Analyse protéomique des enzymes impliquées dans la voie de dégradation du 2-EHN par *M. austroafricanum* IFP 2173.

Le terme protéome est apparu la première fois en 1995 (Wassinger *et al.*, 1995). Il s'agit de l'ensemble des protéines exprimées par le génome d'une cellule ou d'un tissu à un instant donné et dans des conditions données. C'est le produit final de l'expression d'un génome. Alors que le génome reste relativement constant d'une cellule à une autre, le transcriptome est considéré comme dynamique mais son analyse consiste à identifier, à un temps donné les séquences codantes du génome qui sont exprimées. Cependant, celui-ci ne prend pas en compte les fluctuations dues aux régulations post-traductionnelles des protéines. En revanche, le protéome est « dynamique » et varie de façon importante en fonction de différents facteurs comme l'environnement cellulaire, l'état physiologique des cellules, le stress et dans notre cas précis, la présence de produits toxiques comme le 2-EHN (Graves *et al.*, 2002). C'est le protéome de *M. austroafricanum* IFP 2173 que nous étudierons dans ce chapitre. Le génome de cette bactérie n'est pas connu, mais nous avons fait l'hypothèse qu'il était assez proche des génomes d'autres mycobactéries disponibles dans les bases de données pour identifier les protéines induites en présence de 2-EHN.

3.1 Stratégie d'analyse protéomique

Cette première partie présente la démarche expérimentale employée pour identifier les protéines de la souche IFP 2173 impliquées dans la dégradation du 2-EHN. Les outils d'analyse protéomique utilisés dans cette étude sont sommairement décrits car ceux ci se sont considérablement perfectionnés ces dernières années, notamment depuis l'avènement des techniques de spectrométrie de masse pour l'analyse et le séquençage peptidique.

3.1.1 Analyse des protéines induites sur 2-EHN

M. austroafricanum IFP 2173 utilise le 2-EHN comme unique source de carbone et doit donc synthétiser des enzymes spécifiques à son assimilation. L'objectif de l'analyse est de comparer le profil des protéines sur 2-EHN à celui des protéines présentes sur un autre substrat témoin pour cibler et identifier les enzymes responsables de la dégradation du 2-EHN.

3.1.1.1 Choix de la méthode d'analyse : électrophorèse 1D ou 2D ?

L'analyse de base des protéines consiste à les dénaturer en présence de SDS et à les faire migrer sur un gel d'électrophorèse en fonction de leur masse moléculaire. C'est une méthode simple à mettre en œuvre et reproductible. Elle peut être utilisée pour séparer des

protéines ayant des masses comprises entre 10 et 300 kDa. A partir des années 1975, l'introduction de la technique d'électrophorèse bidimensionnelle (2DE) par, entre autres, Klose (1975) et Scheele (1975) a permis de faire évoluer les outils d'analyse des protéines. L'électrophorèse bidimensionnelle consiste à séparer dans un premier temps les protéines en fonction de leur point isoélectrique, sur un gel d'acrylamide contenant un gradient de pH. C'est l'étape d'IEF, pour isoélectrofocalisation. Cette technique a été perfectionnée par les travaux de Görg *et al.* (2000) qui précisent l'importance de la température de l'isoélectrofocalisation et du choix des ampholytes pour la reproductibilité de la séparation des protéines. Dans un second temps, les protéines qui ont migré en fonction de leur point isoélectrique, sont séparées en fonction de leur masse moléculaire en conditions dénaturantes. Cette méthode d'analyse comporte de nombreux avantages. Elle permet d'analyser avec précision la charge et la masse des protéines. Elle permet, d'une part, de visualiser les modifications post-traductionnelles, les phosphorylations et les glycosylations. D'autre part elle permet de faire une carte du protéome. Il existe des bases de données de ces cartes disponibles sur Internet comme par exemple à l'adresse suivante : http://www.expasy.org/ch2d/2d-index.html. Néanmoins, certaines difficultés liées à la solubilisation et à la préparation de certaines protéines font qu'il est impossible de visualiser l'ensemble des protéines. Ainsi les protéines membranaires hydrophobes, les protéines extrêmophiles (hypersalines), les protéines de taille extrême (très petites < 10 kDa ou très grandes > 300 kDa) et les protéines présentant des valeurs de pI extrêmes (basique ou acide) sont rarement détectées sur gel 2D, à moins d'employer des conditions de séparation particulières. De plus, il peut y avoir des biais dans l'analyse liés à l'hyperabondance de certaines protéines, à l'agrégation de glycoprotéines et au mauvais transfert de protéines de la première dimension vers la deuxième.

J'ai dans un premier temps fait l'analyse des protéines sur des gels d'électrophorèse monodimensionnelle (1DE), puis bi-dimensionnelle (2DE). J'ai été confrontée à des problèmes de séparation des protéines membranaires en gels 2DE, comme illustré par la **figure 4.5**. Les protéines ne sont pas correctement séparées en fonction de leur point isoélectrique et ont mal migré dans la deuxième dimension. Bien que certains auteurs aient réussi à améliorer la qualité de ces gels en optimisant la méthode de préparation des protéines membranaires (Sabirova *et al.*, 2006), j'ai préféré analyser les protéines membranaires sur gel SDS-PAGE classique. J'ai améliorer la résolution de ces protéines sur gel SDS-PAGE en utilisant un système de tampon tris-tricine selon la méthode de Schägger *et al.* (2006), plutôt que le système tris-glycine (méthode classique de Laemmli *et al.* (1979)).

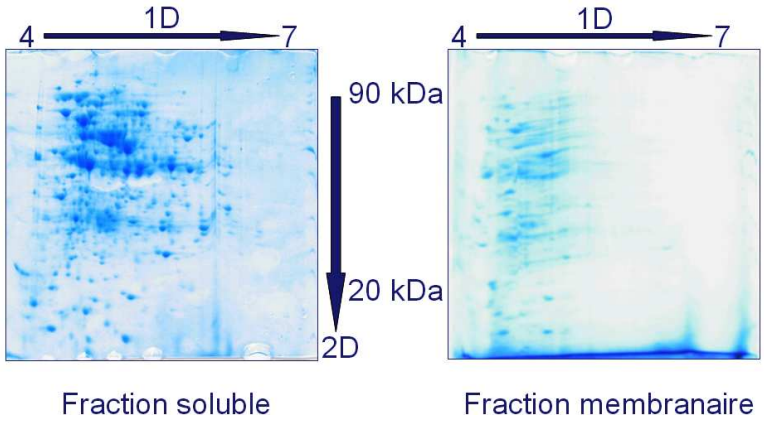

Figure 4.5 : Comparaison de gels d'électrophorèse bidimensionnelle de protéines solubles et membranaires de la souche IFP 2173 incubée sur 2-EHN

Une illustration de la résolution des protéines obtenue dans les deux conditions d'électrophorèse est représentée sur la **figure 4.6**.

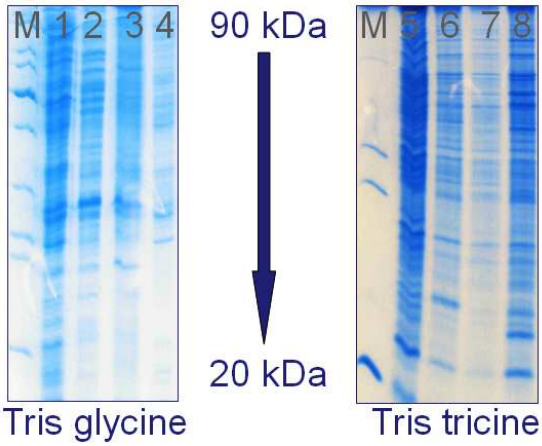

Figure 4.6 : Comparaison de la migration de protéines de la souche IFP 2173 sur gel Tris-glycine et sur gel Tris-tricine.
***M** : Marqueurs de poids moléculaire.*
***1** : Fraction protéique soluble sur Tween 80.*
***2** : Fraction protéique membranaire sur Tween 80.*
***3** : Fraction protéique soluble sur 2-EHN.*
***4** : Fraction protéique membranaire sur 2-EHN*
***5** : Fraction protéique membranaire sur Tween 80.*
***6** : Fraction protéique membranaire sur 2-EHN.*
***7**: Fraction protéique membranaire sur succinate.*
***8** : Fraction protéique membranaire sur 2-EHN*

Ainsi, après la mise au point des protocoles, la stratégie d'analyse adoptée fut la suivante :

- L'analyse des protéines solubles a été effectuée par électrophorèse bi-dimensionnelle.
- Les protéines membranaires ont été séparées sur gels Tris-tricine SDS-PAGE.

3.1.1.2 Choix du substrat témoin

L'objectif est de comparer les protéines induites sur 2-EHN par rapport à un substrat de croissance témoin. Les cultures sur 2-EHN, ont été ensemencées à une $D.O._{600}$ variant de 0,8 à 1 avec des bactéries ayant au préalable poussé sur un substrat de préculture comme l'acétate, le succinate, le Tween 80, le glycérol ou l'isooctane. L'étude de l'influence du substrat de préculture sur la cinétique de dégradation du 2-EHN, décrite p 88-89, montre qu'avec l'acétate, du succinate ou du Tween 80 comme substrat de préculture, la cinétique de dégradation du 2-EHN est similaire. J'ai aussi testé le glucose (qui provoque souvent une répression catabolique) et l'éthanol pour les cultures et les précultures de la souche IFP 2173. J'ai comparé les résultats d'électrophorèse 1D et 2D et j'ai choisi d'utiliser l'acétate comme substrat de référence. C'est un bon substrat pour la souche IFP 2173 et de plus, le métabolisme de l'acétate ne met pas en jeu d'enzymes spécifiques de la dégradation des acides gras. Une illustration de ces résultats pour l'électrophorèse bi-dimensionnelle est représentée par la **figure 4.7**. En comparant les gels 2D sur acétate et sur 2-EHN, j'ai pu identifier une trentaine de protéines spécifiques à la culture sur 2-EHN.

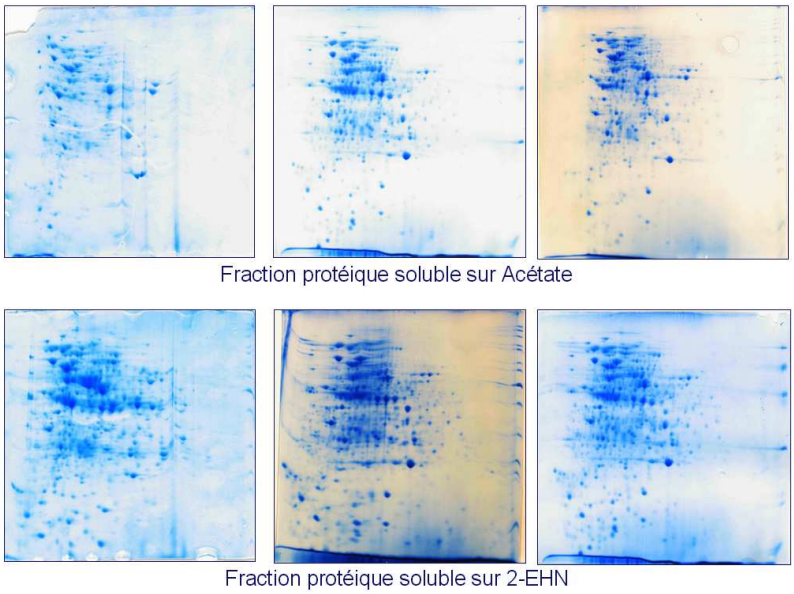

***Figure 4.7** : Analyse bi-dimensionnelle des protéines solubles de la souche IFP 2173 sur acétate et sur 2-EHN.Des échantillons de 450 µg de protéines ont été déposées sur chacun des gels colorés au bleu colloïdal.*

Cette approche protéomique différentielle, basée sur une comparaison visuelle des gels 2DE a permis de sélectionner des spots 2DE spécifiquement induits sur 2-EHN. Cependant, bien que les gels aient été reproduits au moins trois fois, l'analyse n'est pas quantitative. Il aurait fallu pour cela traiter les résultats de réplicats de gels avec un logiciel d'analyse comme Melanie

(Image Master) et faire des analyses statistiques. Cependant, nous avons voulu confirmer nos résultats par un marquage métabolique en présence des acides aminés radioactifs.

3.1.1.3 Marquage métabolique au ^{35}S

Le marquage métabolique consiste à introduire dans le milieu de culture des acides aminés radioactifs, comme par exemple la méthionine marquée au ^{35}S.

Lors de la culture de la souche IFP 2173 sur 2-EHN et sur acétate, la méthionine et la cystéine marquées au ^{35}S ont été ajoutées et se sont incorporées dans les protéines *in vivo* au cours de l'incubation (5h sur acétate et 30 h sur 2-EHN). Les protéines ont ensuite été extraites et analysées en électrophorèse 1D pour la fraction protéique membranaire et 2DE pour les protéines solubles selon le même mode opératoire que celui décrit dans la partie matériel et méthodes (pages 70-75).

Les résultats d'autoradiographie des gels ont révélé que la plupart des spots de protéines spécifiques à l'extrait 2-EHN étaient les mêmes que ceux déjà mis en évidence par coloration au bleu de Coomassie.

3.1.2 Identification des protéines en LC-MS/MS

Le principe d'analyse des spots 2DE ou bandes 1DE est représenté par le schéma de la **figure 4.8**. Les protéines contenues dans le gel sont digérées par la trypsine qui coupe les polypeptides après une lysine ou une arginine. Le mélange de peptides ainsi obtenu est analysé en chromatographie liquide sur une microcolonne puis analysé dans un premier détecteur de masse qui permet de donner une carte de la masse de tous les peptides (ions parents). Chaque peptide est alors fragmenté dans le second analyseur de masse ce qui permet d'établir sa séquence en acides aminés (ions fils). Ces informations de séquence sont confrontées à celles des bases de données pour identifier les protéines dont les peptides sont dérivés.

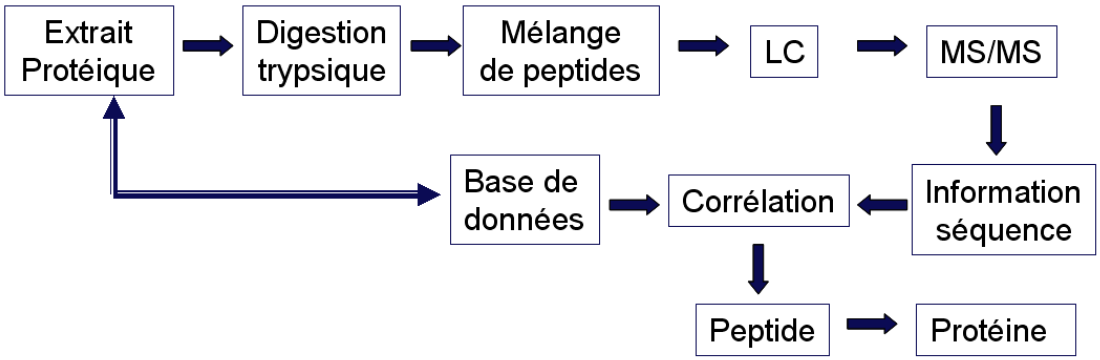

Figure 4.8 : Méthode d'analyse protéomique par LC-MS/MS.

Pour exploiter toutes les séquences peptidiques issues de l'analyse MS/MS, un logiciel, MASCOT, permet d'analyser ces séquences en les confrontant aux bases de données (**figure 4.9**). Grâce au logiciel MASCOT (logiciel d'identification des protéines utilisant des résultats de spectrométrie de masse, Matrix Science, http://www.matrixscience.com/pdf/Brochure_01-2-2005.pdf), les données analytiques sont comparées aux données théoriques obtenues par digestion *in silico* des protéines disponibles dans les bases. A ce jour 18 génomes de mycobactéries séquencés et annotés sont disponibles dans les bases de données et il a 87925 entrées de protéines de mycobactéries dans la base de données UniProtKB/TrEMBL. En général, chaque protéine est identifiée par plusieurs peptides. Les peptides qui ne correspondent pas à 100 % aux peptides théoriques de la protéine de référence ne sont pas inclus dans l'analyse. Un **score** basé sur le résultat d'analyse d'un peptide donné en MS/MS est attribué à chaque peptide. Le score est 10 * Log (P) où P est la probabilité que la corrélation entre le peptide analysé et la séquence peptidique de référence dans les bases de données soit un évènement aléatoire. Si le score individuel d'un peptide est supérieur à 39, cela signifie que la probabilité pour que le peptide analysé et le peptide de référence soit identiques de manière aléatoire est inférieure à 5 %. Les scores des protéines sont calculés en faisant la somme des scores des peptides constitutifs. Enfin, Mascot calcule le % de la protéine couvert par les peptides retrouvés lors de l'analyse en LC-MS/MS ; c'est le taux de recouvrement ou « **coverage** ».

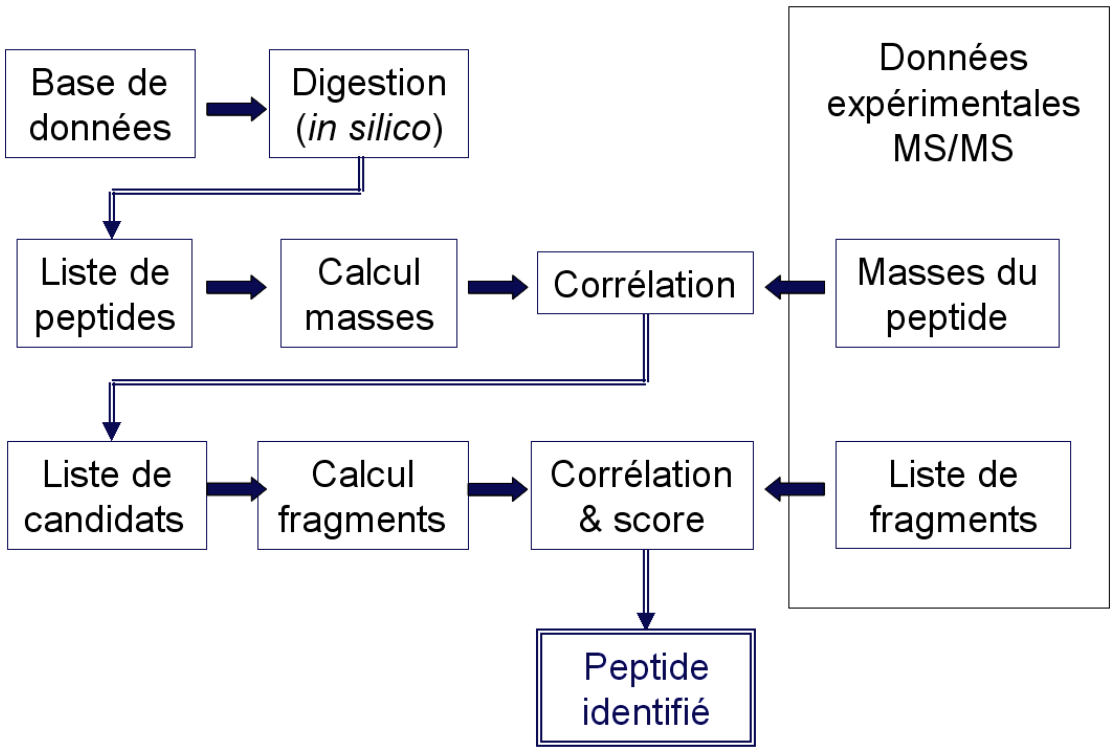

Figure 4.9 : Principe d'identification des protéines par MASCOT

Afin de transférer les données d'identifications générées par Mascot sur Excel, un logiciel libre, élaboré au Laboratoire d'Etude Dynamique des Protéomes (CEA de Grenoble) nommé IRMa (IRMa, CEA/DSV/iRTSV/LEDyP) est utilisé. Ce logiciel permet en outre de vérifier les peptides identifiés lorsqu'ils présentent un faible score d'identification. Il tient compte des mauvaises coupures par la trypsine (« miscleavage »), des modifications des peptides comme l'oxydation de méthionines lors de la préparation des échantillons. Et il indique certains résultats de masse comme le nombre de spectres permettant l'identification d'un peptide et le nombre de peptides servant à identifier une protéine. Ces valeurs permettront de donner une valeur d'abondance relative des protéines.

Une protéine peut être identifiée par au minimum un peptide à condition que son score soit supérieur à 120. Plusieurs protéines sont identifiées (1 à 5 en moyenne) par set de peptides analysé pour un spot 2D. Le score, le nombre de peptides servant à l'identification et le « coverage » de chacune de ces protéines permet de les différencier. La protéine qui a le meilleur score et le plus grand nombre de peptides identifiés est probablement la protéine d'intérêt, ou du moins la plus abondante dans l'échantillon.
Sur un gel SDS-PAGE, toutes les protéines sont séparées uniquement en fonction de leur masse. Dans une bande, il peut y avoir jusqu'à 100 protéines, qui une fois digérées à la trypsine donneront plusieurs centaines de peptides à identifier. Dans le lot de protéines identifiées, il est alors difficile de déterminer celles qui sont effectivement impliquées dans la dégradation du 2-EHN. Dans ce qui suit, nous avons utilisé une approche comparative, extrait 2-EHN versus extrait acétate, pour l'analyse des protéines membranaires.

3.1.3 Une analyse quantitative des protéines membranaires

L'approche mise en œuvre est basée sur une méthode semi-quantitative qui permet de comparer les extraits protéiques de bactéries exposées au 2-EHN à des bactéries témoins (sur acétate) en analysant l'intégralité des protéines séparables par SDS-PAGE. Sur chaque piste du gel d'électrophorèse exactement 5 µg de protéines ont été séparées. Pour chacune des pistes, 13 bandes ont été découpées, et chacune d'elles a été analysée par digestion peptidique et LC/MS-MS comme indiqué précédemment.
A partir des données d'identification de peptides, on peut obtenir des données semi quantitatives en se basant sur le nombre de peptides observés pour identifier une protéine et à la qualité des spectres d'identification des peptides. Il est alors possible de donner une estimation de la quantité relative des protéines présentes dans l'échantillon. Je présente ici

deux méthodes utilisées pour calculer l'abondance relative des protéines dans nos expériences.

3.1.3.1 Protein Abundance Index

Le nombre de peptides observés permettant l'identification d'une protéine est comparé au nombre de peptides observables. Les peptides observables représentent l'ensemble des peptides théoriquement produits par la digestion totale de la protéine sauf ceux qui sont exclus par les paramètres de l'analyse c'est à dire les trop petits, les trop grands et les très hydrophobes. Le rapport entre peptides observés et peptides observables est défini comme étant le « Protein Abundance Index » ou PAI. Le PAI est fonction de l'abondance relative des protéines car, plus une protéine est abondante plus il y a de chances de détecter ses peptides constitutifs par LC/MS. Dans la littérature, on trouve plutôt une forme exponentielle de ce rapport, donnée par la formule ci-dessous pour que les valeurs numériques obtenues soient comprises entre 0 et 9. Le résultat de ce calcul donne à chaque protéine une valeur d'emPAI pour « exponentially modified Protein Abundance Index ».

$$\mathbf{emPAI = 10^{PAI}-1}$$

Cette analyse permet d'une part d'identifier de manière assez exhaustive les protéines présentes dans chacun des échantillons (témoin et essai) et d'autre part de comparer le niveau d'expression des protéines entre les deux conditions expérimentales.

La méthode mathématique utilisée par Mascot utilisée pour calculer le nombre de peptides observés par rapport au nombre de peptides observables est basée sur la publication de Ishihama *et al.* (2005). Les auteurs calculent le nombre de peptides observables en fonction de la masse de la protéine et du nombre de lysine et d'arginine. Concernant le nombre de peptides observés, ils ne tiennent pas compte de leur état de charge. Suite au passage des échantillons au niveau de l'électrospray, les peptides sont chargés une, deux voire trois fois. Ainsi un même peptide peut être compté deux ou trois fois. Ce qui n'est pas représentatif du nombre de peptides observés expérimentalement. De ce fait, la méthode de calcul des emPAI a été modifiée pour ne pas compter plusieurs fois les peptides ayant plusieurs états de charge (nouvelle version du logiciel IRMa).

Afin de comparer les pistes acétate et 2-EHN le pourcentage d'abondance des protéines déduit des valeurs des emPAI est calculé selon la formule : % abondance = emPAI * 100 / Σ(emPAI) de la piste. Il est indiqué dans l'annexe 1 par l'expression « notre emPAI » à la différence de l'emPAI calculé par Mascot.

3.1.3.2 Spectral count

Le spectral count représente le nombre total de spectres peptidiques identifiés pour une protéine. C'est une approche de quantification plus fine que le calcul des emPAI. Elle est citée dans de nombreux articles récents comme Xia *et al.* (2007) et Fu *et al.* (2008).

Un pourcentage d'abondance est aussi calculé. Pour pouvoir faire la comparaison entre la piste acétate et la piste 2-EHN, il est nécessaire de comparer la valeur du spectral count à la somme des valeurs de spectral count de l'échantillon. C'est sur cette base que nous nous baserons pour analyser les résultats de cette étude protéomique.

Dans le cas de l'analyse des protéines de la souche IFP 2173, la bactérie de référence est *M. vanbaalenii* PYR-1, celle qui vient après est *M. gilvum* PYR-GCK, deux bactéries qui ont été isolées pour leur capacité à dégrader le pyrène. Même si la souche *M. austroafricanum* IFP 2173 a de nombreuses protéines proches de celles de *M. vanbaalenii* PYR-1, cette dernière n'est pas capable de dégrader le 2-EHN.

Les résultats de nos travaux sont présentés dans l'article 3 :

3.2 Article 3

Proteomic identification of enzymes involved in 2-EHN biodegradation in *Mycobacterium austroafricanum* IFP2173

Elodie Nicolau[1], Lauriane Kuhn[2], Rémy Marchal[3], Yves Jouanneau[1]

[1] Laboratoire de Chimie et Biologie des Métaux, and [2] Laboratoire d'Étude de la Dynamique des Protéomes, CEA, DSV, IRTSV, 38054 GRENOBLE, France.

[3] Département de Biotechnologie, Institut Français du Pétrole, 92852 Rueil-Malmaison Cedex, France

Keywords *: Mycobacterium austroafricanum*, 2-ethylhexyl nitrate, alkane hydroxylase, AlkB, CYP153, β-oxidation.

Running title: Enzymes involved in 2-EHN degradation

Corresponding author:

Yves Jouanneau
LCBM/iRTSV, CEA-Grenoble
F-38054 Grenoble Cedex 9, France.
Tel. : 33 (0)4 38 78.43.10; Fax : 33 (0)4 38 78.51.85
Email : yves.jouanneau@cea.fr

Abstract

2-Ethyhexyl nitrate (2-EHN) is a synthetic chemical used as a diesel fuel additive, which can be slowly degraded by *Mycobacterium austroafricanum* IFP2173. In this study, the enzymes involved in 2-EHN degradation have been investigated by proteomic analyses. Two-dimensional gel electrophoresis showed that 35 cytoplasmic proteins were up-regulated in cells exposed to 2-EHN. Most proteins were identified as enzymes of the β-oxidation. A thorough analysis of the proteome identified over 1300 proteins, 365 of which were found to be up-regulated on 2-EHN. These proteins include multiple isoenzymes of the β-oxidation, two alcohol and one aldehyde dehydrogenases as well as proteins involved in regulation and response to stress. Interestingly, four cytochromes P450 appeared to be induced on 2-EHN, including one CYP153 which functions as an alkane hydroxylase. Like *Mycobacterium vanbaalenii* PYR-1, strain IFP2173 was found to contain two *alkB*-like genes encoding putative membrane-bound alkane hydroxylases. RT-PCR experiments demonstrated that the gene encoding the CYP153 protein as well as the *alkB* genes were expressed on 2-EHN. Hence, almost all the enzymes required for the biodegradation of 2-EHN by strain IFP2173 have been identified, including three oxygenases that might catalyse the initial hydroxylation step.

1. Introduction

2-ethylhexyl nitrate (2-EHN) is a xenobiotic compound used as a gasoline additive. Due to its explosive properties, 2-EHN is considered as the best cetane improver for diesel oil, including bio-diesels that might be used in the near future [1, 2]. In case of accidental release, 2-EHN is a serious health hazard as humans exposed to this chemical were found to suffer from various symptoms including headache, dizziness, chest discomfort, palpitations or nausea [3]. Although 2-EHN was considered not readily biodegradable by US EPA [4], it was recently reported that it could be degraded by *Mycobacterium austroafricanum* [5].

Soil *Mycobacteria* have been described for their ability to degrade a wide range of aliphatic and aromatic hydrocarbons, including polycyclic aromatic hydrocarbons (PAHs) [6-8], and fuel additives such as methyl *tertio*-butyl-ether (MTBE) [9, 10]. These bacteria are well equipped to degrade hydrocarbons, which they used as carbon sources. Their bacterial wall, rich in mycolic acids, confers resistance to toxic hydrophobic pollutants and may facilitate access to hydrocarbons [11, 12]. Moreover, soil *Mycobacteria* contain oxygenases of different types, which play a crucial role in the degradation of both aliphatic and aromatic hydrocarbons. For instance, monooxygenases catalyze the first step in the degradation of alkanes. C_5-C_{16} alkanes are substrates of two kinds of enzymes, either integral-membrane non-heme diiron monooxygenases (AlkB) [13], or cytochromes P450 [14]. Growth on alkanes requires metabolic adaptation as shown for the marine bacterium *Alcanivorax borkumensis* SK2 [15]. Proteomic analysis revealed that alkane utilisation may proceed via different pathways, involving two AlkB hydroxylases, one putative flavin monooxygenase and three P450 cytochromes. Bacterial adaptation to alkane utilisation resulted in a strongly modified metabolism, with consequences on carbon flux and membrane lipid composition [15].

In previous studies, *M. austroafricanum* IFP 2173 was isolated on iso-octane [16], then selected for its ability to degrade 2-EHN [5]. Degradation of 2-EHN was found to be incomplete, yielding a 6-carbon γ-lactone, which accumulated as a dead-end product. A degradation pathway was proposed involving hydroxylation of the methyl group in distal position, then oxidation to the carboxylic acid, and further metabolism through one cycle of β-oxidation [17]. In order to identify the enzymes involved in this pathway, we have undertaken a proteomic analysis of the proteins up-regulated in cells exposed to 2-EHN. Because the genome sequence of strain IFP 2173 is unknown, we tentatively identified relevant proteins by comparing their peptide sequences to those of orthologs found in the data bases, especially proteins from fast-growing *Mycobacteria*. Besides focusing on enzymes

involved in 2-EHN degradation, this study gives an insight into proteins possibly involved in the response of bacteria to exposure to a toxic and hydrophobic xenobiotic compound.

2. Material and methods

2.1. Bacterial strain and growth conditions

M. austroafricanum strain IFP 2173 was grown on a mineral salts medium at 30°C as described previously [17]. The carbon source was acetate (4 g/l) and 2-EHN (500 mg/l). Growth was monitored by measurements of the optical density (OD) at 600 nm. To prepare 2-EHN-induced cells, acetate grown-cells were washed and resuspended to an OD of 1.5 in culture medium, then incubated for five days with 2-EHN.

2.2. In vivo 35S labelling of proteins

For labelling experiments, bacteria were incubated with 2-EHN or acetate (control) in the presence of a mixture of ^{35}S-labelled methionine and cysteine (0.1 mCi, Easytag Express protein-labelling mix; NEN Life Science Products). Bacteria were incubated for 6 h (acetate) or 30 h (2-EHN) at 30°C. In a control experiment, bacteria were incubated without exogenous C-source for 30h. Protein extracts were prepared as described below and analyzed by 2D electrophoresis and SDS-PAGE.

2.3. Preparation of protein extracts

Cell-free extracts were prepared by ultrasonication as previously described [18]. Ultracentrifugation at 240,000 *g* for I h was performed to separate soluble proteins from the membrane fraction, using an Optima TLX Ultracentrifuge (Berckman Instruments). Supernatant fractions were treated with benzonase (2,000 U; Merck), and subsequently dialysed for 4-5 h at 4°C against 5 mM phosphate buffer, pH 7.5, containing 1 mM $MgCl_2$, then overnight against ultrapure water. Samples were immediately processed as described below or stored at -20°C.

2.4. Two dimensional gel electrophoresis

Two-dimensional (2D) gel electrophoresis was carried out as described previously [18], with minor modifications. Briefly, 400 µg protein samples (or labelled extracts equivalent to 4.2 10^4 cpm) were applied to 18-cm IPG strips (ReadyStrip; Biorad) and fractionated in the pH range 4 to 7 by isoelectric focusing for a total of approximately 70 kVh. Second dimension

electrophoresis was carried out on 12.5 % polyacrylamide gels in a Protean II xi cell (Biorad) at 20 mA per gel for 15-16 h, using a Tris-glycine buffer system [19]. The proteins were visualised by colloidal blue G-250 staining as described by Neuhoff [20], except that ethanol replaced methanol. Gels were photographed with a Canon camera. ^{35}S labelled gels were stained, dried, and exposed to X-ray films for 3 weeks (Kodak BioMax MR). All gels were performed in triplicate, except for gels containing labelled proteins. Comparison of 2D gel patterns and spot intensities was carried out by visual inspection of gels. Only spots that were absent in control extracts or that repeatedly showed an estimated intensity at least twice as high on 2-EHN extracts compared to control extracts were taken into consideration. Proteins up-regulated in acetate-grown cells are not discussed in this study.

2.5. SDS-PAGE of membrane fractions

Protein pellets from high speed ultracentrifugation were homogenized in a volume equivalent to 1/10 the initial volume of extract with 25 mM HEPES pH 7.5, containing 10 % of ethylene glycol. Protein samples were adjusted to 1 % SDS, 2.5 % β-mercaptoethanol, 10% glycerol, 0.001 % bromophenol blue and 150 mM Tris-HCl, pH 7.0, prior to separation by SDS-PAGE on a 12.5 % polyacrylamide gel in the Tris-Tricine buffer system [21]. Proteins were stained with colloidal blue G-250 as described above.

2.6. Protein digestion

Protein bands were manually excised from the gels and automatically prepared (EVO150, Tecan). Samples were washed several times by incubation in 25 mM NH_4HCO_3 for 15 min and then in 50 % (v/v) acetonitrile containing 25mM NH_4HCO_3 for 15 min. Gel pieces were then dehydrated with 100 % acetonitrile and then incubated with 7 % H_2O_2 for 15 min before being washed again with the destaining solutions described above. 0.15 µg of modified trypsin (Promega, sequencing grade) in 25 mM NH_4HCO_3 was added to the dehydrated gel spots for an overnight incubation at 37°C. Peptides were then extracted from gel pieces in three 15 min sequential extraction steps in 30 µL of 50% acetonitrile, 30 µL of 5% formic acid and finally 30µL of 100% acetonitrile. The pooled supernatants were then dried under vacuum.

2.7. Nano-LC-MS/MS analysis.

For nano-LC-MS/MS analysis, the dried extracted peptides were resuspended in water containing 2.5 % acetonitrile and 2.5 % trifluoroacetic acid. A nano-LC-MS/MS analysis was

then performed (Ultimate 3000, Dionex and LTQ-Orbitrap, Thermo Fischer Scientific). The method consisted in a 60-minute gradient at a flow rate of 300 nL/min using a gradient from two solvents: A (5% acetonitrile and 0.1% formic acid in water) and B (80% acetonitrile and 0.08% formic acid in water). The system includes: a 300 µm X 5 mm PepMap C18 precolumn and a 75 µm X 150 mm C18 column (Gemini C18 phase). MS and MS/MS data were acquired using Xcalibur (Thermo Fischer Scientific) and processed automatically using Mascot Daemon software (Matrix Science).

Consecutive searches against the SwissProt/Trembl database were performed for each sample using an intranet version of Mascot 2.0. Peptide modifications allowed during the search were: N-acetyl (protein), dioxidation (M), oxidation (M), cysteic acid (C) and sulphone (M). Proteins showing two peptides with a score higher than the query threshold (p-value <0.05) were automatically validated with an home-made software (IRMa, CEA/DSV/iRTSV/LEDyP). Each protein identified by only one peptide was checked manually using the classical fragmentation rules.

For each identified protein, the emPAI and spectral count values were calculated and reported in the result tables (http://www.matrixscience.com/help/quant_empai_help.html). From these values, abundance rates were then calculated as percentages of the whole set of identified proteins according to the formula: Abundance of proteinX = (spectral count (proteinX) * 100) / (Σ spectral counts for all_proteins).

2.8. RNA extraction and RT-PCR analysis

Total RNA was extracted from 50-mL cultures of strain IFP 2173 using standard procedures [22]. Bacteria were grown to an OD_{600} of 0.7, then centrifuged at 10,000 g, and resuspended in 200 µL of 20 mM Tris-HCl, 5 mM EDTA, pH 8, containing 1.5 mg/mL lysozyme and 25 µg/mL lysostaphine (Sigma Life Science). After 10 min at 37°C, RNA were extracted using the RiboPure™-Bacteria kit (Ambion, Austin, Texas). Crude RNA samples (2 µg) were treated with Turbo DNase (Ambion). Resulting RNA preparations were quantified using a Nanodrop apparatus (NanoDrop Technologies). RT-PCR was performed with 10 ng of RNA preparation using the One step RT-PCR kit (Promega, France). PCR amplification of internal gene sequences was carried out with the following primer pairs: for *alkB1*, alkB1-F (5'-CGTGATCATGGGTGCCTAC-3') and alkB1-R (5'-CCAGAACGTCTCACCGAAG-3'); for *alkB2*, alkB2-F (5'-CCTGATGTTCCTCGTGATCC-3') and alkB2-R (5'-CTTGTCGACGTCGCTCATC-3'); for CYP153, P450fw1 and P450rw3 [23]; for the aldehyde dehydrogenase encoding gene (*alkH*), ALDH1-F (5'-

GCACCGTGCTGATCATCGGTGC-3') and ALDH1-R (5'-CCAGGCGATGCGCTTGGCG-3'), for the 16S RNA gene, P16S-F (5'-GGTCTAATACCGAATACACCCTTCT-3') and P16S-R (5'-CCAGGAATTCCAGTCTCCC-3'). RT-PCR reactions were carried out as follows: 45 min at 45°C, 3 min at 95°C, 32 cycles of 30 s at 95°C, 30 s at 62°C and 30 s at 72°C, 5 min final elongation at 72°C. Products were analyzed by electrophoresis on 2% agarose gels.

2.9. Nucleotide sequences

The nucleotide sequences *alkB1rubA1rubA2tetR*, *alkB2*, CYP153, *alkH* (partial) were deposited under accession number FJ009005, FJ009004, FJ009003, FJ2074720, respectively.

3. Results

3.1. Identification of cytoplasmic proteins up-regulated on 2-EHN

In order to identify proteins up-regulated on 2-EHN, protein profiles of cells incubated with this compound were compared to those of cells grown on actetate. Cytoplasmic proteins were prepared and analysed by 2D gel electrophoresis while membrane-associated proteins present in the high-speed centrifugation pellet of cell extracts were separated by SDS-PAGE.

Comparison of 2D gel protein profiles revealed that more than 30 protein spots were either absent in acetate-grown cells or at least two-fold more abundant in 2-EHN-grown cells (Fig. 1). To confirm these results, we performed ^{35}S-labeling experiments where cells were exposed to 2-EHN for 30 h (6 h for reference cells). Autoradiographies of the 2D gel showed markedly different patterns (Fig.1, C & D). Most labelled proteins uniquely detected in 2-EHN-exposed cells corresponded to spots previously identified based on stained gel comparison, but three additional 2-EHN-specific polypeptides were found. The position of these extra polypeptides has been reported on the 2D image in Fig. 1A. The proteins of interest were tentatively identified by LC-MS/MS analysis of trypsic digests followed by a search for peptide matches in the data bases using Mascot (see Materials and Methods for details). A total of 35 proteins were identified based on close similarities with orthologs found in *M. vanbalenii* PYR1 and related *Mycobacterium* strains from soil (Table 1). One set of induced proteins was clearly associated to the β-oxidation of fatty acids. Some of the enzymes involved in this pathway were found in multiple isoforms, including acetyl-CoA acyltransferase (2 copies), acyl-CoA dehydrogenase (7 copies), enoyl-CoA hydratase/isomerase (4 copies). Consistent with this finding, the reference *Mycobacterium* strains mentioned above were found to contain multiple

gene copies coding for enzymes of the β-oxidation in their genome (Table S1). Other proteins up-regulated on 2-EHN included dehydrogenases, diverse metabolic enzymes and proteins involved in cell response to stress (Table 1).

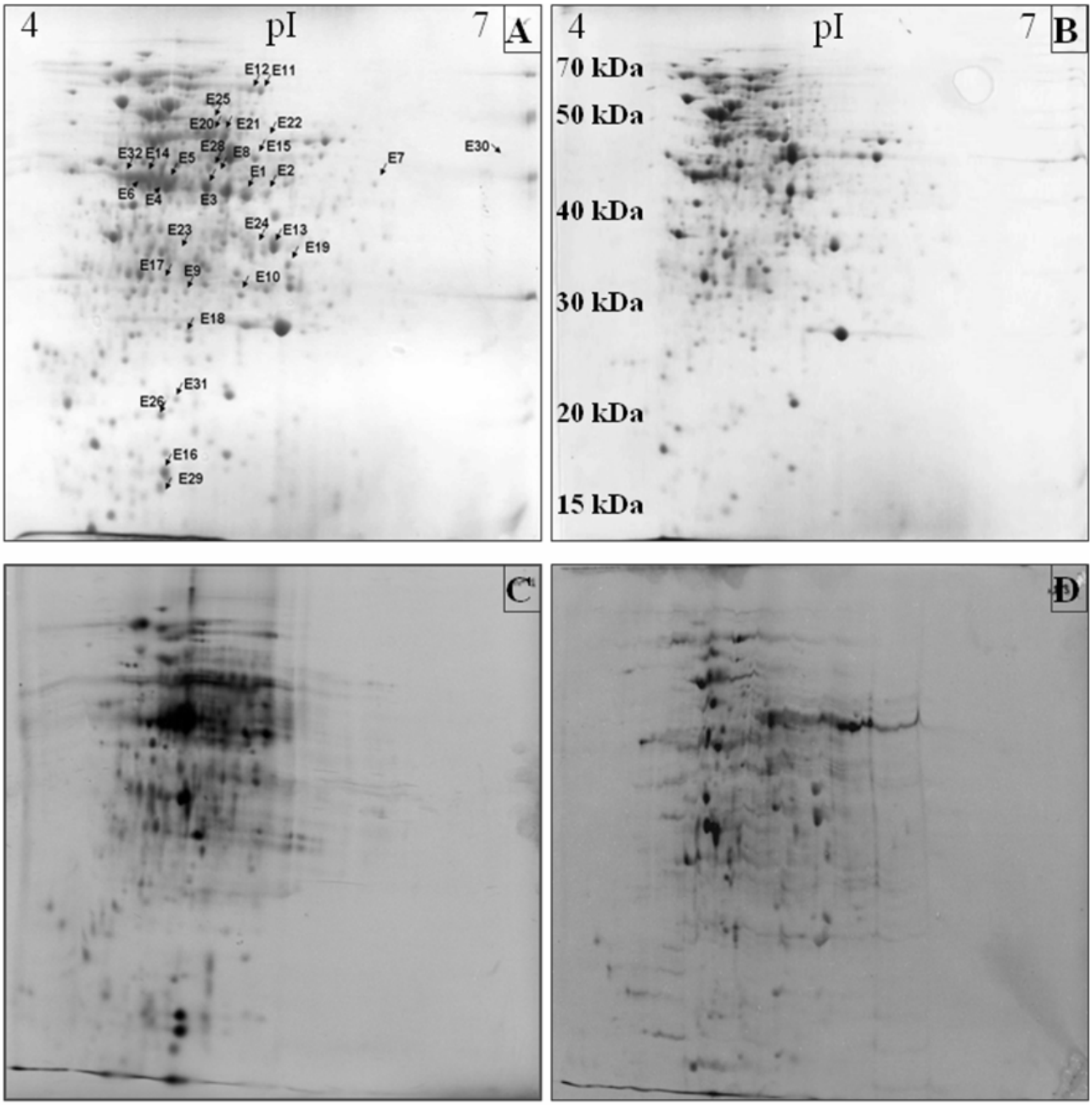

Figure 1: 2D gel map of soluble proteins from *M. austroafricanum* IFP 2173.
Isoelectric focusing was performed in the 4-7 pH range.
A: cells grown on 2-EHN, **B**: cells grown on acetate, **C**: ^{35}S labelled proteins induced on 2-EHN, **D**: ^{35}S labelled proteins induced on acetate. 2-EHN-specific protein spots are numbered in panel A.

Table 1: Cytoplasmic proteins up-regulated upon incubation of *M. austroafricanum* IFP 2173 on 2-EHN

Enzyme or protein function	Spot #	2-EHN induction	^{35}S-labelling on 2-EHN	pI	Mol Mass	Score	Coverage (%)	Peptides	SwissProt/ TrEMBL	Micro-organism
Fatty acid metabolism										
Acyl-CoA dehydrogenase domain protein	E1	+++	+++	5,3	41116	186	9,95	3	A4TFJ1	*M. gilvum* PYR-GC
	E2	+++	+++	5,4		180	8,58	3		
	E1	+++	+++	5,3	40821	133	10,78	3	A3Q1G5	*Mycobacterium* sp. JLS
	E2	+++	+++	5,4	40989	702	31,72	10	A1TCG6	*M.vanbaalenii* PYR-1
	E3	+++	-	5,15	42234	112	9,92	2	A4T8F2	*M. gilvum* PYR-GC
	E4	+++	+++	4,8	39744	265	18,01	4	A1TDA6	*M. vanbaalenii* PYR-1
	E7	++	++	5,87	42108	118	5,74	2	A4TFI8	*M. gilvum* PYR-GC
	E8	+++	+	5,2	41520	585	36,32	9	A1T5U2	*M. vanbaalenii* PYR-1
	E9	+++	+	4,9	29971	133	9,09	2	A1TDW3	*M. vanbaalenii* PYR-1
Enoyl-CoA hydratase/isomerase	E10	-	+++	5,1	27387	121	11,29	2	A4TFL1	*M. gilvum* PYR-GC
	E9	+++	+++	4,85	27478	280	21,69	5	A1TE56	*M. vanbaalenii* PYR-1
	E9	+++	+++	4,85	26666	35	5,37	1	A4TDN8	*M. gilvum* PYR-GC
3-hydroxyacyl-CoA dehydrogenase, NAD-binding	E11	++	++	5,4	76177	586	19,36	11	A1TF87	*M. vanbaalenii* PYR-1
	E12	++	++	5,4		381	10,84	7		
Acetyl-CoA C-acyltransferase (EC 2.3.1.16)	E1	+++	+++	5,4	39836	591	32,41	9	A1TDW4	*M. vanbaalenii* PYR-1
	E13	++	+++	4,9	42130	189	11,49	3	A1TF88	*M. vanbaalenii* PYR-1
Acyl-ACP thioesterase	E10	+++	+++	5,3	70417	117,03	3,28	2	A1T388	*M. vanbaalenii* PYR-1
Dehydrogenases										
Alcohol dehydrogenase GroES domain protein	E14	+++	+++	4,75		370	18,89	5	A4TFL6	*M. gilvum* PYR-GC
	E4	+++	+++	4,77	38681	409	26,39	5		
	E5	+++	+++	4,8		307	17,95	4		
FAD-dependent pyridine nucleotide-disulphide oxidoreductase	E15	+++	+++	5,55	42755	319	17,78	5	A4TFL9	*M. gilvum* PYR-GC
	E7	+++	+	5,8		152	9,21	3		
Short-chain dehydrogenase/reductase SDR	E17	+++	++	4,9	29920	164	15,44	3	A1T1A7	*M. vanbaalenii* PYR-1
	E18	+++	+++	4,9	30143	102	10,22	2	Q1BFX1	*Mycobacterium* sp. MCS

Enzyme or protein function	Spot #	2-EHN induction	^{35}S-labelling on 2-EHN	pI	Mol Mass	Score	Coverage (%)	Peptides	SwissProt/ TrEMBL	Micro-organism
(S)-2-hydroxy-acid oxidase	E30	+++	-	7	42022	169	9,56	3	A1T4N1	*M. vanbaalenii* PYR-1
Dihydrolipoamide dehydrogenase	E22	++	++	5,6	49719	370	17,52	6	A1T382	*M. vanbaalenii* PYR-1
Lysine biosynthesis										
Dihydrodipicolinate synthase	E24	++	-	5,4	31436.7	163	11,81	3	A1T7Q1	*M. vanbaalenii* PYR-1
Dihydrodipicolinate reductase (EC 1.3.1.26)	E23	++	-	4,8	25816	59	7,69	1	A1T7N8	*M. vanbaalenii* PYR-1
Protein synthesis										
Serine-tRNA ligase (EC 6.1.1.11)	E25	+++	+++	4,8	60542.4	128	5,82	2	A1TGX4	*M. vanbaalenii* PYR-1
Ketol-acid reductoisomerase (EC 1.1.1.86)	E3	+++	+++	5,2	36513	158	5,44	2	Q1BAR7	*Mycobacterium sp. CMS*
Nitrogen assimilation										
Alanine dehydrogenase (EC 1.4.1.1)	E28	-	+++	5,2	38907	203	10,76	3	A1T7L9	*M. vanbaalenii* PYR-1
Oxidative phosphorylation										
ATP synthase epsilon chain (EC 3.6.3.14) (ATP synthase F1 sector epsilon subunit)	E29	+++	+++	4,8	13330	85	9,09	1	P45822	*M. leprae* TN
CO_2 hydratation										
Carbonic anhydrase	E31	++	++	4,8	18225	272	32,73	6	A1TDF0	*M. vanbaalenii* PYR-1
Glycolysis / glyconeogenesis										
Phosphoglycerate kinase (EC 2.7.2.3)	E32 E6	+++ +++	+++ +++	4,7 4,75	42102	435 572	24,08 27,75	6 8	A1T8L1	*M. vanbaalenii* PYR-1
Stress response										
Heat shock protein Hsp20	E26	+++	+++	4,8	15648.4	177	26,76	3	A1T4V8	*M. vanbaalenii* PYR-1
UspA	E24	+++	+++	5,4	31354	85	4,91	1	A1T4W2	*M. vanbaalenii* PYR-1
Miscellaneous										
Putative esterase precursor	E18	+++	+++	5	35010	79	6,29	1	A1T6C2	*M. vanbaalenii* PYR-1
Chloride peroxidase (EC 1.11.1.10)	E19	++	-	5,7	30410	173	14,86	3	A1T5E7	*M. vanbaalenii* PYR-1

Enzyme or protein function	**Spot #**	**2-EHN induction**	**^{35}S-labelling on 2-EHN**	**pI**	**Mol Mass**	**Score**	**Coverage (%)**	**Peptides**	**SwissProt/ TrEMBL**	**Micro-organism**
Antibiotic biosynthesis monooxygenase	E16	+++	+++	4,8	11741	56	16,04	1	Q1B2M9	*Mycobacterium* sp. MCS
Allophanate hydrolase subunit 1	E10	+++	+++	5,3	25092	142	9,65	2	A1T1V3	*M. vanbaalenii* PYR-1
Fumarate lyase	E20	+++	++	5,1	49944	579	24,25	10	A1TE24	*M. vanbaalenii* PYR-1
	E21	+++	++	5,1	49944	538	27,37	8		
HpcH/HpaI aldolase	E23	++	-	4,8	29032	202	12,93	4	A1TCG4	*M. vanbaalenii* PYR-1
Ribonuclease PH (EC 2.7.7.56)	E23	++	-	4,8	27449	75	4,82	2	A1T7Q1	*M. vanbaalenii* PYR-1
Glycyl-tRNA synthetase, alpha2 dimer	E22	++	++	5,6	59543	405	18,11	7	A1TBP9	*M. vanbaalenii* PYR-1
3-hydroxyisobutyrate dehydrogenase precursor	E23	++	-	4,8	29262	195	15,79	3	A1T4U4	*M. vanbaalenii* PYR-1
Cyclic nucleotide-binding:regulatory protein, Crp	E27	+++	-	9,6	24776	409	39,56	7	A1T6A5	*M. vanbaalenii* PYR-1
Phosphoribosyltransferase: Erythromycin esterase	E12	++	++	5,4	74587	49	1,62	1	A1T4X7	*M. vanbaalenii* PYR-1

Table S1: Copy number of 2-EHN-induced proteins in strain IFP 2173 compared to the expected copy number of relevant proteins in closest strains as deduced from genome analysis

Enzymes possibly involved in 2-EHN biodegradation	Number of copies induced on 2-EHN		Number of gene copies coding for the enzyme		
	1D analysis	2D analysis	*M. vanbaalenii* PYR-1	*M. gilvum* PYR-GCK	*Mycobacterium* sp. KMS
Cytochrome P450/CYP153	2/2	0	49	51	52
Esterase	4	1	10	8	8
Alcohol dehydrogenase GroES domain protein	2	1	15	13	19
Aldehyde dehydrogenase	1	0	21	18	18
Acetyl-CoA acetyltransferase	1	2	13	17	4
Acyl CoA dehydrogenase domain protein	11	7	71	63	71
Enoyl-CoA hydratase/isomerase	2	4	20	15	40
3-hydroxyacyl-CoA dehydrogenase, NAD-binding	2	1	2	1	2
Thiolase	0	0	3	3	3

3.2. 2-EHN induction of membrane-associated proteins

Since membrane proteins are generally difficult to analyze by regular 2D gel electrophoresis, we chose to separate the insoluble high-speed fractions of cell extracts by one dimension SDS-PAGE. When stained protein profiles of 2-EHN versus acetate-grown cells were compared no obvious differences were observed. However, ^{35}S radioactive labelling revealed that some proteins, including a prominent 45 kDa, became clearly labelled upon exposure to 2-EHN (Fig. S1**)**. In order to identify proteins of interest, protein sets from 2-EHN and acetate grown cells were separated by SDS-PAGE, and subjected to proteomic analysis. Data processing using Mascot and a locally developed software (IRMa) identified over 1300 proteins most of which had counterparts in the proteome of *M. vanbalenii* PYR1. A summary of proteins found to be common or specific to each cell treatment is presented in tables S2 and S3 in supplementary material (at the end of the thesis). From the set of common proteins, a subset was selected based on abundance rates more than twice as high for the 2-EHN treated cells as compared to control cells. This protein subset (65 proteins) as well as those found to be specific to the 2-EHN treatment (300 proteins) were tentatively classified in terms of enzyme category or metabolic function, with special emphasis on enzymes related to alkane degradation. An inventory of the classified proteins is summarized in Table 2. A total of 17 proteins have been identified

as enzymes of the β-oxidation of fatty acids, four of which were already detected in the 2D gel analysis. Consistent with the 2D gel data, numerous proteins up-regulated on 2-EHN were dehydrogenases including 12 short-chain dehydrogenase/reductases (SDR). Several proteins were likely involved in the response to stress, other presumably act as transcriptional regulators. The analysis highlighted two alcohol dehydrogenases and one aldehyde dehydrogenase possibly implicated in the early steps of the 2-EHN degradation pathway (see below). Interestingly, enzymes of the central metabolism previously shown to be essential for alkane assimilation [15] have been detected including two phosphoenolpyruvate synthases involved in gluconeogenesis. Enzymes related to the metabolism of lipids were also identified, suggesting that membrane modifications might occur as part of the bacterial adaptation to growth on a hydrophobic substrate.

Four P450 cytochromes were identified, two of which are potentially capable of alkane hydroxylation. The latter two proteins, which belong to the CYP153 subclass of P450, were exclusively produced on 2-EHN. The most abundant of the two enzymes, identified by 11 peptides (36% coverage), was found to be closely related to the CYP153 enzyme from *Mycobacterium* sp. XHN-1500 [24].

It is to note that many cytoplasmic proteins were recovered in the membrane fractions, indicating that they might have been trapped in membrane vesicles that formed upon cell lysis. On the other hand, membrane proteins relevant to the metabolism of alkanes were not detected in either protein extract. This was the case for the trans-membrane AlkB hydroxylase that is known to catalyze the initial oxidation reaction of alkanes in many bacteria [13].

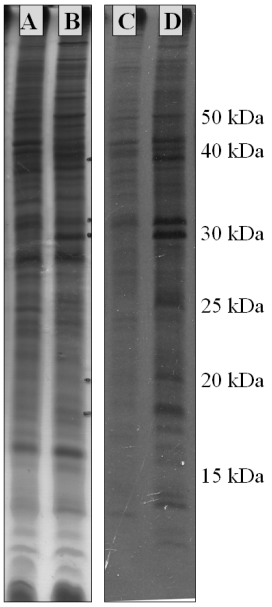

Figure S1: SDS-PAGE analysis of membrane-associated proteins in cells grown on acetate or exposed to 2-EHN.

A: protein extract from acetate grown cells, **B**: protein extract from cells exposed to 2-EHN, **C**: ^{35}S labelled proteins induced on acetate, **D**: ^{35}S labelled proteins induced on 2-EHN

Table 2: Functional classification of proteins more abundant or exclusively detected on 2-EHN

Enzyme or general protein function	Number of proteins	
	exclusively found on 2-EHN	More abundant on 2-EHN
Beta oxidation	17	1
P450 hydroxylases	4	0
Esterases	3	1
Dehydrogenases SDR	2 ADH 10	1 ALDH (1) 2
Stress response	12	9
Regulation	21	3
Lipid metabolism/cell wall	13/1	1/3
Ribosomal proteins	12	2
N-metabolism	3	1
General metabolism	74	22
ABC trasporter proteins	22	3
Other	64	6
Putative uncharacterized	42	10

3.3. Occurrence of several putative alkane monooxygenases in strain IFP2173

Our proteomic analysis revealed that one cytochrome P450 with close similarity with a well characterized alkane hydroxylase (CYP153) was 2-EHN-specific. In order to learn more on this protein, its structural gene was PCR-amplified using genomic DNA from strain IFP2173 and specific oligonucleotides designed after the gene sequence of *CYP153A* from strain XHN-1500 [24]. A 1261 bp gene was obtained, which displayed high sequence similarity with its counterpart from strain XHN-1500 (99 % identities), resulting in a predicted protein having only two amino acid changes compared to CYP153A.

In a previous study, a gene potentially involved in isoalkane degradation was found in strain IFP2173 and identified as an *alkB* gene based on partial sequence determination [16]. A BLAST search showed than this gene was closely related to *alkB* from *M. vanbaalenii* PYR-1. This strain has two alkane monooxygenase genes, one of which is associated with two genes encoding rubredoxins. Primers were designed after the *alkB* gene sequences of *M. vanbaalenii* PYR-1, and used to amplify corresponding genes from strain IFP 2173 genomic DNA. Two *alkB*-like genes were found in two separate loci, which displayed exactly the same gene arrangement as that found in *M. vanbaalenii* PYR-1. In particular, the *alkB1* gene was followed by two rubredoxin genes, named *rubA1* and *rubA2*, almost identical to homologous genes of *M. vanbaalenii* PYR-1 (99% identity). The second *alkB* gene (*alkB2*) was 98% identical to its counterpart in strain PYR-1.

3.4. RT-PCR evidence for the expression of three alkane hydroxylases in 2-EHN fed cells

Since none of the AlkB-like hydroxylase was detected in extracts of strain IFP2173 upon proteomic analysis, we carried out RT-PCR experiments to determine whether the corresponding genes were expressed under the growth conditions used in this study. Transcripts of the *alkB1* and *alkB2* genes were equally detectable in acetate and 2-EHN-fed cells (Fig. 2). Further analysis showed that a transcript specific for the gene encoding the CYP153 hydroxylase described above was also detected in both 2-EHN and acetate-grown cells (Fig. 2). This finding is consistent with the fact that the enzyme was clearly identified by proteomic analysis in 2-EHN-fed cells but contrasted with the finding that it was absent in acetate-grown cells. Perhaps cells growing on acetate produce the CYP153 protein at a low level or in a transient manner during a particular phase of growth, so that it passed undetected.

A single aldehyde dehydrogenase appeared to be up-regulated in 2-EHN-fed cells to a level at least 2-fold as high as in acetate-grown cells. Using primers designed after the gene encoding an orthologous aldehyde dehydrogenase from strain PYR-1 (A1P1A6), a DNA fragment, which perfectly matched the target gene sequence, was PCR-amplified from IFP2173 genomic DNA. RT-PCR demonstrated that this gene was transcribed in both acetate and 2-EHN fed cells (Fig. 2). The deduced sequence of the closely related PYR-1 enzyme displayed 39% sequence identity with the product of the *alkH* gene from *P. putida* GPo1.

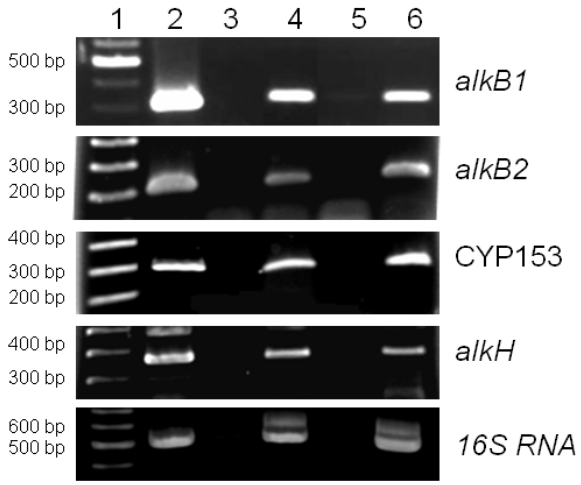

Figure. 2: Expression of genes relevant to 2-EHN degradation as analyzed by Reverse Transcription (RT)-PCR. Lane 1, DNA ladder ; lane 2, PCR amplicon generated from gDNA; lane 3, Control on RNA from 2-EHN-grown cells with reverse transcriptase omitted; lane 4, RNA from acetate-grown cells; lane 5, RNA from 2-EHN-grown cells.

4. Discussion

The present study deals with the metabolic adaptation of a bacterium which was forced to grow on a xenobiotic compound being a poor carbon source and a toxic substance. Our proteomic analysis identified over 1300 proteins based on sequence information available in the data bases even though the genome of strain IFP2173 was unknown. Most proteins were identified as orthologs from strain *M. vanbaalenii* PYR-1 or related *Mycobacterium* species, thus reflecting the high degree of conservation of protein sequences in the proteomes from fast-growing *Mycobacterium* species isolated from various places around the world. Besides catabolic enzymes allowing the bacterium to utilize 2-EHN as carbon source, many up-regulated proteins were found to be involved in lipid metabolism, regulation and response to stress, and might help bacteria to adapt to the toxic and/or hydrophobic character of 2-EHN. The following discussion focuses on enzymes that might be implicated in 2-EHN degradation**.**

In a previous study, we showed that strain IFP2173 partially degraded 2-EHN to a compound identified as 4-ethyltetrahydrofuran-2(3H)-one, and we proposed a degradation pathway outlined in figure 3 [17]. Every step in the pathway can be assigned at least one enzyme found among the proteins up-regulated on 2-EHN, except for the last step of the β-oxidation which is catalyzed by a thiolase. Since three thiolase genes are present in the genomes of three related *Mycobacterium* species (Table S1), at least one thiolase is expected to be produced by strain IFP2173 grown on 2-EHN. Perhaps, the enzyme was synthesized in small amounts and passed undetected in our proteomic analysis.

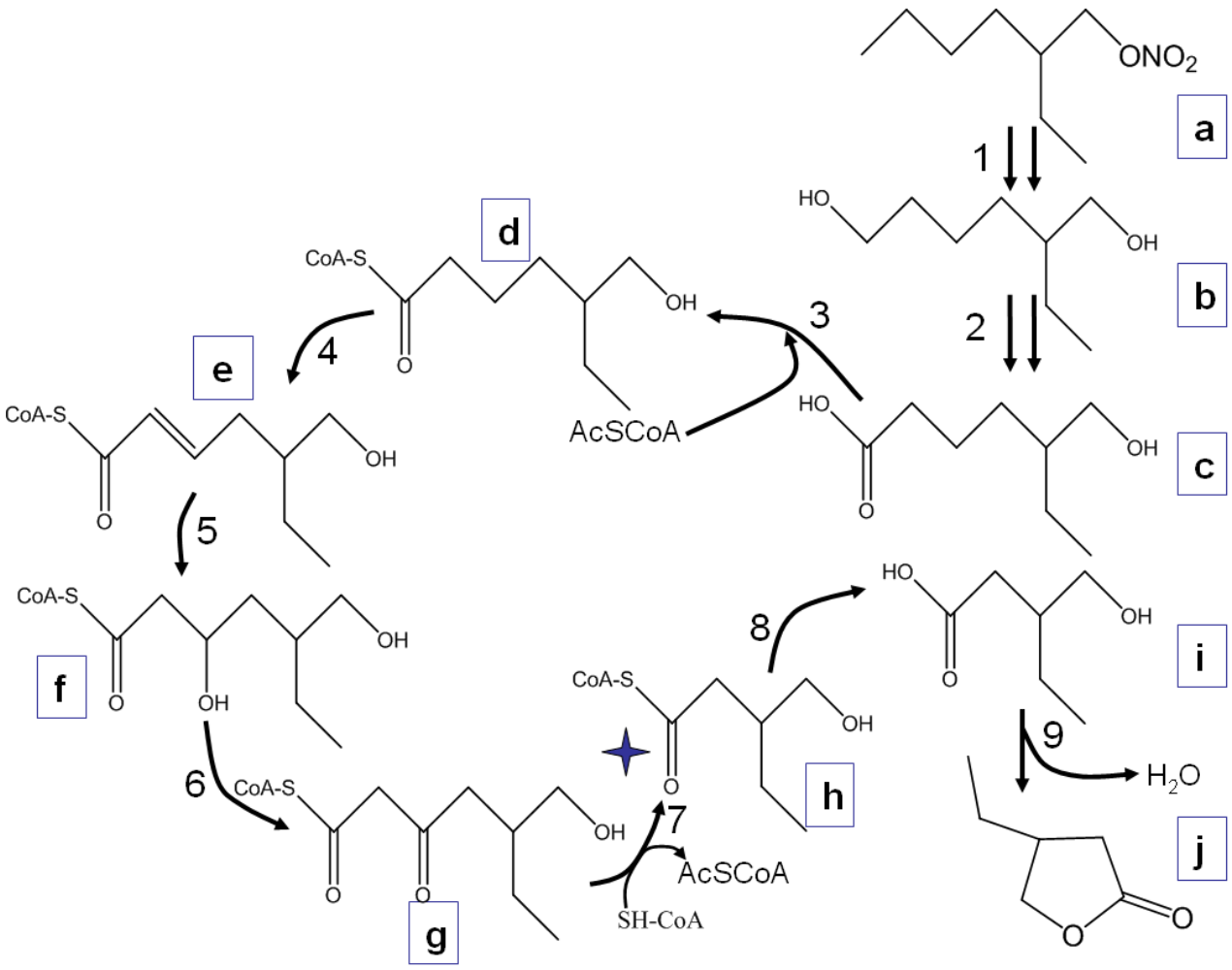

Figure 3: Proposed pathway and enzymes involved in 2-EHN degradation

Products: **a**: 2-ethylhexyl nitrate; **b**: 2-ethylpentane-1,5-diol; **c**: 5-(hydroxymethyl)heptanoic acid; **d**: 5-(hydroxymethyl)heptanoyl CoA; **e**: 5-(hydroxymethyl)heptan-2-enoyl CoA; **f**: 3-hydroxy-5-(hydroxymethyl)heptanoyl CoA; **g**: 5-(hydroxymethyl)-3-ketoheptanoyl CoA; **h**: 3-(hydroxymethyl)-pentanoyl CoA; **i**: 3-(hydroxymethyl)-pentanoic acid; **j**: 4-ethyldihydrofuran-2-(3H)one

Enzymes: **1**: alkane hydroxylase (AlkB1, AlkB2 or CYP153) and esterase, **2**: alcohol dehydrogenase and aldehyde dehydrogenases, **3** and **8**: acyl CoA acyltransferase, **4**: Acyl CoA dehydrogenase, **5**: Enoyl CoA hydratase, **6**: Hydroxyacyl CoA dehydrogenase, **7**: Thiolase, **9**: spontaneous cyclisation.

✦ Indicates that β-oxidation is blocked by the ethyl chain in β position.

We identified four alkane hydroxylases that might catalyze the first step in 2-EHN degradation *i.e*. the hydroxylation of the distal methyl group. Two enzymes are similar to the classical AlkB membrane-bound enzymes that were found to play a pivotal role in alkane degradation by *P. putida* GPo1 [25] and A. *borkumensis* SK2 [26]. The AlkB proteins were not detected in the membrane fraction of IFP 2173 in either growth conditions, even when searching the proteomic data for the expected peptides derived from their deduced protein sequence. Nevertheless, specific transcripts for the corresponding genes were found in cells grown on acetate or exposed to 2-EHN suggesting that the proteins were synthesized. Perhaps, the AlkB proteins were poorly solubilized in SDS or yielded few trypsic peptides detectable by LC-MS/MS. The two other plausible enzymes that can initiate 2-EHN degradation are soluble cytochromes P450 of the CYP153 subfamily. Although many genes encoding cytochromes P450 are present in the genomes of related *Mycobacterium* species (Table 2), sequence alignments indicated that none of the gene products was related to the CYP153 (data not shown). In addition, the CYP153 proteins identified in strain IFP 2173 were found to be mainly produced on 2-EHN, suggesting that at least one of these enzymes takes part in the degradation. CYP153 cytochromes hydroxylate linear or cyclic alkanes with medium chain length [24]. For instance, the CYP153A6 from *Mycobacterium* sp. XHN1500, which is the closest ortholog of the major CYP153 from IFP 2173, preferentially utilizes octane [24], but can also hydroxylate substrates with a bulky structure like limonene [24]. CYP153-like genes have been found in other Actinomycetes as well as in α and β-proteobacteria, and three groups were distinguished based on phylogenetic considerations [13]. Like other bacterial cytochromes, the CYP153 enzymes function with two electron carriers, a NAD(P)H-oxidoreductase and a ferredoxin. In this respect, a FAD-binding oxidoreductase that might be functionally associated with one of the CYP153 from IFP2173 has been identified by the 2D gel approach (Table 1), and four such reductases were detected by the whole proteome approach (Tables S2 & S3).

An esterase is also required to hydrolyze the nitro-ester bond of 2-EHN. A chloride peroxidase was found among the 2-EHN-induced proteins in the 2D gel analysis (A1T5E7), which showed the classical consensus sequence (G*X*S*X*G) typical for the active site of esterases. This type of enzyme is active on carboxylic esters [27], but it is unknown whether it could remove the nitro group of 2-EHN. Our proteomic analysis revealed that four other putative esterases up-regulated on 2-EHN can potentially catalyze this reaction (Tables S2 & S3). Since strain IFP 2173 can use 2-ethylhexanol and transform it to 2-

ethylhexanoic acid and 4-ethyltetrahydrofuran-2(3H)-one [17], it may be inferred that hydrolysis of the ester bond precedes the hydroxylation of the distal methyl group in the biodegradation pathway.

The subsequent step in the degradation pathway is the conversion of the 2-ethylpentane-1,5-diol to an aldehyde by an alcohol dehydrogenase (ADH). Three 38-kDa ADH were apparently associated to 2-EHN metabolism, which are predicted to have a zinc-binding domain and a GroES-like structure. A different and larger alcohol dehydrogenase (AlkJ; 61 kDa) is involved in alkane oxidation in *P. putida* GPo1. However, a deletion of *alkJ* did not affect alkane degradation, indicating that this reaction does not require a specific alcohol dehydrogenase [25].

A single 2-EHN-specific aldehyde dehydrogenase (ALDH) was detected by SDS-PAGE and peptide analysis. Examination of the sequence of the orthologous enzyme of strain PYR-1 (A1T1A6) showed that it might be composed of a catalytic domain and a LuxC-like domain [28]. Interestingly, the detected ALDH showed 39 % sequence identity with AlkH encoded by the *alk* operon, which is involved in alkane biodegradation in *P. putida* GPo1 [29].

The biosynthesis of multiple isoenzymes of the β-oxidation pathway in response to cell exposure to 2-EHN is intriguing. Many *Mycobacterium* species are known to thrive on alkanes, but the redundancy of β-oxidation enzymes had not been previously reported, although it could be predicted from the abundance of genes coding for such enzymes in available genome sequences of *Mycobacteria* (Table S1). In contrast, the hydrocarbonoclastic bacterium *A. borkumensis* SK2 produced relatively few specific enzymes when growing on alkanes [15]. Perhaps, the geater number of isoenzymes found in soil *Mycobacteria* reflects a more versatile metabolism, adapted to a diet made of diverse hydrocarbons present in their environment.

2-EHN is a derivate of a branched alkane, that is a type of organic compound known for its resistance to microbial degradation. Indeed, the size and position of the substituent and the degree of branching all affect microbial alkane utilization, by hindering β-oxidation. Some bacteria may overcome β-oxidation blockage such as *Pseudomonas citronellolis* which uses a β-carboxylase in citronellol degradation [30]. In the case of 2-EHN, the ethyl group in β position prevented its complete degradation by *M. austroafricanum* IFP 2173. To achieve the complete mineralization of 2-EHN, one strategy would consist to isolate a bacterium

able to metabolize the 4-EDF intermediate and then associate this isolate with strain IFP 2173 in co-cultures.

5. References

[1] Szybist, J. P., Boehman, A. L., Taylor, J. D., McCormick, R. L., Evaluation of formulation strategies to eliminate the biodiesel NOx effect. *Fuel Processing Technology* 2005, *86*, 1109-1126.

[2] Ban-Weiss, G. A., Chen, J. Y., Buchholz, B. A., Dibble, R. W., A numerical investigation into the anomalous slight NOx increase when burning biodiesel; A new (old) theory. *Fuel Processing Technology* 2007, *88*, 659-667.

[3] Gołofit-Szymczak, M., *International Journal Of Occupatioinal Safety dand Ergonomics*, CIOP PIP, Warszawa, Czerniakowska 2005.

[4] The American Chemistry Council Petroleum additives panel, 2006, p. 11.

[5] Solano-Serena, F., Nicolau, E., Favreau, G., Jouanneau, Y., Marchal, R., Biodegradability of 2-ethylhexyl nitrate (2-EHN), a cetane improver of diesel oil. *Biodegradation* 2008.

[6] Kim, Y. H., Engesser, K. H., Kim, S. J., Physiological, numerical and molecular characterization of alkyl ether-utilizing rhodococci. *Environ Microbiol* 2007, *9*, 1497-1510.

[7] Dean-Ross, D., Cerniglia, C. E., Degradation of pyrene by *Mycobacterium flavescens*. *Appl Microbiol Biotechnol* 1996, *46*, 307-312.

[8] Wang, R. F., Cao, W. W., Cerniglia, C. E., Phylogenetic analysis of polycyclic aromatic hydrocarbon degrading mycobacteria by 16S rRNA sequencing. *FEMS Microbiol Lett* 1995, *130*, 75-80.

[9] Francois, A., Mathis, H., Godefroy, D., Piveteau, P., *et al.*, Biodegradation of methyl tert-butyl ether and other fuel oxygenates by a new strain, *Mycobacterium austroafricanum* IFP 2012. *Appl Environ Microbiol* 2002, *68*, 2754-2762.

[10] Lopes Ferreira, N., Maciel, H., Mathis, H., Monot, F., *et al.*, Isolation and characterization of a new *Mycobacterium austroafricanum* strain, IFP 2015, growing on MTBE. *Appl Microbiol Biotechnol* 2006, *70*, 358-365.

[11] Lee, M., Kim, M. K., Kwon, M. J., Park, B. D., *et al.*, Effect of the synthesized mycolic acid on the biodegradation of diesel oil by *Gordonia nitida* strain LE31. *J Biosci Bioeng* 2005, *100*, 429-436.

[12] Sokolovska, I., Rozenberg, R., Riez, C., Rouxhet, P. G., *et al.*, Carbon source-induced modifications in the mycolic acid content and cell wall permeability of *Rhodococcus erythropolis* E1. *Appl Environ Microbiol* 2003, *69*, 7019-7027.

[13] van Beilen, J. B., Funhoff, E. G., Alkane hydroxylases involved in microbial alkane degradation. *Appl Microbiol Biotechnol* 2007, *74*, 13-21.

[14] Maier, T., Forster, H. H., Asperger, O., Hahn, U., Molecular characterization of the 56-kDa CYP153 from *Acinetobacter* sp. EB104. *Biochem Biophys Res Commun* 2001, *286*, 652-658.

[15] Sabirova, J. S., Ferrer, M., Regenhardt, D., Timmis, K. N., Golyshin, P. N., Proteomic insights into metabolic adaptations in *Alcanivorax borkumensis* induced by alkane utilization. *J Bacteriol* 2006, *188*, 3763-3773.

[16] Solano-Serena, F., Marchal, R., Heiss, S., Vandecasteele, J. P., Degradation of isooctane by *Mycobacterium austroafricanum* IFP 2173: growth and catabolic pathway. *J Appl Microbiol* 2004, *97*, 629-639.

[17] Nicolau, E., Kerhoas, L., Lettere, M., Jouanneau, Y., Marchal, R., Biodegradation of 2-ethylhexyl nitrate by *Mycobacterium austroafricanum* IFP 2173. *Appl Environ Microbiol* 2008, *74*, 6187-6193.

[18] Krivobok, S., Kuony, S., Meyer, C., Louwagie, M., *et al.*, Identification of pyrene-induced proteins in *Mycobacterium* sp. strain 6PY1: evidence for two ring-hydroxylating dioxygenases. *J Bacteriol* 2003, *185*, 3828-3841.

[19] Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, *227*, 680-685.

[20] Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 1988, *9*, 255-262.

[21] Schagger, H., Tricine-SDS-PAGE. *Nat Protoc* 2006, *1*, 16-22.

[22] Sambrook, J., Fritsch, E. F., Maniatis, T. (Eds.), *Molecular Cloning*, Cold Spring Harbor Laboratory Press 1989.

[23] van Beilen, J. B., Funhoff, E. G., van Loon, A., Just, A., *et al.*, Cytochrome P450 Alkane Hydroxylases of the CYP153 Family Are Common in Alkane-Degrading Eubacteria Lacking Integral Membrane Alkane Hydroxylases. *Appl Environ Microbiol* 2006, *72*, 59-65.

[24] van Beilen, J. B., Holtackers, R., Luscher, D., Bauer, U., *et al.*, Biocatalytic production of perillyl alcohol from limonene by using a novel *Mycobacterium* sp. cytochrome P450 alkane hydroxylase expressed in *Pseudomonas putida*. *Appl Environ Microbiol* 2005, *71*, 1737-1744.

[25] van Beilen, J. B., Eggink, G., Enequist, H., Bos, R., Witholt, B., DNA sequence determination and functional characterization of the OCT-plasmid-encoded alkJKL genes of *Pseudomonas oleovorans*. *Mol Microbiol* 1992, *6*, 3121-3136.

[26] Schneiker, S., Dos Santos, V. A., Bartels, D., Bekel, T., *et al.*, Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nat Biotechnol* 2006, *24*, 997-1004.

[27] Bornscheuer, U. T., Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev* 2002, *26*, 73-81.

[28] Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., *et al.*, CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* 2007, *35*, D237-240.

[29] van Beilen, J. B., Panke, S., Lucchini, S., Franchini, A. G., *et al.*, Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the alk genes. *Microbiology* 2001, *147*, 1621-1630.

[30] Fall, R. R., Brown, J. L., Schaeffer, T. L., Enzyme recruitment allows the biodegradation of recalcitrant branched hydrocarbons by *Pseudomonas citronellolis*. *Appl Environ Microbiol* 1979, *38*, 715-722.