramp-1). *J Exp Biol* 202 (Pt 14):1909-1915
- Fleming MD, Trenor CC, III, Su MA, Foernzler D, Beier DR, Dietrich WF, Andrews NC (1997) Microcytic anaemia mice have a mutation in Nramp2, a candidate iron transporter gene. *Nat Genet* 16:383-386

Forbes JR, Gros P (2001) Divalent-metal transport by NRAMP proteins at the interface of host- pathogen interactions. *Trends Microbiol* 9:397-403

Gaucher EA, Gu X, Miyamoto MM, Benner SA (2002) Predicting functional divergence in protein evolution by site-specific rate shifts. *Trends Biochem Sci* 27(6):315-321

Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A (2002) Extreme genome reduction in *Buchnera spp.*: toward the minimal genome needed for symbiotic life. *Proc Natl Acad Sci U S A* 99:4454-4458

Glockner G, Eichinger L, Szafranski K, Pachebat JA, Bankier AT, Dear PH, Lehmann R, Baumgart C, Parra G, Abril JF, Guigo R, Kumpf K, Tunggal B, Cox E, Quail MA, Platzer M, Rosenthal A, Noegel AA (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature* 418:79-85

Govoni G, Canonne-Hergaux F, Pfeifer CG, Marcus SL, Mills SD, Hackam DJ, Grinstein S, Malo D, Finlay BB, Gros P (1999) Functional expression of Nramp1 in vitro in the murine macrophage line RAW264.7. *Infect Immun* 67:2225-2232

Gruenheid S, Pinner E, Desjardins M, Gros P (1997) Natural resistance to infection with intracellular pathogens: the Nramp1 protein is recruited to the membrane of the phagosome. *J Exp Med* 185:717-730

- Gunshin H, Mackenzie B, Berger UV, Gunshin Y, Romero MF, Boron WF, Nussberger S, Gollan JL, Hediger MA (1997) Cloning and characterization of a mammalian proton-coupled metal-ion transporter. *Nature* 388:482-488
- Gu X, Zhang J (1997) A simple method for estimating the parameter of substitution rate variation among sites. *Mol Biol Evol* 14(11):1106-1113
- Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177:4121-4130
- Hackam DJ, Rotstein OD, Zhang W, Gruenheid S, Gros P, Grinstein S (1998) Host resistance to intracellular infection: mutation of natural resistance-associated macrophage protein 1 (Nrampl) impairs phagosomal acidification. *J Exp Med* 188:351-364
- Heine HS, England MJ, Waag DM, Byrne WR (2001) In vitro antibiotic susceptibilities of *Burkholderia mallei* (causative agent of glanders) determined by broth microdilution and E-test. *Antimicrob Agents Chemother* 45:2119-2121
- Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77:51-59
- Hou YM (1999) Transfer RNAs and pathogenicity islands. *Trends Biochem Sci* 24:295-298

- Hu WS, Yang CC (2001) IS1550 from *Mycoplasma fermentans* is transposable in *Escherichia coli*. FEMS Microbiol Lett 198:159-164
- Jabado N, Jankowski A, Dougaparsad S, Picard V, Grinstein S, Gros P (2000) Natural resistance to intracellular infections: natural resistance-associated macrophage protein 1 (Nramp1) functions as a pH-dependent manganese transporter at the phagosomal membrane. J Exp Med 192:1237-1248
- Jones DT, Taylor WR, Thornton JM (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 33:3038-3049
- Kehres DG, Zaharik ML, Finlay BB, Maguire ME (2000) The NRAMP proteins of *Salmonella typhimurium* and *Escherichia coli* are selective manganese transporters involved in the response to reactive oxygen. Mol Microbiol 36:1085-1100
- Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55:709-742
- Korbel JO, Snel B, Huynen MA, Bork P (2002) SHOT: a web server for the construction of genome phylogenies. Trends Genet 18:158-162
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244-1245

Lawrence JG, Hendrix RW, Casjens S (2001) Where are the pseudogenes in bacterial genomes?. *Trends Microbiol* 9:535-540

Mahillon J, Leonard C, Chandler M (1999) IS elements as constituents of bacterial genomes. *Res Microbiol* 150:675-687

Makui H, Roig E, Cole ST, Helmann JD, Gros P, Cellier MF (2000) Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter. *Mol Microbiol* 35:1065-1078

Meeks JC, Elhai J (2002) Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiol Mol Biol Rev* 66:94-121

Minerdi D, Fani R, Gallo R, Boarino A, Bonfante P (2001) Nitrogen fixation genes in an endosymbiotic *Burkholderia* strain. *Appl Environ Microbiol* 67:725-732

Moulin L, Munive A, Dreyfus B, Boivin-Masson C (2001) Nodulation of legumes by members of the beta-subclass of Proteobacteria. *Nature* 411:948-950

Pinner E, Gruenheid S, Raymond M, Gros P (1997) Functional complementation of the yeast divalent cation transporter family SMF by NRAMP2, a member of the mammalian natural resistance-associated macrophage protein family. *J Biol Chem* 272:28933-28938

- Que Q, Helmann JD (2000) Manganese homeostasis in *Bacillus subtilis* is regulated by MntR, a bifunctional regulator related to the diphtheria toxin repressor family of proteins. *Mol Microbiol* 35:1454-1468
- Rost B (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol* 266:525-539
- Rouhbakhsh D, Baumann P (1995) Characterization of a putative 23S-5S rRNA operon of *Buchnera aphidicola* (endosymbiont of aphids) unlinked to the 16S rRNA-encoding gene. *Gene* 155:107-112
- Rubens CE, Heggen LM, Kuypers JM (1989) IS861, a group B streptococcal insertion sequence related to IS150 and IS3 of *Escherichia coli*. *J Bacteriol* 171:5531-5535
- Sanchez-Beato AR, Garcia E, Lopez R, Garcia JL (1997) Identification and characterization of IS1381, a new insertion sequence in *Streptococcus pneumoniae*. *J Bacteriol* 179:2459-2463
- Silva FJ, Latorre A, Moya A (2001) Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends Genet* 17:615-618
- Smoot JC, Barbian KD, Van Gompel JJ, Smoot LM, Chaussee MS, Sylva GL, Sturdevant DE, Ricklefs SM, Porcella SF, Parkins LD, Beres SB, Campbell DS, Smith TM, Zhang Q, Kapur V, Daly JA, Veasy LG, Musser JM (2002) Genome sequence and comparative microarray analysis of serotype M18 group A

Streptococcus strains associated with acute rheumatic fever outbreaks. Proc Natl Acad Sci U S A 99:4668-4673

Thomine S, Wang R, Ward JM, Crawford NM, Schroeder JI (2000) Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to Nramp genes. Proc Natl Acad Sci U S A 97:4991-4996

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-4882

Ulett GC, Currie BJ, Clair TW, Mayo M, Ketheesan N, Labrooy J, Gal D, Norton R, Smith CA, Barnes J, Warner J, Hirst RG (2001) *Burkholderia pseudomallei* virulence: definition, stability and association with clonality. Microbes Infect 3:621-631

Utaisincharoen P, Tangthawornchaikul N, Kespichayawattana W, Chaisuriya P, Sirisinha S (2001) *Burkholderia pseudomallei* interferes with inducible nitric oxide synthase (iNOS) production: a possible mechanism of evading macrophage killing. Microbiol Immunol 45:307-313

Website references

PredictProtein interface [<http://cubic.bioc.columbia.edu/predictprotein/>]

Psipred v2.4 protein structure prediction server [<http://bioinf.cs.ucl.ac.uk/psipred/>]

ExpASy Molecular Biology Server [<http://ca.expasy.org/>]

The Preliminary Directory Of Dictyostelium Genes - Version 2.1 Gene

[<http://dicty.sdsc.edu/annotationdicty.html>]

Berkeley Drosophila Genome Project port and the Gadfly database

[<http://www.fruitfly.org/annot/index.html>]

Wormbase [<http://www.wormbase.org/>]

[http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/euk_g.html]; Anopheles gambiae

Figure legends

Figure1. Phylogenetic analysis of the Nramp/MntH family. The circle representation establishes the most probable tree topology based on sequence relationships calculated using the software MEGA2 (Kumar et al. 2001) and multiple sequence alignments performed with the program ClustalX (Thompson et al. 1997). Pairwise distance matrix calculations were performed using a Poisson correction for multiple substitutions, and the dendrogram generated using the Minimal Evolution calculation (Kumar et al. 2001) and tested for statistical significance by bootstrapping (Kumar et al. 2001). Only the weakest nodes are indicated for clarity; their presence indicate possible alternative local topologies. Systematic names for proteins are composed of the initial of genus name in capital, one to three letters from the species names followed by NR for eukaryotic Nramp proteins or A, B, or C α , C β , and C γ for bacterial MntH proteins. Numbers have been added in multigenic organisms.

Archaea: Nphar: *Natromonas pharaonis*.

Bacteria : Atum: *Agrobacterium tumefaciens*; Bant: *Bacillus anthracis*; Bce:

Burkholderia cepacia; Bcer: *B. cereus*; Bfra: *Bacteroides fragilis*; Bfu: *B.*

fungorum; Bma: *B. mallei*; Bps: *B. pseudomallei*; Bsub: *B. subtilis*; Bsui:

Brucella suis; Cace: *Clostridium acetobutylicum*; Cdip: *Corynebacterium*

diphtheriae, plasmid; Chyd: *Carboxydotherrmus hydrogenoformans*; Ctep:

Chlorobium tepidum; Drad: *Deinococcus radiodurans*; Echr: *Erwinia*

chrysanthemii; Ecol: *Escherichia coli*; Efa: *Enterococcus faecalis*; Efc: *E.*

faecium; Lbre: *Lactobacillus brevis*; Lino: *Listeria innocua*; Llac: *Lactococcus*

lactis; Lmon: *L. monocytogenes*; Mavi: *Mycobacterium avium*; Mlep: *M. leprae*;

Mlot: *Mesorhizobium loti*; Mmag: *Magnetospirillum magnetotacticum*; Msme: *M.*

smegmatis; Mtub: *M. tuberculosis*; Nost: *Nostoc* sp. PCC 7120; Npun: *N. punctiforme*; Pae: *Pseudomonas aeruginosa*; Pflu: *P. fluorescens*; Psyr: *P. syringae*; Rpal: *Rhodopseudomonas palustris*; Saur: *Staphylococcus aureus*; Scri: *Streptococcus criceti*, plasmid; Sepi: *S. epidermidis*; Sgor: *S. gordonii*; Smi: *S. mitis*; Smu: *S. mutans*; Stym: *Salmonella typhimurium*; Tten: *Thermoanaerobacter tengcongensis*; Wbre: *Wigglesworthia brevialpilis*; Xaxo: *Xanthomonas axonopodis*; Xcam: *X. campestris*; Xfas: *Xylella fastidiosa*; Yent: *Yersinia enterocolitica*; Ypes: *Y. pestis*.

Eukarya: Ag: *Anopheles gambiae*; At: *Arabidopsis thaliana*; Bt: *Bos taurus*; Ca: *Candida albicans*; Ce: *Caenorhabditis elegans*; Cre: *Chlamydomonas reinhardtii*; Dd: *Dictyostelium discoideum*; Dm: *Drosophila melanogaster*; Gg: *Gallus gallus*; Hs: *Homo sapiens*; Ncr: *Neurospora crassa*; Om: *Oncorhynchus mykiss*; Os: *Oryza sativa*; Mm: *Mus musculus*; Sc: *Saccharomyces cerevisiae*.

Figure 2. Genomic organization of *mntH C β* locus in *Wigglesworthia*. **A.**

Wigglesworthia MntH C β protein is more similar to proteins found in Gram positive bacteria (indicated in blue color), whereas genes surrounding *mntH C β* encode sequences and/or proteins more similar to homologs found in other Enterobacteriaceae (indicated in red for closeness to *E. coli*, and pink for closeness to *Buchnera* spp. (Gil et al. 2002)). They encode various functions, including uptake systems for metal cations: a *znu* operon for Zn(II) and *trkH* for K⁺; redox active enzymes: a protoporphyrinogen oxidase of the biosynthetic pathway of the haem group, and other enzymes involved in ubiquinone synthesis deriving from the shikimate pathway in bacteria (e.g. 3-octaprenyl-4-hydroxybenzoate carboxy-lyase). The gene *ruvC*,

involved in the resolution of crossing over intermediates was found together with genes encoding other ubiquitous functions, the stress proteins GrpE and Gapdh. **B.** Variation of GC content along the DNA fragment shows that high GC percentage is restricted to the region encoding ribosomal genes, similarly to *Buchnera* ribosomal operon (Rouhbakhsh et al. 1995). **C.** Schematic representation of *E. coli* chromosomal fragments that are co-linear with *Wigglesworthia* DNA fragment (represented by a thin line and *E. coli* chromosome coordinates; crosses indicate probable discrete deletions) and of the previously unknown gene block *Asn-tRNA^{mntH} C β* (indicated by a hatched box).

Figure 3. Linkage of oral streptococcaceae *mntH* C β genes with different IS-like elements. (A) *S. mutans* single chromosomal *mntH* locus (coordinates : 712001-724000 bp); *mntH* gene is flanked by an IS861-like sequence in opposite orientation. IS861 was identified in Lancefield's Group B streptococcaceae (GBS, *S. agalactiae* (Rubens et al. 1989)) and is shared by Group A and oral streptococcaceae. IS861-like elements encode putative transposases sharing similarity to those of *E. coli* IS3 and IS150 (Rubens et al. 1989). (B) One *S. mitis* *mntH* gene is flanked by a distinct IS861-like sequence in the same orientation, which contains the sequence AAAAAAG located near the 3' end of transposase Orf A that may induce a -1 translational frameshift resulting in the fusion of Orfs A and B (Chandler et al. 1993; Hu et al. 2001). A second *S. mitis* *mntH* gene (C) is preceded by an IS1381-like element in the same orientation. IS1381 was identified in *S. pneumoniae* and few homologs have been identified in selected species including GBS, *Lactobacillus*, *Cyanobacteria* spp. (e.g. *N. punctiforme*). (D) A *S. gordonii* gene is followed by an IS1562-like element in

same orientation. These insertion sequences are known to be present in different genomes that do not possess *mntH* genes: IS1562 and IS861 in GAS, IS1381 in GBS.

Figure 4: A. Consensus transmembrane topology predicted for Nramp and MntH proteins. Symbols in blue color indicate the intracytoplasmic end of the predicted transmembrane domain (TMD; vertical bar), and the outer membrane end of the TMD (filled symbol). The presence of a 12th TMD changes does not affect the predicted orientation of the N-terminus of the protein. B. Alignment of the hydrophobicity plots (Kyte-Doolittle scale, window 19) calculated with the full length protein sequences NpunC α , CreNR, LlacC β , DdNR1, EcolA, LlacB. A black asterisk on the profile of LlacB and C β proteins indicate the position of truncation performed, which correspond to the premature stop codon found in LlacB; a grey asterisk on the profile of LlacC β protein symbolises the position of the N-terminal truncation performed on the close homolog EfaC β 1 protein. All the profiles display significant similarity with variations in the regions of TMD 1, 3, 6, 8-10, and 12. MntH A and B proteins profiles are more compact due to shorter sequences.

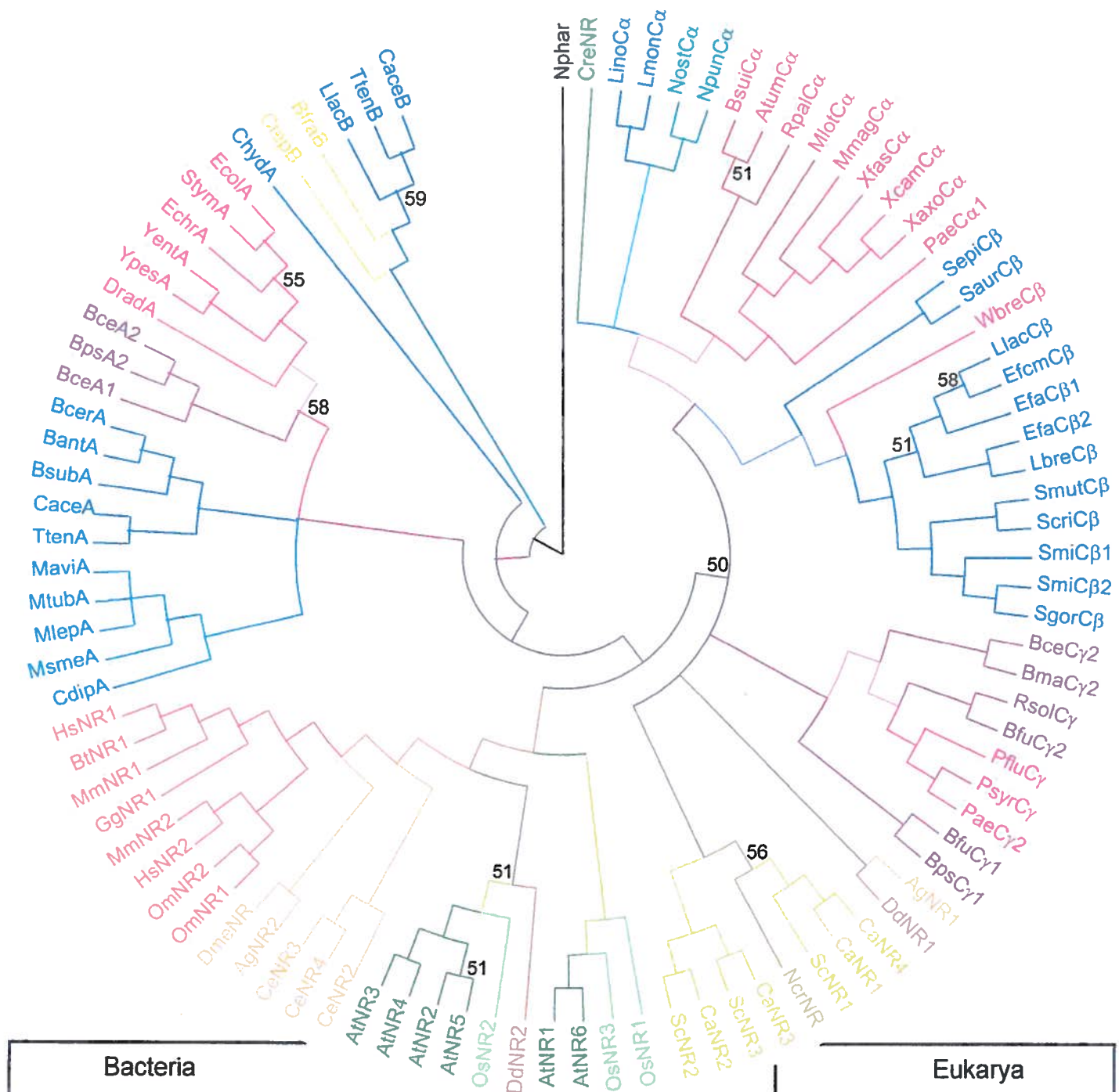
Tables

Table 1 Disk assay of metal sensitivity conferred by expression of MntH A B, C in *E. coli*

Protein	MnCl ₂ 1M	FeSO ₄ 1M	CoCl ₂ 0,25M	CdCl ₂ 0,25M
none	0	13,5±0,4	10,75±0,25	20,6±0,3
EcolA	14,5±0,5	27,5±0,6	30±0,2	37±0,4
EcolA Blam-179	0	15±0,5	16±0,5	23±0,5
CtepB	0	20,25±0,5	21,5±0,2	22±1
LlacB	0	14±0,5	11,5±0,5	35,5±0,7
LlacB Δ-COOH	0	14±0,5	13,25±0,75	28±1,2
LlacCβ	0	13±0,2	12±0,2	25±0,8
LlacCβ Δ-COOH	0	13,75±0,5	10,5±0,5	31,25±0,75
Efa Cβ1* Δ-NH ₂	0	14,25±0,25	15,5±0,5	20±0,2
Efa Cβ1*	0	20,5±0,7	25±0,2	37,5±0,5
Efa Cβ2*	0	14±0,5	29,75±0,25	30,6±5

* bacterial growth less dense compared to controls (pBAD, EcolA)

6×10^7 cells were seeded in top agar on GTA plates containing 0.06% arabinose (0.0006 % for Efa Cβ1 and 2 proteins) and a filter disk impregnated with 10 μl of metal ion solution and were incubated for 16 h at 37°C. The diameter of metal-induced growth inhibition are indicated in mm (mean and standard error of three independent experiments).

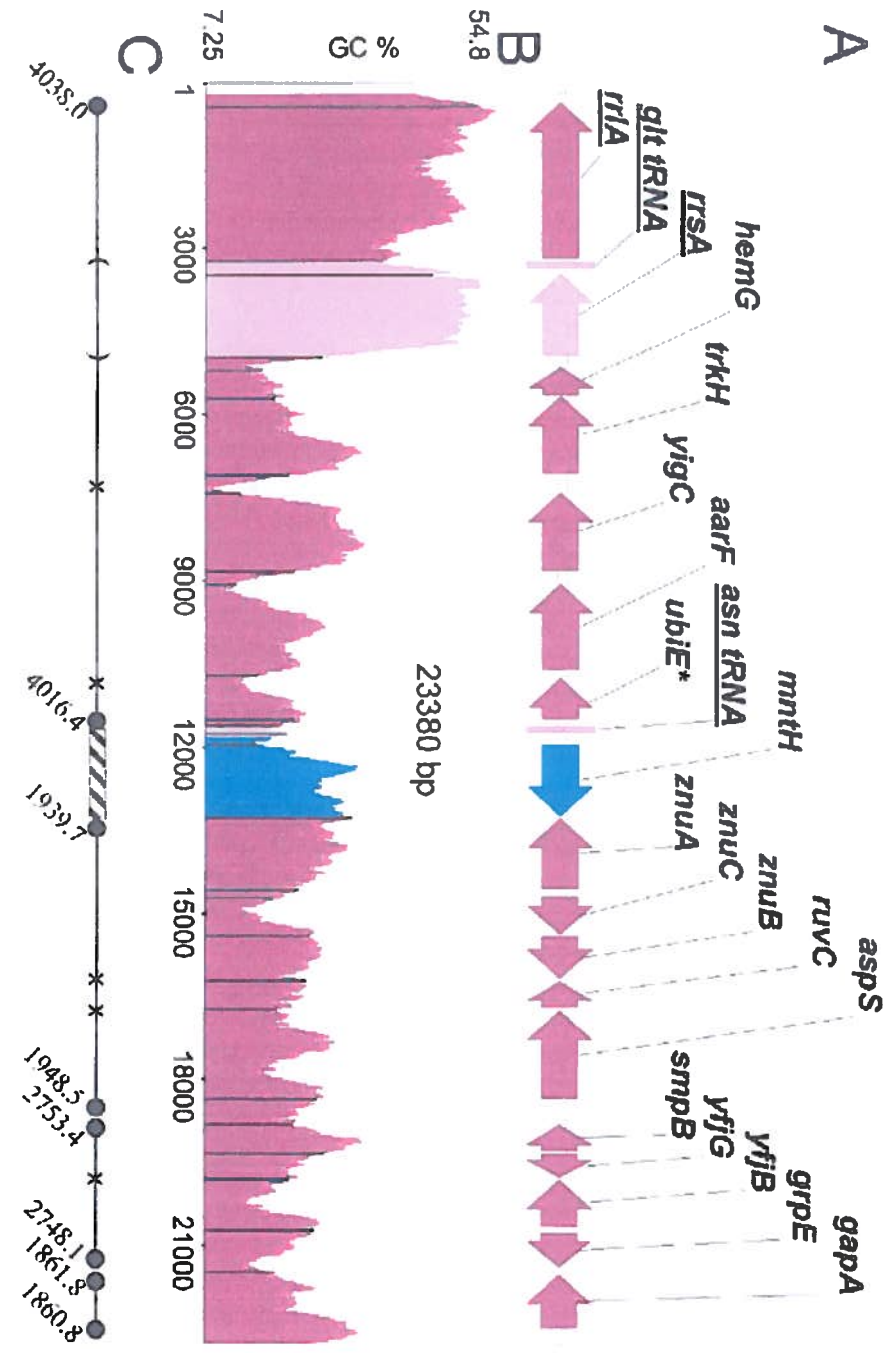


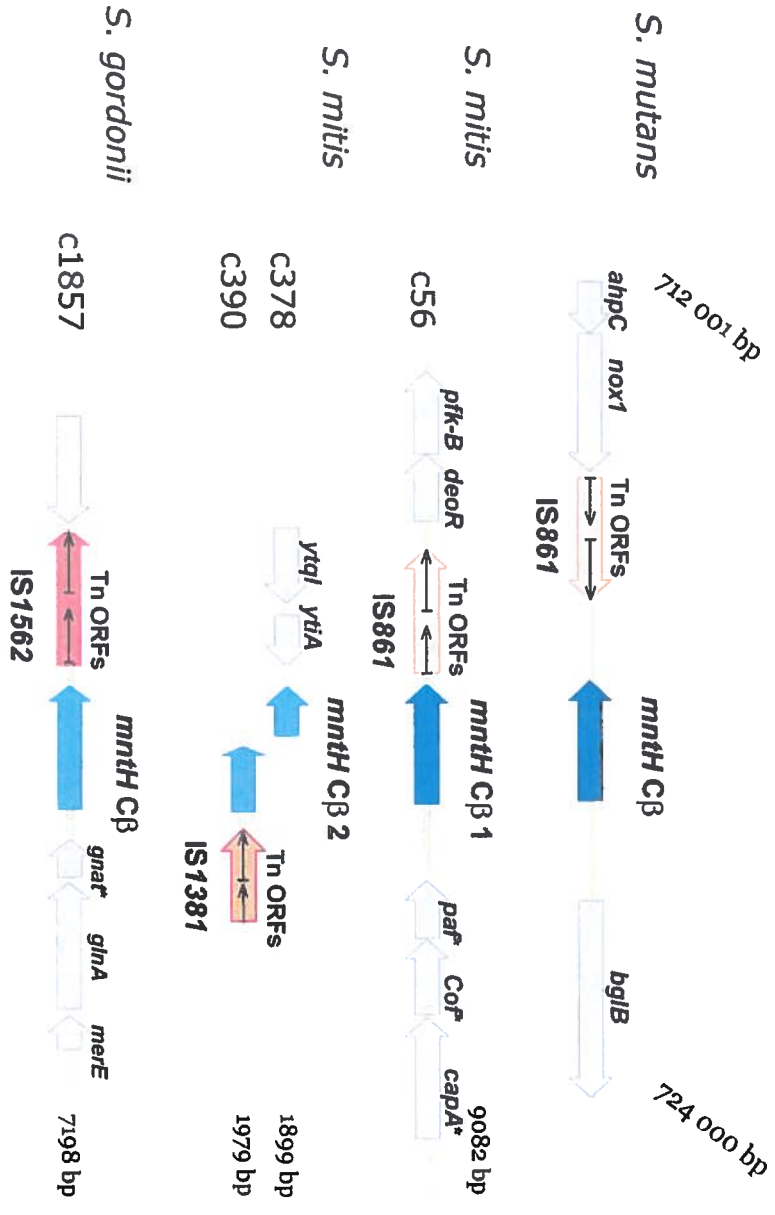
Bacteria

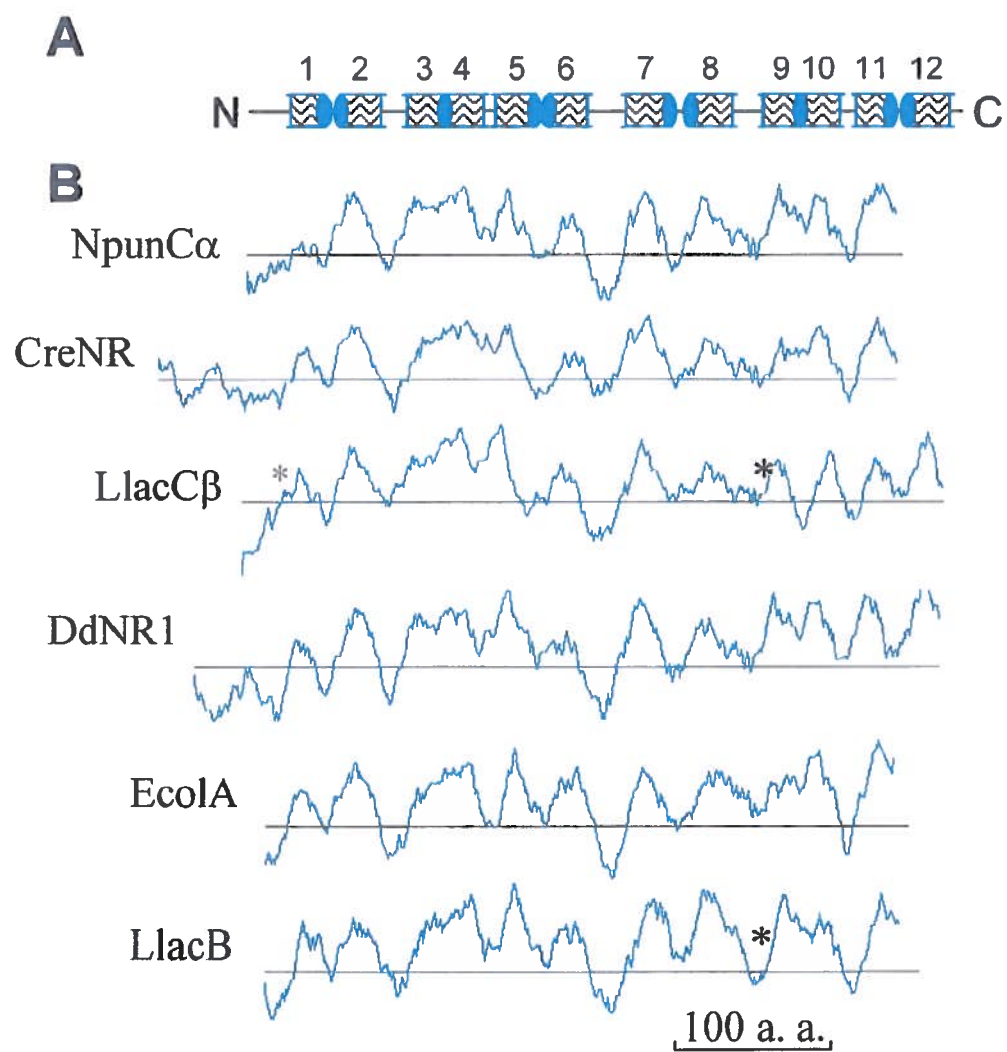
- CFR/Green Sulfur Bacteria Group
- Cyanobacteria
- Thermus/Deinococcus group
- Actinobacteria (high G+C gram-positive bacteria)
- Bacillus/Clostridium (high G+C gram-positive bacteria)
- α-Proteobacteria
- β-Proteobacteria
- γ-Proteobacteria

Eukarya

- Mycetozoa
- Fungi (Sordariomycetes)
- Fungi (Saccharomycetes)
- Chlorophyta
- Streptophyta (monocot)
- Streptophyta (eudicot)
- Insecta
- Vertebrata







Annexe 3

Article 2: Determination of the transmembrane topology of *Escherichia coli* Nramp ortholog (2003) Journal of Biological Chemistry in press.

Determination of the transmembrane topology of *Escherichia coli*

Nramp ortholog

Courville Pascal ^a, Chaloupka Roman ^{a,b}, Veyrier Frédéric ^a and Mathieu FM Cellier ^{a,*}.

^a INRS-Institut Armand-Frappier, 531 Bd. des prairies, Laval, Québec, CANADA H7V 1B7

^b Institute of Physics, Charles University in Prague, Ke Karlovu 5, 121 16 Prague 2, Czech Republic

Running title: *E. coli* MntH A transmembrane topology.

Keywords: Type III integral membrane protein; transmembrane topology; computer-aided prediction; transport; divalent metal; proton.

* To whom correspondence should be addressed:

INRS-Institut Armand-Frappier

531, Bd. des Prairies

Laval, Québec, CANADA H7V 1B7

e-mail: mathieu.cellier@inrs-iaf.quebec.ca

Tel: (450) 687-5010 ext. 4681; Fax:(450) 686-5301

Abstract

The Natural resistance-associated macrophage protein (Nramp) defines a conserved family of secondary metal transporters. Molecular evolutionary analysis of Nramp family revealed the early duplication of an ancestral eukaryotic *Nramp* gene which was likely derived from a bacterial ortholog, characterized as a proton-dependent manganese transporter (1). *Escherichia coli* MntH represents a model of choice to study structure function relationship in Nramp protein family. Here we report *E. coli* MntH transmembrane topology using a combination of in silico predictions, genetic fusion with cytoplasmic and periplasmic reporters, and MntH functional assays. Blam constructs revealed extraloops between TMD 1/2, 5/6, 7/8, 9/10 and placed the C-terminus periplasmic; Cat constructs indicated cytoplasmic loops 2/3, 6/7, 8/9 and 10/11. Two intraloops for which no data were produced (N-terminus, intraloop 4/5) both display composition bias supporting their deduced localization. The extraloop 5/6, 6/7 and periplasmic exposure of the C-terminus were confirmed by targeted reporter insertion. Three of them preserved MntH function measured by disk-assay of divalent metal uptake and fluorescence assay of divalent metal-dependent proton transport whereas a truncated form lacking TMD 11 was inactive. These results demonstrate EcolIA is a type III integral membrane protein with 11 transmembrane domains transporting both divalent metal ions and protons.

Introduction

Divalent metal ions such as ferrous iron, Fe^{2+} , Mn^{2+} , Co^{2+} and Zn^{2+} are vital nutrients for living cells that participate as metabolic cofactors to a variety of biochemical processes involving electron transfers, including respiration and photosynthesis. The Natural Resistance Associated Macrophage Protein (Nramp) belongs to a highly conserved family of integral membrane proteins found in a large spectrum of organisms, including mammals (2; 3), plants (4), yeast (5), and bacteria (6; 7). Eukaryotic Nramp proteins were implicated in pH-dependent transport of divalent metals including Fe^{2+} and Mn^{2+} (8; 9). Mammalian Nramp2 facilitates transferrin-independent iron absorption in the intestine, and transferrin-dependent iron uptake by peripheral tissues (e.g. bone marrow, erythrocytes, kidney), whereas Nramp1 is devoted to host resistance, acting at the level of the membrane of the phagosome in macrophages and neutrophils (10). Likewise, three yeast homologs contribute to Mn and Fe homeostasis by facilitating Mn acquisition either at the plasma membrane, or from intracellular vesicles, and iron mobilization from the vacuole (11). The phenotypes associated with *Nramp1* and *Nramp2* knockout in mice, innate susceptibility to various intracellular pathogens and microcytic anemia, respectively, demonstrate non-redundant roles of mammalian Nramp proteins in divalent metal homeostasis. Deregulation of their expression may also perturb normal metabolism due to the cytotoxic effects of excessive metal accumulation (12).

Bacterial homologs of eukaryotic Nramp were subsequently characterized in Gram positive and negative species as a proton-dependent manganese transporter and were denominated MntH. Detailed sequence analyses revealed the existence of three phylogenetic groups of MntH proteins that showed distinct evolutionary patterns and were designated MntH A, B and C (13). One of three phylogenetic groups of bacterial homologs (1) is distributed among Gram positive and negative Bacteria and shows congruency between function, phylogenetic and taxonomic relationships and amino acid substitution rate pattern, consistent with an early evolutionary origin (14). Several MntH A proteins were functionally

characterized: In Gram positive species, *Bacillus subtilis* MntH A is necessary for growth in minimal medium not supplemented with manganese and *mntH* gene expression is regulated by the Mn-dependent repressor, MntR (15). Mycobacterial MntH A were characterized by heterologous expression in *Xenopus* oocytes (16) and yeast (17). MntH A proteins were also studied in Gram negative species; *E. coli* and *S. typhimurium* MntH A were characterized in our laboratory as proton-dependent transporters of divalent metals with strong preference for Mn^{2+} (1; 18). MntH proteins B and C have also been studied recently; one MntH C protein was reported to contribute to *S. aureus* virulence (19) and several *mntH* genes from groups B and C were functionally expressed in *E. coli* and shown to confer sensitivity to divalent metals (14). Therefore, despite significant distance between currently known Nramp homologs, their sequence preserved structural features that correspond to a conserved function in proton-dependent divalent metal uptake.

Based on their evolutionary features, the prokaryotic genes of *mntH* group A represent more likely precursors of eukaryotic *Nramp* genes (13; 14); the corresponding proteins should thus exhibit similar structure because they perform similar function (20). Determination of the transmembrane topology is a prerequisite to study the structural and functional properties of a membrane transporter (21) and a combination of in silico predictions and experimental validation is desirable to establish a reliable topological model (22). Initially, in absence of functional hint, extensive sequence analysis of eukaryotic Nramp homologs yielded a consensus transmembrane topology that was based on several conserved features: the presence of ten hydrophobic segments showing a remarkable degree of amino acid sequence conservation, including for some the presence of charged residues and overall amphiphilic character, the strong net positive charge for hydrophilic loops predicted to be cytoplasmic, and the presence of a potential glycosylation signal in a predicted extracytoplasmic loop and a putative conserved transport motif. This putative topology placed the N-terminus cytoplasmic, followed by a conserved hydrophobic core of ten transmembrane domains (TMD), and either one or two none conserved highly hydrophobic

TMDs (placing the C-terminus extra- or intra-cytoplasmic, respectively).

The proposed topology has been verified in different regions of different homologs. The orientation of yeast Smf3p TMD 11 has been determined by inserting a hemagglutinin (HA) tag at either end of this segment. Only the carboxy-terminal insertion was removed by vacuolar proteases consistent with the location of Smf3p in the membrane of yeast vacuole. The data indicated that the last TMD in yeast Smf proteins is oriented from the cytoplasm toward the exterior, placing the C-terminus extracytoplasmic (11). Both the N and C-termini of the mammalian protein Nramp1 were localized on the cytoplasmic side of the membrane using either a polyclonal serum generated against the N-terminal portion of the protein or by creating a c-myc tag insertion at the C-terminus (23). The same group found, using the paralogous Nramp2, that the loop between the TMD 7/8 was extracellular after inserting a HA tag (24). Recently, a study indicated that 24-residue peptide corresponding to Nramp2 putative TMD 4 (residues 179-202) adopted a well folded α -helical conformation from residue V₁₈₇ to K₂₀₁ in detergent micelles (25).

Alternative topologies have been proposed; one study suggested that mouse Nramp1 could adopt a reverse transmembrane topology (N and C-termini extra-cytoplasmic) based on findings that incubation with antibodies directed against either the predicted extra-loop between TMD 7/8 or the C-terminus inhibited or not, respectively, temperature-dependent association of iron with purified phagosomes (26).

Concerning bacterial homologs, a study of the reactivity of the unique cysteine of *Mycobacterium leprae* MntH A protein proposed a transmembrane topology inferred from sequence-based predictions that included 12 TMD (17). These data raise the issue of the number of TMD in prokaryotic MntH proteins and of the similarity of prokaryotic MntH and eukaryotic Nramp and Smfp transmembrane organisation.

In the present study, we investigated the transmembrane topology of *E. coli* Nramp ortholog MntH A (EcolIA) which constitutes an ideal structural and functional model of eukaryotic Nramp to study using

different genetic approaches. We employed a combination of in silico predictions, reporter gene fusions and insertion mutagenesis, including bi-functional hybrid proteins for which metal uptake and metal-dependent proton transport were evaluated.

Experimental procedures

Bacterial strains and growth conditions

DH11S *mntH E. coli* cells (Kan^r) were used in the standard cloning procedures. Bacteria were grown at 37°C and 250 rpm in Luria-Bertani (LB) broth supplemented with the appropriated antibiotics (ampicillin 100 µg/mL, kanamycin 30 µg/mL or gentamicin 15 µg/mL). The GTA medium was used for the disc assay and complementation experiment using SL93 *mntH E. coli* strain. (Proposed location for table 1 and 2)

Sequence analyses

Three programs were used to formulate a consensus prediction topology of MntH A protein: TMHMM (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>), PHDhtm (<http://cubic.bioc.columbia.edu/predictprotein/>), and MEMSAT (<http://bioinf.cs.ucl.ac.uk/psiform.html>). Hydrophobicity profiles were generated using the method of Kyte and Doolittle (e.g. <http://www.bio.davidson.edu/courses/compbio/flc/home.html>) and implementing a sliding window of 19 residues.

Construction and analysis of random Blam fusions

pBAD NX1.1 (27) was subcloned in pYZ4 (28). Nested 3' deletions of gene *mntH* were obtained by digesting pYZ4 NX1.1 with *Xba* I and *Sph* I and treated with Exonuclease III using the Exo-size™ Deletion Kit (NEB biolabs, Mississauga, ON) according to the manufacturer instructions. The fusion at the amino acid A165 was obtained by digesting pYZ4 NX1.1 with *Sac* II, treating with T4 DNA polymerase to create blunt ends conserving the reading frame. The different preparation were digested with *Sac* I, purified and ligated with the 850 bp secreted Blam ORF, previously obtained by digestion with *Pvu* II and *Sac* I of the plasmid pYZ5 (27). The ligation product was transformed in *E. coli* competent cell

DH11S and Kan^r transformants were selected. The colonies were scratched on three plates containing respectively Kan. (30µg/mL), Amp. (20 and 100µg/mL). All plates contained 10 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and Amp^r, Kan^r colonies were selected. The Minimal Inhibitory Concentration (MIC) of Amp for the MntH-Blam fusions were determined by spotting 4 µl of a diluted 16h liquid culture (10⁻⁵ and 10⁻⁶) on plates containing different concentrations of ampicillin and supplemented with IPTG. Plasmids were isolated from selected clones (>45µg/mL amp.) and analyzed with restriction enzymes to determine the approximate position of the fusion. Those corresponding to novel positions were sequenced using automatic sequencer CEQ 2000XL from Beckman Coulter (Mississauga, ON).

Construction and analysis of targeted Cat fusions

Cat constructions were created by targeted PCR fusion inserting a *Bgl* II sites after the amino acid L324 and K387. The products were digested with *Sac* II and *Bgl* II and ligated to the vector pBADNX1.1 digested with the *Sac* II and *Xba* I and to the *cat* gene amplified by PCR and containing 5' *Bam*H I and a 3' *Xba* I sites. The fusions at residue R91, G226 and L412 were constructed using *Nco* I and *Bam*H I sites except for R91 (*Bgl* II site). The different fusions were tested for resistance to chloramphenicol (Cml) using a 16 hour liquid culture to inoculate 2 mL of LB medium (1/10) and grown for 3 hours under agitation. Cultures were then induced to express the constructs with 0.06% L-arabinose for 1 hour and 4 µl of a dilution 10⁴ were spotted on plates containing different Cml concentrations and 0.2% L-arabinose (except for the construct L412-Cat, 6x10⁻⁵ % L-arabinose was used). Constructs BlaM-371 and BlaM-404 were subcloned in pBADNX1.1 and pBADNX7.1, respectively, using *Nco* I and *Sph* I restriction sites.

Construction and analysis of targeted insertions

Targeted Blam insertions were obtained by PCR overlap extension to introduce a *Bam*H I site (29) after the amino acid L179, F263 and L412. The final products were cloned in pBADGmΔ*Bam*H I using *Nco* I

and *Xba* I restriction sites. The mutants were digested with *Bam*H I and ligated to *Bam*H I Blam ORF. The hybrid proteins were subcloned in pBADNX7.1 and ampicillin MIC were obtained as described above using subtoxic concentration of L-arabinose as inducer of expression. The PCR overlap extension approach was also used to generate a C-terminal c-myc fusion (EQKLISEEDL). The toxicity of the constructs was estimated by growing a LB culture for 16 h at 37°C and 250rpm which was then diluted in fresh LB broth to obtain a final optical density (OD₆₀₀) of 0.5 and subcultured in presence of different concentrations of L-arabinose during 6h. The level of toxicity of the mutants was estimated from the final OD₆₀₀.

Western blot analysis

Cytoplasmic and membrane fractions were prepared for each random fusion (30). The samples were solubilised in SDS dissolving buffer [4% (w/v) SDS, 40% (v/v) glycerol, 0.5M mercaptoethanol, 40mM Tris-Hcl, pH7.2, 0.005% (w/v) bromphenol blue] and incubated at 37°C for 30 min. Proteins were separated by electrophoresis using a 12% polyacrylamide gel (SDS-Page). The protein were transferred on a sequi-blot (polyvinylidene fluoride) PVDF membrane from Bio-Rad (Mississauga, ON) at 60 mA during 1 hour, washed with distilled H₂O and blocked for 16h at 4°C in a PBS 1X-Tween 0.1% milk 5% solution. Membrane were incubated with the primary antibody, during 1 hour, washed four times for 12 minutes in PBS Tween 0.1%, and incubated for 1 hour with the secondary antibody (anti-IgG1-HRP, (Santa Cruz, CA); 1/1000) and four washes of 12 minutes in PBS Tween 0,1%. Blots were revealed using the ECL Western blotting reagent from Amersham (Baie d'Urfé, Qc). The primary antibody used are anti β -lactamase monoclonal antibody (Biodesign intl., Saco, ME); 1/500), anti c-myc monoclonal (antibody Biodesign intl.; 1/100), anti Cat polyclonal antibody (Sigma; 1/700).

Protein over-expression and preparation of mixed membranes

Membranes extracts were obtained by preparation of *E. coli* mixed membranes using a water lysis

protocol (31). In brief, 50mL of *E. coli* culture induced for 16h with 0.06% of L-arabinose were harvested and lysed in sucrose buffer containing EDTA and lysozyme. The spheroplasts were harvested and incubated for 30 min in deionized water. The mixed membranes were then washed and resuspended in 400 μ L of membrane resuspension buffer (0.1M sodium phosphate pH 7.2, 1mM 2-mercaptoethanol). 120 μ g of total proteins were separated on SDS-PAGE and Western blot was performed as described above. Protein concentrations were determined using the Non-Interfering protein assay by Genotech (MO, USA)

MntH A functional assays

MntH-dependent metal uptake was evaluated by disk assay of metal sensitivity. Cells were grown during 16h at 37°C and 250rpm. The cultures were then diluted in the GTA broth to obtain a final optical density (OD₆₀₀) of 0.55. 100 μ L of bacteria was mixed to 3 mL of molted GTA agar and spread onto GTA plates supplemented with the appropriate concentration of arabinose prior to placing a disk impregnated with 10 μ L of metallic solution and incubating for 16h at 37°C or 42°C (1).

MntH- and divalent metal-dependent intracellular accumulation of protons was determined by measuring fluorescence intensity ratio of pH-dependent ratiometric GFP (pHluorin) (32; 33).

Electro-competent cells of *E. coli* DH11S *mntH* were transformed with pGBM6-pH constructed as follows: pHluorin (GenBank accession AF058694) was amplified by PCR from pGFP_{ratiometric} (32) using the primers Gratio2F (5'-AAAAGCATGCGTAAAGGAGAAGAAGACT) and Gratio2R (5'-TATTAAGCTTTTATTTGTATAGTTCATCC). Cycling parameters with the *Pfu* polymerase (Stratagene, Cedar Creek, TX), were 5 cycles of 96 °C for 45s, 55°C for 45s, 72 °C for 2 min 30, followed by 25 cycle of 96 °C for 45s, 62°C for 45s, 72 °C for 2 min 30. This fragment was cloned into pJBA111 (34) using *Sph* I and *Hind* III restrictions sites. Subsequently, the cassette P_{A1/04/03}-RBSII-gfp ratio- T0-T1, was excised using *Not* I, blunt-ended using T4 DNA polymerase, and ligated to pGBM6 (ATCC 87502)

previously digested by *Sma*I.

For transport activity determination, TSS-competent cells *E. coli* DH11S *mntH* (pGBM6-pH) were transformed with pBAD plasmids encoding different MntH variants. Clones were grown in the presence of appropriate antibiotics at 37°C in LB broth which was supplemented with 0.06% L-arabinose 1 h before fluorescence measurement to allow the expression of MntH proteins. Cells were then harvested by centrifugation, washed once in citrate-potassium phosphate buffer (50mM, pH 5.0) and finally resuspended in the same buffer to an O.D₆₀₀ of 0.2.

Fluorescence was measured on a Cary Eclipse fluorescence spectrophotometer (VARIAN, Inc.), pHluorin fluorescence was excited by a xenon lamp at 410nm and 470nm respectively, the emission was detected at 520nm. Fluorescence excitation ratio R(410/470nm) was continuously monitored for 500 seconds and transformed into pH values (or intracellular concentration of protons) according to the calibration curve. Calibration was performed by use of carbonyl cyanide-*m*-chlorophenylhydrazone (CCCP) as previously described (32). Cadmium (10μM or 100μM) or magnesium (100μM) were added to the samples 80 seconds after beginning of the acquisition.

Results

A consensus topology prediction (35) was generated to verify whether prokaryotic MntH could assume a transmembrane organization similar to eukaryotic Nramp. Algorithms using different approaches that performed best in test predictions (36-38) were employed. Memsat is a dynamic program using statistical amino acid distribution derived from well-characterized membrane proteins, and five structural states (inside loop, inside helix end, helix middle, outside helix end, outside loop), to recognise possible topology models and to produce a list of all possible topologies and number of TMDs (39). TMHMM uses Hidden Markov simulations that implement a model defined by seven structural classes of residues constituting a membrane protein (40). To increase the accuracy of our prediction we performed a PHDhtm transmembrane topology prediction using the evolutionary information (41) contained in a representative set of MntH A proteins from Gram positive and negative species and *Deinococcus* group. In addition, *E. coli* MntH A (EcoliA) hydropathy profile was inspected using the amino acid hydrophobicity scale of Kyte and Doolittle and a sliding window of 19 residues which discriminates best potential transmembrane helices (42). The combined results of these approaches were very consistent with those obtained with eukaryotic Nramp sequences (7), and yielded a predicted topology with the N-terminus cytoplasmic, followed by 11 TMD and the C-terminus extracytoplasmic (Fig. 1). [Proposed location for Fig. 1]

To validate this prediction experimentally we used different enzymes as topological reporters of the periplasmic and cytoplasmic sides of the membrane, the secreted form of the β -lactamase (Blam) and the chloramphenicol acetyl transferase (Cat), respectively. Blam has been used as an alternative to the alkaline phosphatase (22) to determine the topology of different membrane proteins (27; 43); this reporter is active in the periplasm and therefore confers resistance to ampicillin when fused to or inserted in a region accessible to the periplasmic space (21). The cytoplasmic reporter Cat has been used in several studies to complement data obtained with either Blam or PhoA fusions (44-46); it confers resistance to

chloramphenicol when fused on the cytoplasmic side of the membrane.

Blam has been reported to be active as in-frame insertion which can generate more useful topological information concerning a protein that is functional; on the other hand random-based Blam fusions enable to produce data without any assumption but can also produce false positive results (25; 21; 47). Both methods were used, limiting the analysis to fusions that conferred strong resistance to ampicillin i.e. demonstrating periplasmic exposure. The location of Blam insertions was chosen based on the topological prediction (Fig. 1) and targeted to the N-terminal portion of predicted extracytoplasmic loops showing relatively low level sequence conservation.

[Proposed location for Table 3]

The results obtained with Blam are presented in Table 3 and Fig. 2 A-B. All the fusions conferring ampicillin resistance above 45 $\mu\text{g/mL}$ were determined to be in frame; the position of the preceding residue is indicated in Table 3. Most fusions and insertions were detected in membrane preparations by Western blot analysis using an α -Blam antibody. Some constructs resulted in significant amount of low molecular weight form of immuno-reactive Blam (similar to secreted Blam), suggesting proteolysis after MntH-Blam fusions were inserted in the membrane (F38, G49, V350, M377, G378, L404, Fig. 2A and F263, L412, Fig. 2B). In one case, low amount of fusion protein correlated with low level Blam activity (A165), in accordance with previous studies showing that the level of ampicillin resistance does not correlate with the level of expression of Blam fusion proteins (43; 48; 49). Hence the two fusions that conferred lower levels of ampicillin resistance (A165 and T371) differed significantly in their expression levels while both were detected predominantly as Blam fusions with negligible proteolysis. This suggested the two fusions points were located within TMD resulting in sub-optimal exposure of the Blam in the periplasm and consequently in relatively lower level of Amp resistance. Three independent in-frame fusions indicated a result in contradiction with the prediction that placed the loop between TMD 10-11 on

the periplasmic vs. cytoplasmic side of the membrane (ambiguous location, Table 3). [Proposed location for Fig. 2]

The three targeted Blam insertions obtained all conferred strong Amp resistance; two of them retained some functional activity as their expression conferred sensitivity to Fe^{2+} (supplemental data 1) and suppressed *hfIB1* (Ts) defect (F263, L412; Table 3). They differed in their level of toxicity and of functional activity which seemed proportional to the expression levels detected using the α -Blam antibody (Fig. 2B). The targeted insertion L179 was hardly detectable with α -Blam and no secreted Blam was detected either (Fig. 2B); however strong resistance to ampicillin (Table 3) implies expression of the hybrid protein probably at very low level. No functional activity was detected with L179 insertion; further excision of the Blam reporter -still resulting in a GS dipeptide insertion- did not restore EcoliA activity (supplemental data 1). Insertion before the stop codon of either the Blam polypeptide (L412) or a c-myc tag (L412-c-myc; Fig. 2C) had similar impact as both constructs exhibited close to wild type MntH activity (Table 3). The data indicate that the (fourth) loop 6/7 (F263) and the C-terminus (L412) are periplasmic, consistent with the topological prediction and most Blam fusions data.

To prove the predicted cytoplasmic exposure of the loop between TMD 10-11 (showing that M371, M377, G378 in Table 3 represent methodological artefacts) a targeted fusion of Cat after residue 387 (K387-Cat) was constructed; additional Cat fusion points in the loops 2/3, 6/7, 8/9 were used as positive controls (R91-Cat, G226-Cat, L324-Cat), based on the prediction (Fig. 1) and an insertion before the stop codon was used as negative control (L412-Cat). This C-terminal Cat insertion did not confer resistance to chloramphenicol above the background (data not shown); however, L412-Cat was strongly expressed and toxic for cells expressing it which limited the possible significance of the result. Addition of an N-terminal His-tag was found to strongly reduce MntH-c-myc protein expression level; (Fig. 2C) and to diminish the toxicity of L412-Cat construct. Although this hybrid protein (His-L412-Cat) was barely detectable in

Western analysis of membrane preparations using an antiserum against Cat (Fig. 2D), it retained metal uptake activity and did not confer resistance to chloramphenicol (Table 4). These results confirmed that EcoliA C-terminus is periplasmic.

[Proposed location for Table 4]. Cat fusions after residues predicted to be cytoplasmic conferred accordingly strong chloramphenicol resistance (R91-, G226-, L324- and K387-Cat, Table 4). The shorter constructs conferred higher levels of resistance. Antibody reactivity against Cat was detected in membrane preparations of all the clones expressing fusions (Fig. 2D). Proteolysis of these fusions was significant, in particular for R91-Cat; this was confirmed by metabolic labeling and observation of relatively low level ³⁵S-Met incorporation in a polypeptide of expected size, and decreased protein stability after chasing with cold methionine (supplemental data 2). The deduced cytoplasmic exposure of K387 is consistent with the prediction and all other results except those of the Blam fusions T371, M377 and G378 (Table 3). These three residues precede the putative topological signal (K R v K₃₈₇; Fig. 1) believed to anchor the cytoplasmic end of TMD11 based on the 'positive inside' rule (47; 48) and on other Cat fusion data obtained with EcoliA (Table 4). Expression of L324- and K387-Cat did not confer increased sensitivity to Mn and Fe; however residual sensitivity to the toxic metals, cobalt and cadmium persisted. Expression of the Blam construct T371 (Table 3) conferred similar low level sensitivity to the heavy metals Co and Cd (data not shown), although in this construct, TMD10 is predicted to be inserted incorrectly in the membrane.

Disk assay of MntH-dependent metal sensitivity indicates alteration of metal transport capacity secondary to mutagenesis (Tables 3 and 4). However bacterial sensitivity to heavy metals such as Cd and Co may be overestimated in this assay due to their inherent cyto-toxicity (14). Therefore, to demonstrate the functional difference between MntH hybrid proteins (F263, L412, Table 3 and His-L412-Cat, Table 4) and a truncated form of MntH (fused to Cat) that still conferred sensitivity to Cd and Co (L324-Cat, Table 4),

we studied a mechanistic aspect of MntH transport. Proton-dependent transporters can use the proton electrochemical gradient to drive substrate transport against its concentration gradient. Electrophysiological measurements of eukaryotic Nramp 2 activity showed that intracellular acidification accompanied divalent metal intracellular accumulation, implying proton-metal co-transport (9). To demonstrate EcoliA MntH transport function we examined divalent metal-dependent intracellular acidification subsequent to MntH expression using a pH-dependent ratiometric green fluorescent protein (32; 33).

[Proposed location for Fig. 3]. The data presented in Fig. 3 show that expression of wild-type EcoliA MntH induced an intracellular acidification after addition of Cd^{2+} but not after addition of Mg^{2+} , indicating that a divalent metal substrate for MntH transport (1) is required for proton transport. MntH insertion mutants exhibited similar activity but with stronger dependence on the concentration of Cd^{2+} added (e.g. Fig. 3AB, L412 and F263, L412-c-myc and His-L412-c-myc, His-L-412-c-myc). In contrast, expression of the mutant L324-Cat did not modulate the intracellular pH (Fig. 3), indicating the growth sensitivity to Cd observed with this mutant is not likely due to proton-dependent transport. Expression of the construct containing the dipeptide GS inserted after L₁₇₉ showed little intracellular acidification (supplemental data 1). These results demonstrate that Blam insertions within MntH polypeptide can preserve transport function and validate the topology determination.

Discussion

To our knowledge, a topological analysis demonstrating the orientation relative to the membrane of both termini, and the total number of TMD, had not been previously reported for any member of Nramp family. In this work, we used functional reporters to study the TMD of Nramp proteins using the *E. coli* MntH A (EcoliA) ortholog as a model. The results show EcoliA is a type III integral transmembrane protein with the N-terminus cytoplasmic followed by 11 TMD and the C-terminus exposed on the periplasmic side of the membrane (Fig. 4). [Proposed location for Fig. 4].

The periplasmic location of the C-terminus of *E. coli* MntH A was demonstrated with the C-terminal insertions of the Blam and Cat reporters that were both functional in conferring bacterial sensitivity to Fe^{2+} and Mn^{2+} and preserving divalent metal-dependent H^+ -transport. Indeed, only the Blam insertion conferred resistance to ampicillin whereas L412-Cat did not confer chloramphenicol resistance. The intracytoplasmic location of EcoliA N-terminus was revealed by two random fusion constructs (F38 and G49) that conferred strong resistance to ampicillin; F38 appeared predominantly in a fused form, indicating that Blam membrane insertion was due to TMD1. As the first TMD is preceded by the sequence R K M R₁₈ (that resembles the likely topological signal preceding TMD11, K R V K₃₈₇), these two fusions demonstrate that the first and little hydrophobic segment of the protein constitutes TMD1. Although P and G are generally viewed as “helix-breaking” residues, their structure-based distribution in known transmembrane helices suggest a likely important structural role (50) further underscored by their conservation in Nramp TMDs, e.g. TMD1, 6, 9 and 10.

Periplasmic exposure of the predicted extraloop 6/7 was also established with the targeted Blam insertion F263 that resulted in a bi-functional hybrid protein. This result added support to that of the Blam fusion obtained within predicted TMD7 (F248) which suggested that the carboxy-end of this transmembrane

segment protruded in the periplasm. Both constructs F248 and F263 were expressed at significant levels with apparently low level of proteolysis, indicating a relative stability within the membrane. The bi-functional MntH Blam insertion (F263) provided strong support to believe that the model presented in Fig. 4 is correct, because it preserved MntH transport of both divalent metals and protons which are hallmarks of Nramp function. Detection of both Cd^{2+} -dependent proton transport and increased metal sensitivity (Fe, Co and Cd) in cells expressing the Blam insertions F263 and L412 imply the transmembrane topology deduced from these results describes a functional MntH A protein.

The actual number of 11 TMDs constituting EcoliA was derived from the analysis of the 16 fusions or insertions constructs that were obtained in five out of six extraloops or end, and in four out of six intraloops or end. Altogether, the Blam constructs revealed extraloops between TMD 1/2, 5/6, 7/8, 9/10 and the Cat constructs indicated cytoplasmic exposure of the loops 2/3, 6/7, 8/9 and 10/11. The two intraloops for which no data were produced (N-terminus, intraloop 4/5) both display a significant net positive charge (+5, +3), whereas the extraloop 3/4 has a net positive charge of +1. The amino acid composition bias of these segments thus supports their deduced localization (51).

The TMD for which no fusion was produced (TMD4) is one of the most hydrophobic in *E. coli* MntH A protein. TMD4 is devoid of charged residues in MntH B and most MntH A proteins, as well as in plant type I sequences, albeit other eukaryotic sequences and most MntH C sequences possess a conserved negative charge (D or E). Nevertheless rat Nramp 2 TMD4 incorporated as an α -helix into micelles made of phospholipids and detergent (25), suggesting the negative charge when present must be located on a side of the transmembrane helix that contributes to inter-helices contacts (50). However, natural or induced mutation introducing a negatively charged residue in the N-terminal portion of TMD4 suffice to knockout both Nramp1 (2) and Nramp2 function (52). In any case, both the strong hydrophobicity of EcoliA TMD4 (including 14 strongly hydrophobic residues L, I, V, F and five residues frequently found in

lipid exposed face of transmembrane helices, A and G) and its strong potential for internal helix packing (increased with residues S₁₂₇, T₁₃₅, T₁₃₉) together imply, consistent with the overall topology prediction, the existence of EcoliA TMD4.

The targeted Blam insertion in the extraloop 5/6 confirmed the result obtained with the targeted Blam fusion within TMD5 (A₁₆₅). This fusion was expressed at low level and conferred low level of ampicillin resistance; similarly, Blam insertion L₁₇₉ was expressed at very low level under the conditions tested and lost MntH A functional activity. Insertion of the dipeptide GS (vs. Blam) after the residue L₁₇₉ was also associated with limited EcoliA activity (both metal sensitivity and Cd-dependent proton transport, data not shown). These results are consistent with low level expression of L₁₇₉GS construct and indicate that the carboxy end of TMD5 is sensitive to mutation. The sensitivity of TMD5 carboxy end to insertion mutagenesis would contrast with the results obtained with TMD7, in which similar mutations (F₂₄₈, F₂₆₃) resulted in rather stable polypeptides. L₁₇₉ is located two residues after a P residue that is conserved among MntH A proteins; the sixth residue after A₁₆₅ is either a polar or a charged residue, predicted to be embedded within the lipid bilayer. Hence, both the amphiphilic character of TMD5 and the presence of potential conserved structural determinants may explain the relative sensitivity of this region to insertional mutagenesis.

All the fusions and insertions that were obtained supported the consensus prediction consistently except three random fusions that produced an artefact (T₃₇₁, M₃₇₇ and G₃₇₈). This interpretation is supported by the observation that these fusion points are localized 5 to 12 residues upstream of a possible topogenic signal that would anchor TMD11 N-terminal end to the cytoplasmic side of the membrane. Therefore all the results obtained strongly support the model derived from the consensus prediction (Fig. 1 & 4).

The 11 TMD topology (N_{in} , C_{out}) of EcoliA resembles more that of yeast proteins vs. animal or plant

homologs which have 12 TMD. Nevertheless, the model established in this study of *E. coli* MntH A protein is consistent with the predictions and data that were generated studying eukaryotic homologs. The hydrophobic core that is most conserved in Nramp proteins encompasses TMD 1-10 and implies strong structural and functional constraints to maintain both selective divalent metal transport and proton transport. The findings are consistent with the proposition that eukaryotic Nramp transporters derived from an ancestral *mntH* A gene.

Despite relatively low level sequence relationship with each other, MntH proteins from groups A and B display very similar hydropathy profiles suggesting MntH B proteins could adopt a transmembrane topology similar to EcolIA. Also, based on their similarity to eukaryotic Nramp, MntH C proteins should accordingly display a topology similar to the model presented, including some MntH C β 1 proteins found in lactic Bacteria predicted to contain 12 TMD and the C-terminus extracytoplasmic. In conclusion, the model established in this study will be useful for structural and functional studies of both bacterial MntH and eukaryotic Nramp transporters of divalent metals and proton.

Acknowledgments

This work was supported by a research grant from the Canadian Institutes of Health Research (MOP-78014-MI). RC was supported by a short term NATO Science Fellowship (CZ 3/2003). MFMC is a Scholar of the Fond pour la Recherche en Santé du Québec. We thank past and present lab members for their support, as well as Drs MA. Mandrand-Berthelot, H. Ingmer, C. Dupont and F. Shareck who supplied reagents or shared equipment that were used in this study.

Reference List

1. Makui, H., Roig, E., Cole, S. T., Helmann, J. D., Gros, P., and Cellier, M. F. (2000) *Mol.Microbiol.* **35**, 1065-1078
2. Vidal, S. M., Malo, D., Vogan, K., Skamene, E., and Gros, P. (1993) *Cell* **73**, 469-485
3. Gruenheid, S., Cellier, M., Vidal, S., and Gros, P. (1995) *Genomics* **25**, 514-525
4. Belouchi, A., Cellier, M., Kwan, T., Saini, H. S., Leroux, G., and Gros, P. (1995) *Plant Mol.Biol.* **29**, 1181-1196
5. West, A. H., Clark, D. J., Martin, J., Neupert, W., Hartl, F. U., and Horwich, A. L. (1992) *J.Biol.Chem.* **267**, 24625-24633
6. Cellier, M., Prive, G., Belouchi, A., Kwan, T., Rodrigues, V., Chia, W., and Gros, P. (1995) *Proc.Natl.Acad.Sci.U.S.A* **92**, 10089-10093
7. Cellier, M., Belouchi, A., and Gros, P. (1996) *Trends Genet.* **12**, 201-204
8. Supek, F., Supekova, L., Nelson, H., and Nelson, N. (1996) *Proc.Natl.Acad.Sci.U.S.A* **93**, 5105-5110
9. Gunshin, H., Mackenzie, B., Berger, U. V., Gunshin, Y., Romero, M. F., Boron, W. F., Nussberger, S., Gollan, J. L., and Hediger, M. A. (1997) *Nature* **388**, 482-488
10. Forbes, J. R. and Gros, P. (2001) *Trends Microbiol.* **9**, 397-403
11. Portnoy, M. E., Liu, X. F., and Culotta, V. C. (2000) *Mol.Cell Biol.* **20**, 7893-7902
12. Muckenthaler, M., Roy, C. N., Custodio, A. O., Minana, B., deGraaf, J., Montross, L. K., Andrews, N. C., and Hentze, M. W. (2003) *Nat.Genet.* **34**, 102-107
13. Cellier, M. F., Bergevin, I., Boyer, E., and Richer, E. (2001) *Trends Genet.* **17**, 365-370
14. Richer, E., Courville, P., Bergevin, I., and Cellier, M. (2003) *J.Mol.Evol.* **57**, 363-376
15. Que, Q. and Helmann, J. D. (2000) *Mol.Microbiol.* **35**, 1454-1468
16. Agranoff, D., Monahan, I. M., Mangan, J. A., Butcher, P. D., and Krishna, S. (1999) *J.Exp.Med.* **190**, 717-724
17. Reeve, I., Hummel, D., Nelson, N., Voss, J., and Hummel, D. (2002) *Proc.Natl.Acad.Sci.U.S.A* **99**, 8608-8613
18. Boyer, E., Bergevin, I., Malo, D., Gros, P., and Cellier, M. F. (2002) *Infect.Immun.* **70**, 6032-6042

19. Horsburgh, M. J., Wharton, S. J., Cox, A. G., Ingham, E., Peacock, S., and Foster, S. J. (2002) *Mol.Microbiol.* **44**, 1269-1286
20. Saier, M. (2001) Families of Transporters: A Phylogenetic Overview. In Winkelmann, G., editor. *Microbial Transport Systems*, Wiley-VCH, Weinheim
21. van Geest, M. and Lolkema, J. S. (2000) *Microbiol.Mol.Biol.Rev.* **64**, 13-33
22. Drew, D., Sjostrand, D., Nilsson, J., Urbig, T., Chin, C. N., de Gier, J. W., and von Heijne, G. (2002) *Proc.Natl.Acad.Sci.U.S.A* **99**, 2690-2695
23. Vidal, S. M., Pinner, E., Lepage, P., Gauthier, S., and Gros, P. (1996) *J.Immunol.* **157**, 3559-3568
24. Picard, V., Govoni, G., Jabado, N., and Gros, P. (2000) *J.Biol.Chem.* **275**, 35738-35745
25. Traxler, B., Boyd, D., and Beckwith, J. (1993) *J.Membr.Biol.* **132**, 1-11
26. Kuhn, D. E., Lafuse, W. P., and Zwilling, B. S. (2001) *J.Leukoc.Biol.* **69**, 43-49
27. Broome-Smith, J. K. and Spratt, B. G. (1986) *Gene* **49**, 341-349
28. Zhang, Y. B. and Broome-Smith, J. K. (1990) *Gene* **96**, 51-57
29. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene* **77**, 51-59
30. (2002) *Current protocols in protein science*, John Wiley and Sons, Inc.,
31. Ward, A., Sanderson, N. M., O'Reilly, J., Rutherford, N. G., Poolman, B., and Henderson, P. J. F. (2000) The amplified expression, identification, purification, assay, and properties of hexahistidine-tagged bacterial membrane transport proteins. In Baldwin, S., editor. *Membrane Transport*, Oxford university press, New York
32. Olsen, K. N., Budde, B. B., Siegumfeldt, H., Rechinger, K. B., Jakobsen, M., and Ingmer, H. (2002) *Appl.Environ.Microbiol.* **68**, 4145-4147
33. Miesenbock, G., De Angelis, D. A., and Rothman, J. E. (1998) *Nature* **394**, 192-195
34. Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M., and Molin, S. (1998) *Appl.Environ.Microbiol.* **64**, 2240-2246
35. Nilsson, J., Persson, B., and von Heijne, G. (2000) *FEBS Lett.* **486**, 267-269
36. Moller, S., Croning, M. D., and Apweiler, R. (2001) *Bioinformatics.* **17**, 646-653
37. Lao, D. M., Okuno, T., and Shimizu, T. (2002) *In Silico.Biol.* **2**, 485-494

38. Melen, K., Krogh, A., and von Heijne, G. (2003) *J.Mol.Biol.* **327**, 735-744
39. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1994) *Biochemistry* **33**, 3038-3049
40. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) *J.Mol.Biol.* **19;305**, 567-580
41. Rost, B., Fariselli, P., and Casadio, R. (1996) *Protein Sci.* **5**, 1704-1718
42. Kyte, J. and Doolittle, R. F. (1982) *J.Mol.Biol.* **157**, 105-132
43. Benoit, S., Abaibou, H., and Mandrand-Berthelot, M. A. (1998) *J.Bacteriol.* **180**, 6625-6634
44. Zelazny, A. and Bibi, E. (1996) *Biochemistry* **20;35**, 10872-10878
45. Adler, J. and Bibi, E. (2002) *J.Bacteriol.* **184**, 3313-3320
46. Hirata, T., Fujihira, E., Kimura-Someya, T., and Yamaguchi, A. (1998) *J.Biochem.(Tokyo)* **124**, 1206-1211
47. Ott, C. M. and Lingappa, V. R. (2002) *J.Cell Sci.* **115**, 2003-2009
48. Gunn, F. J., Tate, C. G., Sansom, C. E., and Henderson, P. J. (1995) *Mol.Microbiol.* **15**, 771-783
49. Tate, C. G. and Henderson, P. J. (1993) *J.Biol.Chem.* **268**, 26850-26857
50. Ulmschneider, M. B. and Sansom, M. S. (2001) *Biochim.Biophys.Acta* **1512**, 1-14
51. Juretic, D., Zoranic, L., and Zucic, D. (2002) *J.Chem.Inf.Comput.Sci.* **42**, 620-632
52. Fleming, M. D., Romano, M. A., Su, M. A., Garrick, L. M., Garrick, M. D., and Andrews, N. C. (1998) *Proc.Natl.Acad.Sci.* **95**, 1148-1153
53. Herman, C., Lecat, S., D'Ari, R., and Bouloc, P. (1995) *Mol.Microbiol.* **18**, 247-255

Figure 1. Sequence analysis of selected MntH homologs from group A.

The open boxes above the alignment indicate putative transmembrane segments, and their orientation relative to the membrane as predicted by different algorithms. *Escherichia coli* (EcoliA), *Erwinia chrysantemi* (EchryA), *Bacillus subtilis* (BsubtA), *Mycobacterium leprae* (MleprA) and *Deinococcus radiodurans* (DradiA).

Figure 2. Detection of MntH mutant proteins by Western blot analysis. (A) *E. coli* MntH-Blam fusions. (B) *E. coli* MntH-Blam targeted insertions. (C) Epitope-tagged MntH (L412-c-myc, His-L412-c-myc). (D) *E. coli* MntH-Cat targeted fusions. Equal protein amounts (120 µg per lane, unless indicated otherwise) of membrane extracts prepared as described in Material and Methods were loaded for SDS-PAGE on 12% polyacrylamide gels.

Figure 3. EcoliA MntH-dependent intracellular accumulation of protons in presence of cadmium (Cd²⁺). The intracellular accumulation of protons (i.e. decrease of intracellular pH) after addition of Cd²⁺ was determined by a fluorescence assay using ratiometric GFP as described in Material and Methods. According to the calibration curve, the Cd-dependent proton transport by wild type and mutant forms of EcoliA MntH was depicted as an increase of intracellular concentration of protons. The addition of magnesium (Mg²⁺) to cells expressing Ecoli MntH was used as negative control as well as the addition of 100µM Cd to cells without MntH (vector alone). The arrows indicate the addition of (A) 10µM CdCl₂ and (B) 100µM CdCl₂ (or MgCl₂).

Figure 4. EcoliA MntH consensus transmembrane topology. Residues upstream of Blam fusion point that were obtained by random fusion and targeted fusion of Cat are surrounded by circles and squares, respectively. Sites of targeted Blam insertions are indicated by inverted triangles. The last residue (L412) is indicated by a star symbolizing either Blam, c-myc or Cat targeted insertions. Numbers correspond to the residue positions. CTM indicates a predicted location of the conserved transport motif.

Table 1: sequence of the oligonucleotide primers used in this study.

Primer	Sequence ^{a,b}	Restriction site
NcoI25	5' TCACCATGGCGAACTATCGC	<i>Nco</i> I
XbaIR	5' AGGTCGACTCTAGAGGA	<i>Xba</i> I
BlaM179	5' AACCTGGGATCCGCGCAGCTG	<i>Bam</i> H I
BlaM263	5' ACTTTGGATCCTCCGGTCATACT	<i>Bam</i> H I
Cat91 Δ	5' ACTACAGATCTACGCGGATAGT	<i>Bgl</i> II
Cat226	5' ATGGCGGATCCGGTTCGCGTC	<i>Bam</i> H I
Cat373	5' TCAACTTGCTGAATCCGTCG	<i>Bam</i> H I
Cat226Δ	5' CGCGTCTAGAATCCAAACTGC	<i>Xba</i> I
Cat373 Δ	5' AACTTCTAGAATCCAAACTGCA	<i>Xba</i> I
Cat324Δ	5' GCACAGATCTCAGCGGGATAT	<i>Bgl</i> II
Cat387 Δ	5' CA GCC <u>AGA TCT</u> TTT TAC GCG TTT	<i>Bgl</i> II
PCTAG1	5' CCACCAAATGGGATGTGGCTATC	<i>Xcm</i> I
PCTAG2	5' ATTGGATCCCAATCCCAGCGCCGTCCCC	<i>Bam</i> H I
PCTAG3a	5' CGCTGGGATCCTAGTTGAATGAGCGTC	<i>Bam</i> H I
PCTAG4a	5' GCCCGAGGCATAGACTGTACAAA	<i>Bsr</i> G I
FT1	5' <u>ACTCTAGATCAGATCCTCTTCTGAGATGAGTTTTTGTTCC</u> CAATCCCAGCGCCGTCCCCA	<i>Xba</i> I
FT2	5' TTGCCGCGGCTTACATTGTCGAGT	<i>Sac</i> II
Bla1	5' CTCGTGCACCCA ACTGA	
BamBla1	5' GTTGGATCCCACCCAGAAACGCTGGTGAAA	<i>Bam</i> H I
BamBla2	5' TGAGGATCCCAATGCTTAATCAGTGAGGCAC	<i>Bam</i> H I
CatF	5' GCTAAGGGATCCAGCCTGCAGAAAAAATCACTG	<i>Bam</i> H I
CatR	5' ACTGGATCCAAACTGCAGCGCCAGCGCCCCGCCCTGCC	<i>Bam</i> H I

^a The underline sequence represents the localisation of the restriction site

^b The italicized sequence represent the c-myc tag epitope

Table 2: Bacterial strains and plasmids used in this study.

Strains and plasmids	Genotype or other relevant characteristic	Reference source
DH11S	F ⁻ <i>mcrA</i> Δ(<i>mrr hsdRMS mcrBC</i>) Δ(<i>lac-proAB</i>) Δ(<i>recA1398</i>) <i>deoR supE rpsL srl thi/F' proAB⁺ lacI^qZ</i> ΔM15	Gibco BRL.
SL93	<i>lac-3350, galK2, galT22, rpsL179, zgj-203::Tn10 hflB1</i> (Ts)	(53)
SL93 <i>mntH</i>		(1)
pBADNX1.1 ¹	Amp ^r and induction with L-arabinose	(1)
pBADNX7.2 ²	Gm ^r and induction with L-arabinose	This study
pBADGmΔBamHI	Gm ^r and not inducible	This study
pYZ4NX1.1	Kan ^r and induction with IPTG	(27)
pYZ5	Tc ^r Blam fusion vector	(27)

¹ This plasmid was used for most cloning experiments

² This plasmid was used to clone the Blam insertions

Table 3. Characterization of EcoliA Blam constructs

Blam Fusion Point	Methods ¹	Amp ^R (µg/mL)	Sensitivity to Mn(18)	Suppress <i>HflB1</i> (Ts) defect	Location ²
F38	RF	>300	nt	nt	P
G49	RF	>300	nt	nt	P
A165	TF	45	nt	nt	TM
L179	TI	>300	-	-	P
F248	RF	>300	nt	nt	P
F263	TI	>300	-	+	P
V350	RF	>300	nt	nt	P
T371	RF	100	nt	-	A
M377	RF	200	nt	nt	A
G378	RF	200	nt	nt	A
L404	RF	>300	-	+/-	P
L412	TI	>300	+	++	P
EcoliA	na	na	++	+++	na
L412-c-myc	TI	na	+	++	nt

¹ RF, random fusion; TI, targeted insertion; TF, targeted fusion

² P, periplasmic; TM, transmembrane; A, ambiguous
na, not applicable; nt, not tested

Table 4. Analysis of Cat targeted fusions

Constructs	CmI ^R (µg/ml)	Disc assay of metal sensitivity (mm)				Suppress. <i>HflB1</i> (Ts) Defect
		FeSO ₄ 1M	MnCl ₂ 1M	CoCl ₂ 0,25M	CdCl ₂ 0,25M	
EcoliA	4	26 (+/-0,5)	14 (+/-0,5)	27 (+/-1)	37 (+/-1)	+++
Vector alone	4	14 (+/-0,5)	0	10 (+/-0,5)	21 (+/-0,5)	-
R91-Cat	16	nt	nt	nt	nt	nt
G226-Cat	16	nt	nt	nt	nt	nt
L324-Cat	12	15 (+/-0,2)	0	20 (+/-0,5)	28 (+/-1)	+/- ¹
K387-Cat	12	15 (+/-0,2)	0	20 (+/-0,5)	29 (+/-1)	+/- ¹
His-L412-Cat	4	23 (+/-0,5)	0	24 (+/-1)	32 (+/-0,5)	+/-
His-L412-c-myc	4	22 (+/-1)	0	31 (+/-1)	32 (+/-0,5)	+

nt, not tested

¹His-L324-Cat, His-K387-Cat

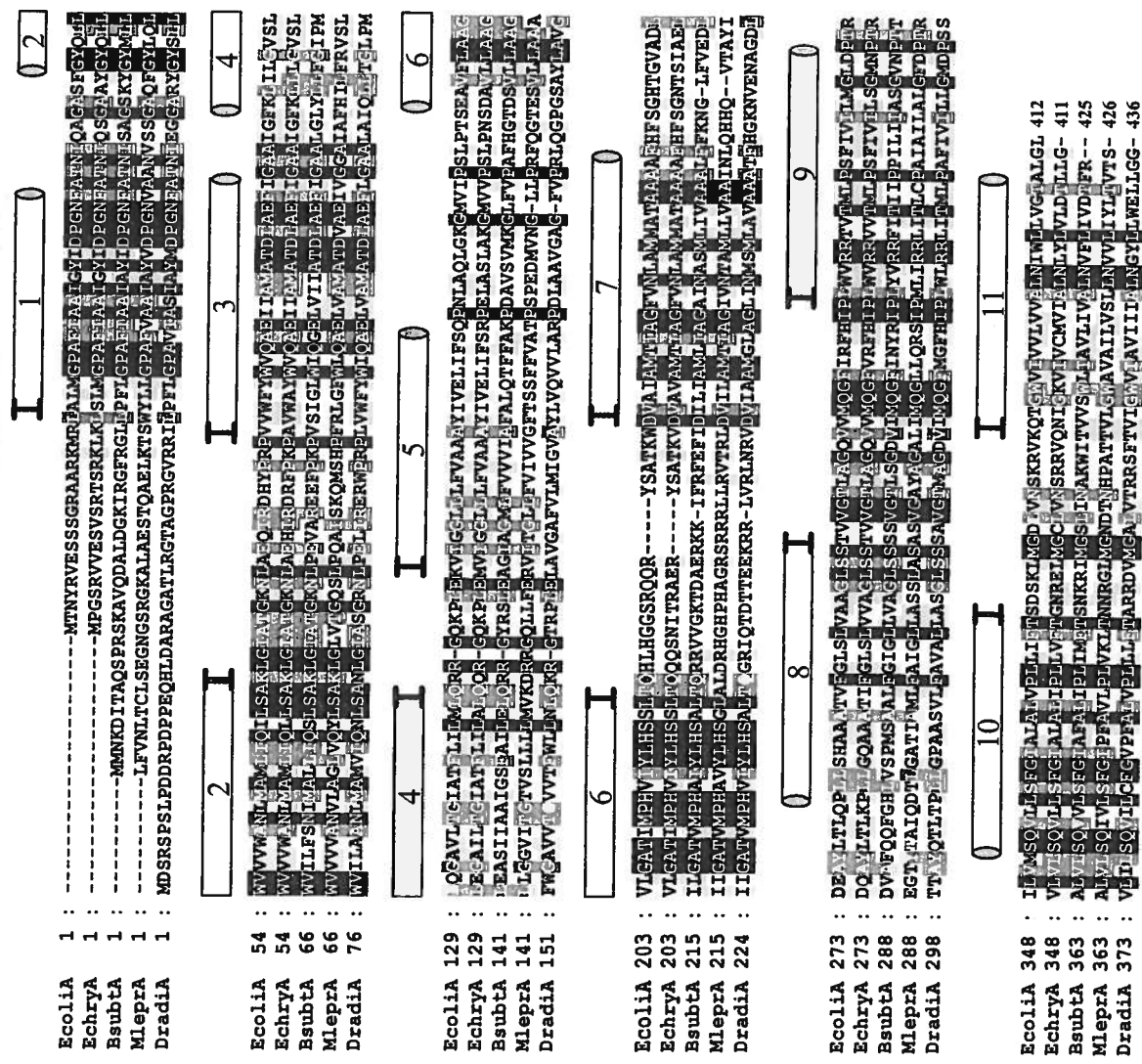


Figure 1
 I Inside. cytoplasmic
 O Outside. periplasmic

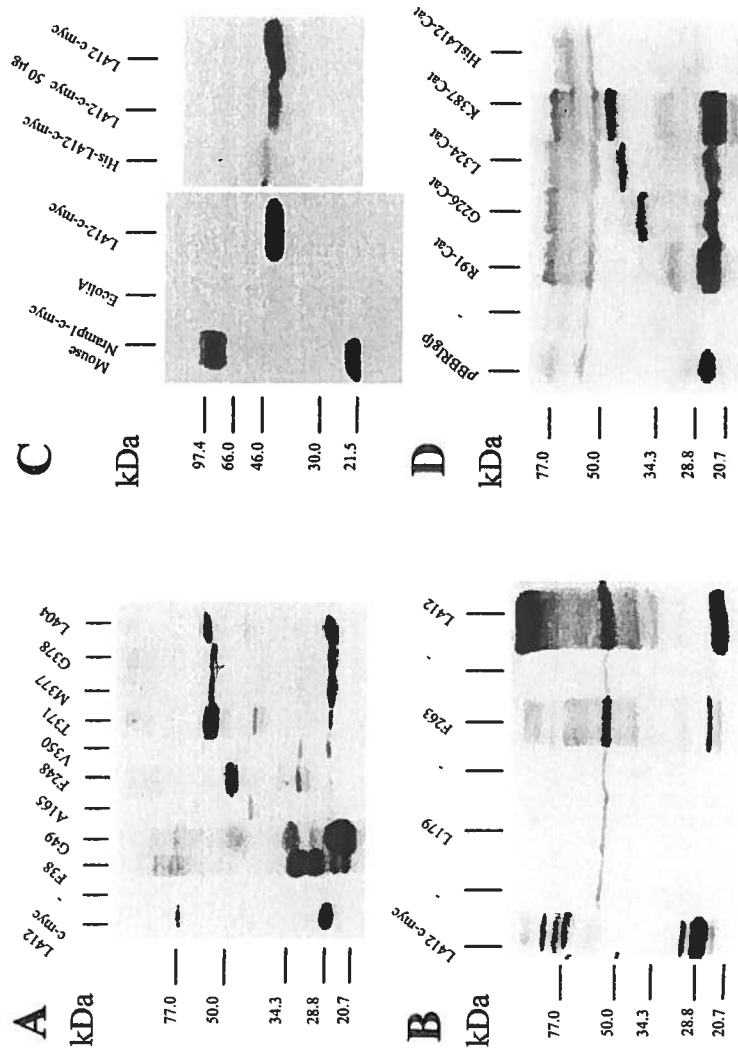


Figure 2

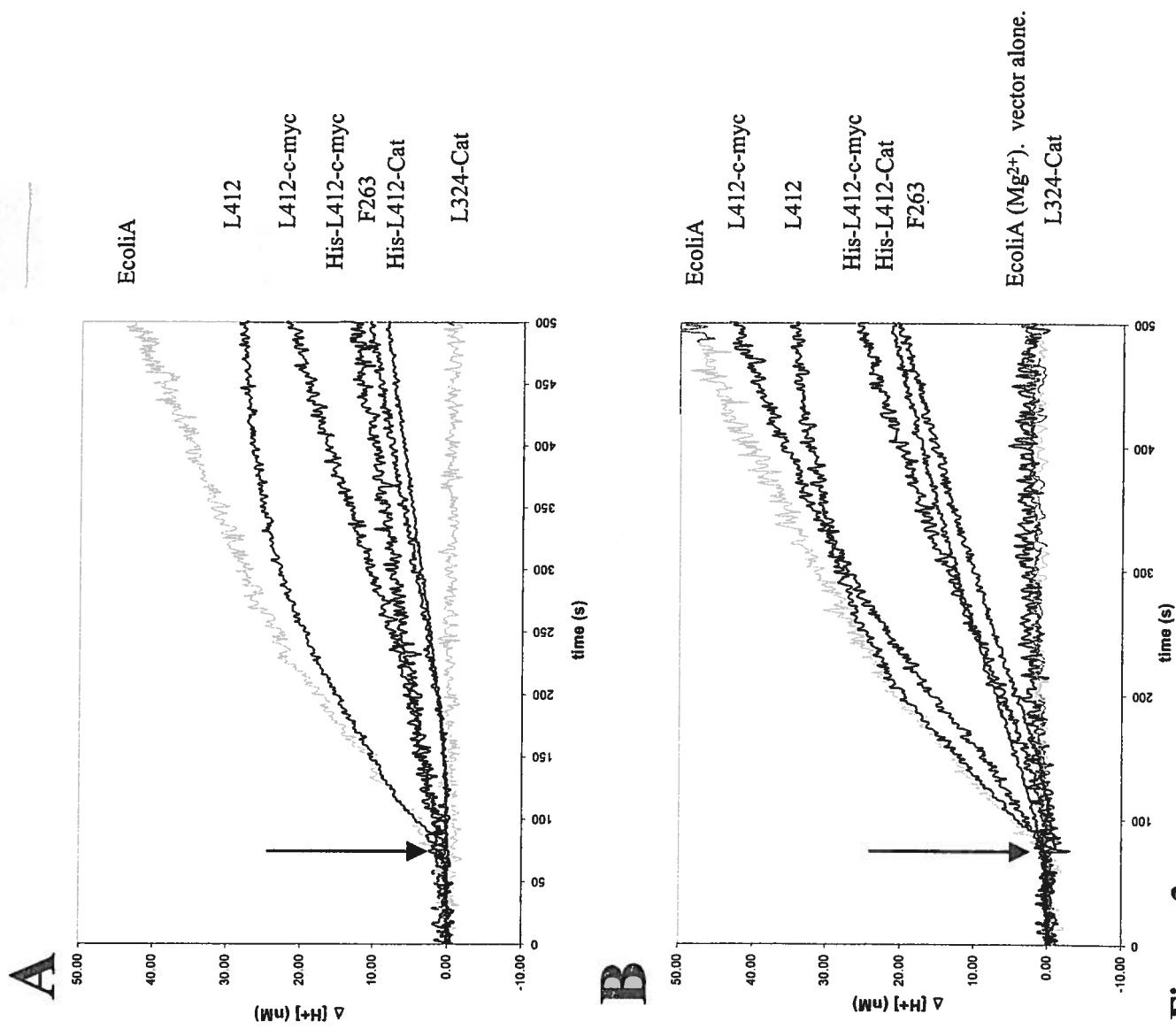


Figure 3

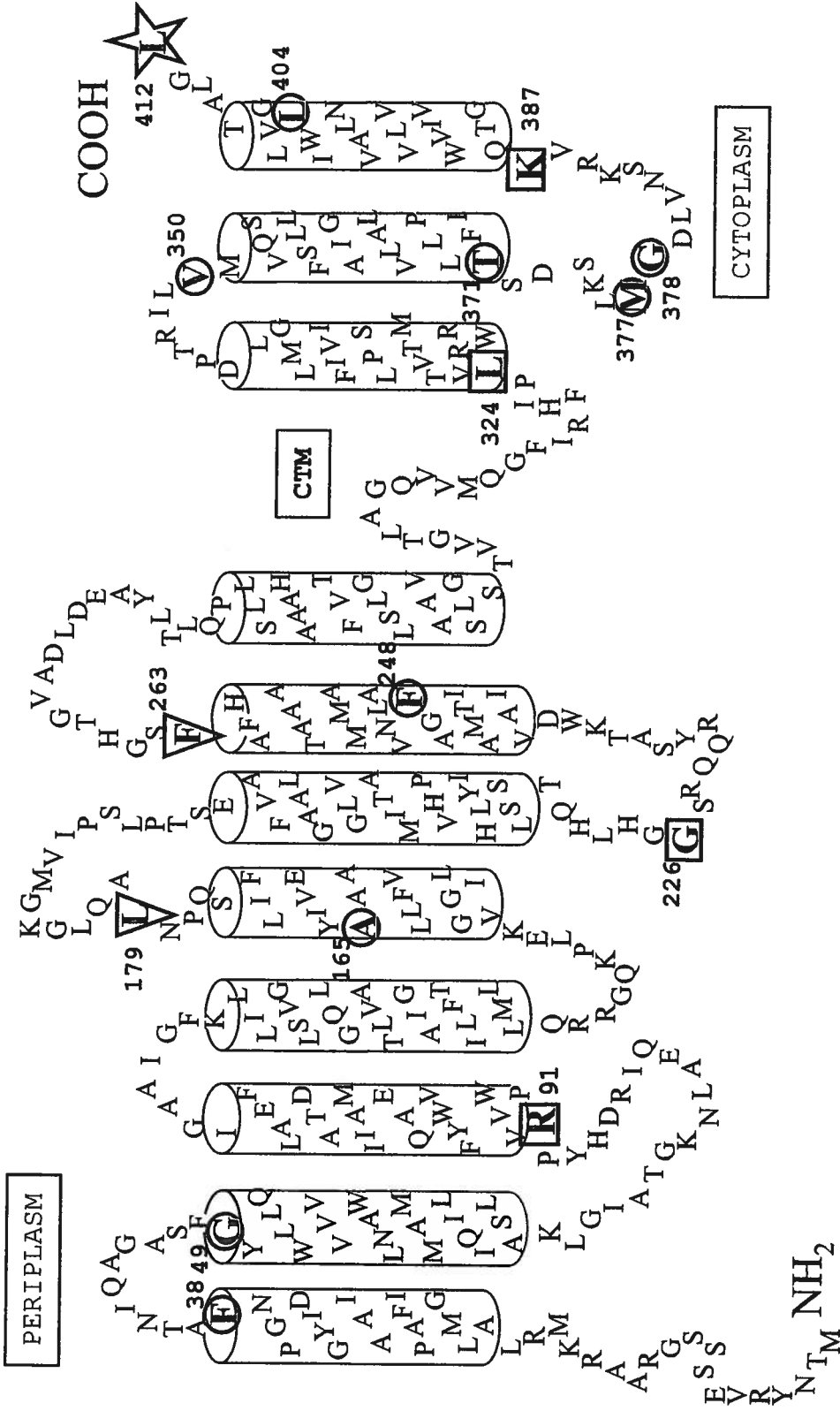


Figure 4

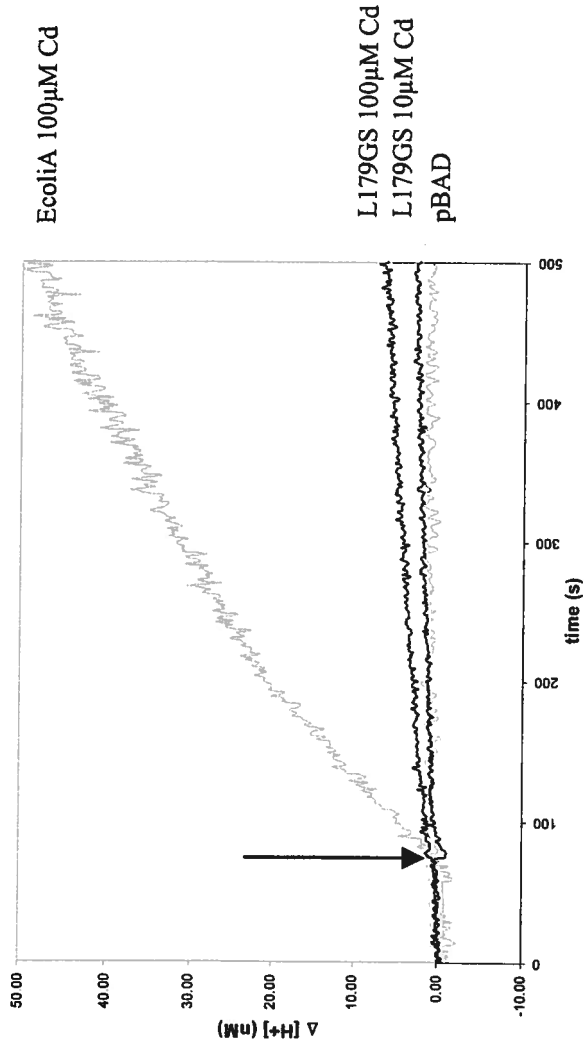
Supplemental data 1 Transport activity of EcoliA is impaired by insertion of the dipeptide GS after L₁₇₉

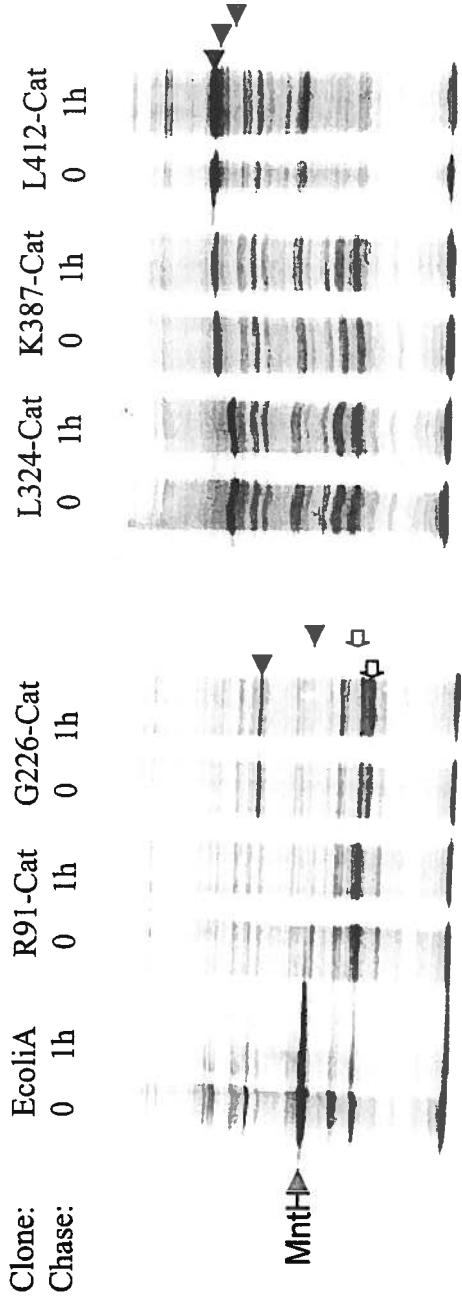
A

Clones	Disk assay of metal sensitivity (mm)					
	MnCl ₂ (1M)	FeSO ₄ (1M)	CoCl ₂ (0.25M)	CdCl ₂ (0.25M)	Complémentation 42°C	
EcoliA	14±0.5	25±0.5	27±1	37±1	+++	
pBAD	0	14±0.5	10±0.5	21±0.5	-	
L412	14	26±1	28±2	37±1	++	
F263	0	21±1	21.5±0.5	36.3±0.5	+	
L179	0	15±1	21±1	22±1	-	
L179 GS	0	22±1*	22.5±0.5*	31.5 ±0.5*	+	
L179GS- L412-c-Myc	n.t	16±1*	13.3 ±0.5*	29.3 ±0.5*	n.t	
His-DHFR-L412-c-Myc	0	14±0	18±0.5	27±1	n.t.	

*: weak growth inhibition

B





Supplemental Data 2: detection of R91-Cat. G226-Cat. L324-Cat. K387-Cat and L412-Cat by ³⁵S-Met metabolic labelling (for 30 min.. starting 15 min. after addition of 0.2 % arabinose as inducer of pBAD expression). Labelling was followed by a 1 min. or 1h. chase. by addition of 100-fold excess cold methionine before cell lysis in Laemmli buffer and SDS-PAGE analysis in 10 % gel. 3.10⁵ cpm/lane.

The expected protein sizes are indicated with black arrowheads; the efficiency of ³⁵S-Met incorporation was by decreasing order: EcoliA. L412-Cat. L324-Cat. K387-Cat. G226-Cat and R91-Cat.

The likely accumulation of proteolysis by-products corresponding to proteins expressed at low level (K387-Cat. G226-Cat. R91-Cat) is indicated by □ .