# De novo structure determination of 3-((3-aminopropyl)amino)- 4-hydroxybenzoic acid, a novel and abundant metabolite in Acinetobacter baylyi ADP1

Marion Thomas, Lucille Stuani, Ekaterina Darii, Christophe Lechaplais, Emilie Pateau, Jean-Claude Tabet, Marcel Salanoubat, Pierre-Loïc Saaidi, Alain Perret

Metabolomics _####################_
https://doi.org/10.1007/s11306-019-1508-3

ORIGINAL ARTICLE

Check for updates

# De novo structure determination of 3-((3-aminopropyl)amino)-4-hydroxybenzoic acid, a novel and abundant metabolite in *Acinetobacter baylyi* ADP1

**Marion Thomas[1] · Lucille Stuani[2] · Ekaterina Darii[1] · Christophe Lechaplais[1] · Emilie Pateau[1] · Jean-Claude Tabet[3,4] · Marcel Salanoubat[1] · Pierre-Loïc Saaidi[1] · Alain Perret[1]**

Received: 4 October 2018 / Accepted: 7 March 2019


## Abstract

**Introduction** Metabolite identification remains a major bottleneck in the understanding of metabolism. Many metabolomics AQ1 studies end up with unknown compounds, leaving a landscape of metabolites and metabolic pathways to be unraveled. Therefore, identifying novel compounds within a metabolome is an entry point into the 'dark side' of metabolism.

**Objectives** This work aimed at elucidating the structure of a novel metabolite that was first detected in the soil bacterium AQ2 *Acinetobacter baylyi* ADP1 (ADP1).

**Methods** We used high resolution multi-stage tandem mass spectrometry for characterizing the metabolite within the metabolome. We purified the molecule for 1D- and 2D-NMR (1H, $^{13}$C, $^{1}$H-$^{1}$H-COSY, $^{1}$H-$^{13}$C-HSQC, $^{1}$H-$^{13}$C-HMBC and $^{1}$H-$^{15}$N-HMBC) analyses. Synthetic standards were chemically prepared from MS and NMR data interpretation.

**Results** We determined the de novo structure of a previously unreported metabolite: 3-((3-aminopropyl)amino)-4-hydroxybenzoic acid. The proposed structure was validated by comparison to a synthetic standard. With a concentration in the millimolar range, this compound appears as a major metabolite in ADP1, which we anticipate to participate to an unsuspected metabolic pathway. This novel metabolite was also detected in another γ-proteobacterium.

**Conclusion** Structure elucidation of this abundant and novel metabolite in ADP1 urges to decipher its biosynthetic pathway and cellular function.

**Keywords** Functional genomics · Bacterial metabolism · LC/MS-orbitrap · NMR · Structure elucidation

Marion Thomas and Lucille Stuani have contributed equally to this work.

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11306-019-1508-3) contains supplementary material, which is available to authorized users.

✉ Pierre-Loïc Saaidi
plsaaidi@genoscope.cns.fr

✉ Alain Perret
aperret@genoscope.cns.fr

[1] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

[2] INSERM, Institut National de la Santé et de la Recherche Médicale - CNRS - UPS - Centre de Recherche en Cancérologie de Toulouse (CRCT), Toulouse, France

[3] Sorbonne Université, UPMC Univ Paris 06, CNRS, Institut Parisien de Chimie Moléculaire, Paris, France

[4] CEA, iBiTec-S, SPI, LEMM, Gif-sur-Yvette, France

## 1 Introduction

Extensive and accurate knowledge of bacterial metabolism is critical for developing a comprehensive and detailed understanding of cellular physiology. This requires that, *in fine*, within a studied organism the function of all genes is established and the identity of all metabolites known. This goal is today probably out of reach given the complexity of the chemistry of life, and represents a long-term objective. On the one side 30–40% of genes of a typical genome remain unannotated or associated with a putative function (Chang et al. 2016; Galperin and Koonin 2004) and on the other side, many metabolites remain unidentified in most metabolomics studies (Bingol et al. 2016; Dias et al. 2016; Domingo-Almenara et al. 2018; Dunn et al. 2013b; Viant et al. 2017). Nuclear magnetic resonance (NMR) is currently the most powerful spectroscopic method for elucidating the chemical structure of organic compounds





| Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019 |
|---|---|---|---|---|

due to the development of high field magnets as well as capillary and cryogenic probes that significantly increased the sensitivity of NMR spectroscopy (Dias et al. 2016). However, this technique most preferably needs a pure sample in the upper μM range that requires laborious and time consuming work. Moreover, since annotating compounds to most likely structures using high-resolution mass spectrometry (HRMS) is a faster and feasible first step, electrospray-mass spectrometry (ESI-MS) and tandem mass spectrometry combined with liquid chromatographic (LC) separation is the most used technique for characterizing unknown compounds present in metabolomes. Nevertheless, many metabolomic studies end up with a mere list of mass-to-charge ratio (*m*/*z*) values and chromatographic retention times. As a consequence, a landscape of metabolites and metabolic pathways are yet to be discovered. In metabolomics, identification of metabolites remains one of the biggest bottlenecks (Dunn et al. 2013a; Kind and Fiehn 2010; Viant et al. 2017), but of importance, since identifying novel compounds within a metabolome is an entry point into the 'dark side' of bacterial metabolism. The use of HRMS instruments with mass accuracy in the low parts per million (ppm) range can provide the elemental composition of the detected ions. HRMS also provides a useful method for putative annotation querying databases (Kanehisa et al. 2012; Smith et al. 2005; Wishart et al. 2013) for metabolites whose molecular masses are within a specified mass deviation range. However, the majority of the observed peaks continue to remain unidentified or associated with multiple putative identifications. Without authentic compounds used for comparison of retention time and product ion spectra, the identity of a metabolite cannot be established. Structural information can also be obtained by performing MS/MS experiments, where the ion is fragmented and the *m*/*z* of the resulting fragment ions recorded. However, fragmentation interpretation of an unknown metabolite for its structural elucidation remains hampered by our incomplete understanding of the complex gas phase ion chemistry. Alternatively, these spectra can be matched with existing spectral databases for identity assignment or similarity search (Zhu et al. 2013). These databases, which are increasing in number and size (Horai et al. 2010; https://www.mzcloud.org; Kale et al. 2016; Sawada et al. 2012; Smith et al. 2005; Wishart et al. 2013) are very useful for identifying known compounds, but are far from being complete. Only a small proportion of all known metabolites are commercially available, leaving the majority of them not included in mass spectral libraries. It is estimated that less than 5% of all known metabolites are $MS^2$ spectra available (Frainay et al. 2018; Tsugawa et al. 2016). Another drawback of ESI-MS/MS experiments is that the tandem mass spectra generated exhibit high instrument-dependent variability which interferes with the constitution of universal databases as done with electron ionization mass spectrometry (Roux et al. 2011). Compared to $MS^2$ approaches, multistage MS ($MS^n$) fragmentation in which ions are specifically selected for sequential fragmentation, results in deeper and more detailed fragmentation pathways and provides more structural information (Rojas-Cherto et al. 2012; van der Hooft et al. 2011). In this case, again, complete identification is possible only when a fragmentation tree of the same compound is present in the database. Otherwise, putative candidate structures can be proposed using computer assisted structure elucidation (Peironcely et al. 2012; Wolf et al. 2010) but the identity of the unknown remains non-validated. Four levels of reporting metabolite annotation and identification are defined by the metabolomics standards initiative (Sumner et al. 2007). "Identification" is the most rigorous level (level 1), where two orthogonal properties of the metabolite are matched to the same properties of an authentic compound analyzed by the same method. Levels 2 and 3 provide annotation as metabolites or metabolite classes, respectively, without comparison to authentic compounds. Level 4 considers the metabolite as unidentified. Therefore, only information from rigorously 'level 1' identified metabolites can be used for investigating metabolic pathways and accessing other levels of biological hierarchy such as gene function annotation.

*Acinetobacter baylyi* ADP1 (ADP1) is a nutritionally versatile strictly aerobic bacterium (Young et al. 2005). Its extraordinary competence for natural transformation and the ease with which it can be genetically engineered (de Berardinis 2009; Metzgar et al. 2004) make ADP1 a key organism for the study of bacterial metabolism. We recently used metabolomics to reinvestigate quinate metabolism in ADP1, which was thoroughly studied for many years (Young et al. 2005). We reported that shifting the carbon source from succinate to quinate elicited a global metabolic reprogramming and led to the production of almost 30 unidentified and unexpected compounds that we anticipate to participate to unsuspected metabolic pathways (Stuani et al. 2014). Here, we report the complete de novo structure elucidation of one these unknown metabolites. We conducted its extensive LC/HRMS/$MS^n$ characterization within the metabolome and purified the molecule for 1D- and 2D-NMR (1H, $^{13}C$, $^{1}H$-$^{1}H$-COSY, $^{1}H$-$^{13}C$-HSQC, $^{1}H$-$^{13}C$-HMBC and $^{1}H$-$^{15}N$-HMBC) analyses. The putative structure deduced from MS and NMR data interpretation was eventually confirmed by means of chemical synthesis to be 3-((3-aminopropyl) amino)-4-hydroxybenzoic acid (APAH). This compound was discovered to be a major metabolite in ADP1 with intracellular concentration in the millimolar range. Its presence in another γ-proteobacterium such as *Pseudomonas putida* suggests a wider taxonomic distribution and urges to investigate its metabolic role.





Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019

## 2 Materials and methods

### 2.1 Chemicals and reagents

Chemicals, HPLC-grade solvents, ammonium acetate, ammonium carbonate, ammonium hydroxide and deuterated water were from Sigma-Aldrich. 5-((3-aminopropyl)amino)-2-hydroxybenzoic acid (SN0079), 4-((3-aminopropyl)amino)-2-hydroxybenzoic acid (SN0080), 2-((3-aminopropyl)amino)-5-hydroxybenzoic acid (SN0123), 4-((3-aminopropyl)amino)-3-hydroxybenzoic acid (SN0124) and 3-((3-aminopropyl)amino)-4-hydroxybenzoic acid (SN0140) were synthesized by Synthenova (http://www.synthenova.com).

### 2.2 Strains and medium

*Acinetobacter baylyi* ADP1 strain (DSM 24193) was provided by Dr. Nicholas Ornston (Yale University). ADP1, *Pseudomonas putida* KT2440 (ATCC 47054) and *Ralstonia eutropha* H16 (DSMZ 428) were routinely grown on medium for *Acinetobacter* (MA) minimal medium [31 mM $Na_2HPO_4$, 25 mM $KH_2PO_4$, 18 mM $NH_4Cl$, 41 µM nitrilotriacetic acid, 2 mM $MgSO_4$, 0.45 mM $CaCl_2$, 3 µM $FeCl_3$, 1 µM $MnCl_2$, 1 µM $ZnCl_2$, 0.3 µM ($CrCl_3$, $H_3BO_3$, $CoCl_2$, $CuCl_2$, $NiCl_2$, $Na_2MoO_2$, $Na_2SeO_3$)] supplemented with 20 mM (or 5 mM for protocatechuate) of the desired carbon source.

### 2.3 Metabolome preparation

#### 2.3.1 Analytical scale

Cells were grown in 5 ml at 30 °C. Metabolite extraction was adapted from the Metabolomics Service Protocols from the University of Glasgow (http://www.polyomics.gla.ac.uk/assets/downloads/MSMetabolomicsPrepCells-Aug2013.pdf). A saturated overnight culture was diluted in a fresh medium at an $OD_{600}=0.05$. 5 ml were transferred in a well of a 24 deep-well plate (Whatman; reference 7701-5110). The cells were further grown to an $OD_{600}$ between 0.6 and 0.8 in a shaking incubator (Climo Shaker ISF1-X Kühner). The plate was then centrifuged at 2700×*g* at 4 °C for 10 min and the supernatant removed. The cell pellet was resuspended in 200 µl of water/methanol/acetonitrile (1:3:1 ratio) and placed in a cold bath (−80 °C) composed of dry ice and ethanol. After cell freezing, the mixture was let at 25 °C to complete unfreezing of the cells. This procedure was repeated twice to completely break the cells. The metabolites were next rocked on a shaker for 1 h at 4 °C. The metabolites were centrifuged at 5000×*g* at 4 °C for 10 min. The supernatant was dried and stored at −80 °C. Before LC/MS analysis, the metabolites were suspended with 20 µl water and 42 µl 80% acetonitrile and 20% 10 mM ammonium carbonate (pH 9.9), centrifuged at 5000×*g* at 4 °C for 10 min. The supernatant was finally filtered on 0.22 µm (PTFE, Acroprep Advance, Pall).

#### 2.3.2 Preparative scale

For purification of metabolite of *m/z* 211 (M211), a saturated overnight culture was diluted in a fresh medium at an $OD_{600}=0.05$. 5 ml were distributed in each well of a 24 deep-well plate. Four plates were prepared per day. The cells were further grown to an $OD_{600}$ between 0.6 and 0.8 in a shaking incubator. The four plates were then centrifuged at 2700×*g* at 4 °C for 10 min and the supernatant removed. Cell pellets were resuspended in 200 µl of water/methanol/acetonitrile (1:3:1 ratio). The plates were placed in a cold bath (−80 °C) composed of dry ice and ethanol. After cell freezing, the plates were let at 25 °C to complete unfreezing of the cells. This procedure was repeated twice to completely break the cells. The plates were next rocked on a shaker for 1 h at 4 °C. The extracted metabolites were pooled in a 50 ml Falcon tube and lyophilized and stored at −80 °C. All this procedure was repeated three times, for a total of twelve 24 deep-well plates processed (1.5 l of cell culture). The metabolome was resuspended in 7 ml 80% acetonitrile and 20% 10 mM ammonium acetate (pH 6.8) before purification.

### 2.4 Metabolome derivatization

1,3-Diaminopropane (DAP) was labelled with benzoyl chloride for improving LC/MS detection (Aflaki et al. 2014). Dried metabolomes from 5 ml cultures were resuspended in 250 µl water and basified with 0.5 µl 2M NaOH. 250 µl of 4% benzoyl chloride in acetonitrile were added. After 15 min, the mixture was filtered on 0.22 µm (PTFE, Acroprep Advance, Pall) and injected for LC/MS analysis.

### 2.5 Analytical chromatography

Analyses were conducted using a Dionex Ultimate 3000 Rapid Separation LC (Thermo Fisher Scientific).

Chromatographic Method 1 used a ZIC®-*p*HILIC column (150×4.6 mm$^2$; 5 µm; Merck) and was previously reported (Stuani et al. 2014).

Chromatographic Method 2 was used for hydrogen/deuterium exchange (HDX) experiments, with deuterated ($D_2O$) mobile phase. Elution was carried out on a ZIC®-*p*HILIC column (100×2.1 mm$^2$; 5 µm; Merck) to minimize the use of heavy water. Column temperature was set to 15 °C using a mobile phase gradient with a flow rate of 0.2 ml/min. Mobile phase A consisted of 10 mM ammonium carbonate





Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019

in deuterated water and mobile phase B of acetonitrile. The gradient started at 80% B for 1 min, followed by a linear gradient to 40% B at 8 min and remained 4 min at 40% B. The system returned to the initial solvent composition in 3 min and re-equilibrated under these conditions for 15 min. 2.5 µl were injected. All experiments done on labelled samples with these new settings were replicated on light samples with the same chromatographic column for comparison.

Chromatographic Method 3 used a Syncronis HILIC column ($50 \times 2.1$ mm$^2$; 1.7 µm; Thermo Fischer Scientific) and was conducted at 30 °C with a mobile phase gradient at a flow rate of 0.3 ml/min. A consisted of 10 mM ammonium acetate (pH 6.8) and B of acetonitrile. The gradient started at 80% B for 0.3 min, followed by a linear gradient to 40% B at 1.8 min and remained 1.5 min at 40% B. The system returned to the initial solvent composition in 0.5 min and re-equilibrated under these conditions for 3.70 min. 2 µl were injected.

Chromatographic Method 4 used a ZIC®-*p*HILIC column ($150 \times 2.1$ mm$^2$; 5 µm; Merck) with a mobile phase gradient at a flow rate of 0.2 ml/min at 40 °C. A consisted of 10 mM ammonium carbonate with pH adjusted to 9.9 with $NH_4OH$ and B of acetonitrile. The gradient started at 80% B for 2 min, followed by a linear gradient to 40% B at 20 min and remained 3 min at 40% B. The system returned to the initial solvent composition in 2 min and re-equilibrated under these conditions for 12.2 min. 5 µl were injected.

Chromatographic Method 5 used a ZIC®-*p*HILIC ($50 \times 2.1$ mm$^2$; 5 µm; Merck) with a mobile phase gradient at a flow rate of 0.2 ml/min at 40 °C. A consisted of 10 mM ammonium carbonate with pH adjusted to 9.9 with $NH_4OH$ and B of acetonitrile. The gradient started at 80% B for 0.5 min, followed by a linear gradient to 40% B at 6 min and remained 1.5 min at 40% B. The system returned to the initial solvent composition in 1.5 min and re-equilibrated under these conditions for 4.3 min. 2 µl were injected.

Chromatographic Method 6 used an Acquity UPLC BEH C18 column ($100 \times 2.1$ mm$^2$; 1.7 µm; Waters) and was conducted at 35 °C with a mobile phase gradient at a flow rate of 0.3 ml/min. Mobile phase A consisted of $H_2O$ and mobile phase B consisted of methanol. The gradient started at 10% B for 1 min, followed by a linear gradient to 70% B at 19 min and remained 5 min at 70% B. The system returned to the initial solvent composition in 5 min and re-equilibrated under these conditions for 10 min. 5 µl were injected. For all chromatographic analysis, the autosampler was kept at 4 °C.

## 2.6 Metabolite purification

Metabolite purification was performed on a semi-preparative HPLC system Prominence LC-20AP (Shimadzu, Kyoto, Japan) fitted with a fraction collector FRC-10A (Shimadzu, Kyoto, Japan). The extracted metabolome was first dissolved and sonicated for 10 min at 4 °C in 6 ml 80% acetonitrile and 20% 10 mM ammonium acetate (pH 6.8). The sample was centrifuged at 4000×*g* for 10 min at 4 °C and the supernatant filtered on 0.22 µm (Millipore Millex-GV 13 mm). The filter was further washed with 1 ml of the same solution. The first purification step was conducted using a ZIC®-HILIC column ($250 \times 10$ mm$^2$; 5 µm; Merck). A mobile phase gradient was used with a flow rate of 4.6 ml/min. Mobile phase A consisted of 10 mM ammonium acetate (pH 6.8) and mobile phase B consisted of acetonitrile. The gradient started at 80% B for 2 min, followed by a linear gradient to 40% B at 22 min and remained 10 min at 40% B. The system returned to the initial solvent composition in 20 min and re-equilibrated under these conditions for 20 min. 200 µl were injected (33 injections of 200 µl were performed). The compound was monitored at 223 nm and collected in 2.3 ml fractions. Due to the weak UV signal of M211, the eluted fractions were analyzed for the presence of m M211 by LC/MS using Chromatographic Method 3. The fractions of interest were pooled and dried, yielding 1.1 mg of a yellow/pale brown solid residue. The partially purified compound was dissolved in 2 ml 72% acetonitrile and 28% 10 mM ammonium acetate (pH 6.8) and centrifuged at 4000×*g* for 10 min at 4 °C and the supernatant filtered on 0.22 µm (Millipore Millex-GV 13 mm). The filter was further washed with 1 ml of the same solution. The second purification step was conducted using the same ZIC®-HILIC column ($250 \times 10$ mm$^2$; 5 µm; Merck) with an isocratic separation (mobile phase: 72% acetonitrile and 28% 10 mM ammonium acetate, pH 6.8) at a flow rate of 4.6 ml/min. 200 µl were injected (13 injections of 200 µl were performed). The eluted 2.3 ml fractions were analyzed for the presence of M211 by LC/MS using Chromatographic Method 3. 100 µg of pure M211 were obtained as a pale yellow solid matter.

## 2.7 HRMS analyses

High-resolution measurements were obtained with a Velos Pro Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) fitted with a heated electrospray ionization source (HESI) operating in positive and negative ionization modes. In the positive ion mode, the ion spray (IS) was set to +3.5 kV and the capillary temperature at 275 °C. Sheath gas and auxiliary gas flow rates were set at 60% arbitrary units (a.u.) and 44 a.u., respectively. The S-lens RF level was set to 55%. In the negative ion mode, the IS was set to −4.0 kV and the S-lens RF level to 60. Mass spectra were acquired over an *m*/*z* range from *m*/*z* 50 up to *m*/*z* 1000 with the mass resolution set to 60,000 FWHM at *m*/*z* 400. LC/MS data were acquired in raw files and processed with the Qualbrowser module of Xcalibur 2.2 (Thermo Fisher Scientific) to access to elemental compositions. Fragmentation experiments were performed under collision induced dissociation





| Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019 |
|---|---|---|---|---|

(CID) and higher energy dissociation (HCD) conditions. For CID and HCD, normalized collision energies (NCE) of 22% and 28% in positive mode and 33% and 75% in negative mode, respectively were applied.

### 2.8 NMR analyses

Samples were dissolved in $D_2O$:$H_2O$ (10:90). NMR experiments were performed on a Bruker Advance 600-MHz NMR spectrometer equipped with a 1.7 mm cryogenic probe.

The $^1H$ NMR water signal from the polar fraction was suppressed by means of excitation sculpting (Bruker ZGESGP pulse program). Additional $^{13}C$ NMR spectra were acquired for each sample. $^1H$-$^1H$ COSY, $^1H$-$^{13}C$ HSQC, $^1H$-$^{13}C$ HMBC and $^1H$-$^{15}N$ HMBC experiments was only carried out for the purified metabolite sample.

Chemical shifts (expressed in ppm) of $^1H$ NMR spectra were referenced to the solvent peaks δ H 4.79 for $D_2O$.

3-((3-Aminopropyl)amino)-4-hydroxybenzoic acid (M211; APAH):

$^1H$ NMR ($H_2O/D_2O$ (90:10), 600 MHz): 7.22 (1H, d, $J=2.0$ Hz, $H_2$), 7.20 (1H, dd, $J=8.5$ and 2.0 Hz, $H_6$), 6.77 (1H, d, $J=8.5$ Hz, $H_5$), 3.20 (2H, t, $J=7.4$ Hz, $H_1'$), 3.03 (2H, t, $J=7.4$ Hz, $H_3'$), 1.92 (2H, quint, $J=7.4$ Hz, $H_2'$); $^{13}C$ NMR ($H_2O/D_2O$ (90:10), 150 MHz): 176.0 ($C_1''$), 148.5 ($C_4$), 136.2 ($C_3$), 128.6 ($C_1$), 121.7 ($C_6$), 114.5 ($C_5$), 114.5 ($C_2$), 41.6 ($C_1'$), 38.4 ($C_3'$), 26.8 ($C_2'$).

5-((3-Aminopropyl)amino)-2-hydroxybenzoic acid (SN0079):

$^1H$ NMR ($D_2O$, 600 MHz): 7.25 (1H, d, $J=2.9$ Hz), 6.96 (1H, dd, $J=8.7$ and 2.9 Hz), 6.84 (1H, d, $J=8.7$ Hz), 3.16 (2H, t, $J=7.1$ Hz), 3.09 (2H, t, $J=7.6$ Hz), 1.94 (2H, quint, $J=7.4$ Hz); $^{13}C$ NMR ($D_2O$, 150 MHz): 176.0, 153.8, 140.5, 123.0, 118.9, 117.6, 116.7, 43.0, 38.3, 26.9.

4-((3-Aminopropyl)amino)-2-hydroxybenzoic acid (SN0080):

$^1H$ NMR ($D_2O$, 600 MHz): 7.51 (1H, d, $J=8.6$ Hz), 6.2 (1H, dd, $J=8.6$ and 2.1 Hz), 6.07 (1H, d, $J=2.1$ Hz), 3.13 (2H, t, $J=6.7$ Hz), 2.96 (2H, t, $J=7.7$ Hz), 1.85 (2H, quint, $J=7.3$ Hz); $^{13}C$ NMR ($D_2O$, 150 MHz): 176.0, 161.6, 153.3, 131.9, 107.6, 105.7, 98.3, 40.0, 37.5, 26.5.

2-((3-Aminopropyl)amino)-5-hydroxybenzoic acid (SN0123):

$^1H$ NMR ($D_2O$, 600 MHz): 7.29 (1H, d, $J=3.0$ Hz), 6.96 (1H, dd, $J=8.8$ and 3.0 Hz), 6.86 (1H, d, $J=8.8$ Hz), 3.25 (2H, t, $J=6.8$ Hz), 3.10 (2H, t, $J=7.6$ Hz), 1.97 (2H, quint, $J=7.2$ Hz). $^{13}C$ NMR ($D_2O$, 150 MHz): 175.3, 147.1, 142.3, 123.0, 119.7, 117.2, 116.0, 41.7, 37.6, 26.3.

4-((3-Aminopropyl)amino)-3-hydroxybenzoic acid (SN0124):

$^1H$ NMR ($D_2O$, 600 MHz): 7.44 (1H, dd, $J=8.2$ and 1.5 Hz), 7.34 (1H, d, $J=1.6$ Hz), 6.79 (1H, d, $J=8.2$ Hz), 3.33 (2H, t, $J=6.8$ Hz), 3.12 (2H, t, $J=7.5$ Hz), 2.00 (2H, quint, $J=7.2$ Hz). $^{13}C$ NMR ($D_2O$, 150 MHz): 175.5, 143.0, 139.8, 125.3, 122.9, 115.1, 110.7, 40.2, 37.6, 26.2.

### 2.9 APAH quantification

A calibration line for APAH (3-((3-aminopropyl)amino)-4-hydroxybenzoic acid) was required to calculate its in vivo concentration. The calibration line was obtained using calibration solutions containing 0, 1, 2, 5, 10 and 19 µM of synthetic APAH spiked in metabolomes from succinate-grown cells (devoid of APAH), to take into account the matrix effect. These spiked metabolomes were analyzed on our Velos Pro Orbitrap Elite in the positive mode using Chromatographic Method 4. The peak areas from extracted ion chromatograms of APAH were integrated and the values used to build the calibration line. Next, to estimate APAH concentration in the cell, metabolomes from 13 independent 5 ml cultures of quinate-grown cells ($OD_{600}=0.47\pm0.02$) were used. Cell cultures (succinate and quinate) and metabolome preparation were conducted as described above. The lyophilized metabolomes were resuspended with 20 µl water and 42 µl 80% acetonitrile and 20% 10 mM ammonium carbonate (pH 9.9). Samples were filtered on 0.22 µm prior to injection.

## 3 Results and discussion

### 3.1 LC/HRMS/MS$^n$ analysis of an unknown metabolite of m/z 211

Among the compounds exclusively formed in quinate-grown cells, we considered here an ion that, analyzed with LC–HRMS, has a long retention time (14.3 min) on the hydrophilic zwitterionic ZIC®-*p*HILIC column (Chromatographic Method 1, see "Material and methods" section) and led to intense ionization signal in the positive mode at *m/z* 211.1074 (Fig. 1a, b). The compound was also detected in the negative mode at *m/z* 209.0934. These two *m/z* values were anticipated to be the protonated $[M+H]^+$ and deprotonated form $[M-H]^-$ of the same molecule, respectively. The monoisotopic mass of the protonated form of this unknown metabolite, which we called M211 (metabolite of *m/z* 211), was consistent with an elemental composition of $C_{10}H_{15}O_3N_2^+$ (− 1.6 ppm off the theoretical mass) and with its first natural isotope abundance (M+1 = 11.1%; Fig. 1b). Elemental composition was validated by applying the Seven Golden Rules (Kind and Fiehn 2007). The long retention time of M211, associated with detection in both ionization AQ4 modes suggest a very polar structure carrying acidic and basic functional groups. Photodiode array (PDA) detection indicated UV absorption (Fig. 1c) with maxima at 224, 259,





| Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019 |
|---|---|---|---|---|

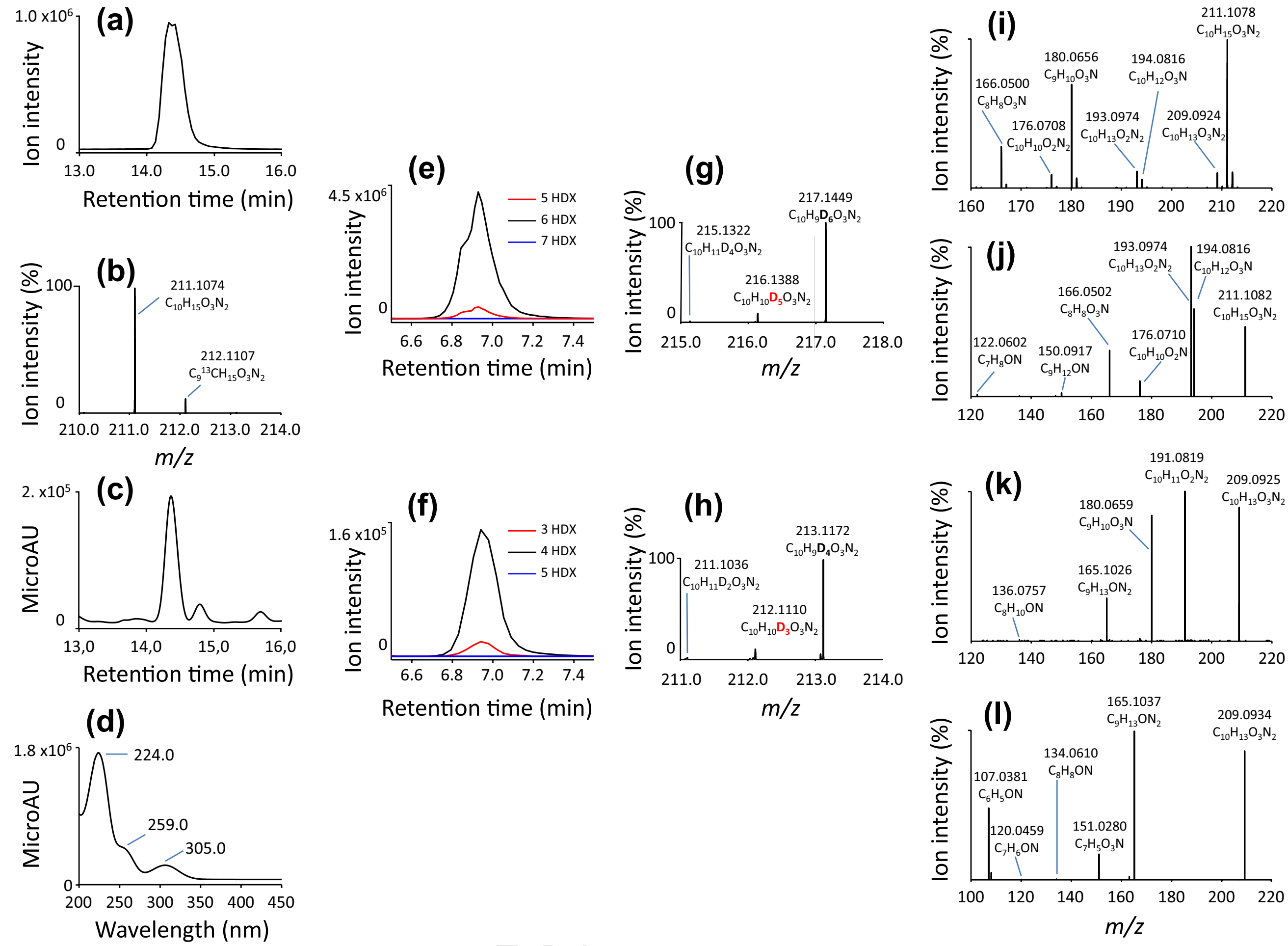


**Fig. 1** LC/MS analysis of M211 in the metabolome of *Acinetobacter baylyi* ADP1. **a** EIC corresponds to the protonated metabolite molecule $[M+H]^+$ at *m/z* 211.1077 in a metabolome from ADP1 quinate-grown cells (5 ppm accuracy). **b** Zoom on the mass spectrum of M211 (close up) in the positive mode. **c** LC/PDA chromatogram of M211. **d** UV–visible spectrum of M211. **e** EICs of deuterated species of M211 after HDX in the positive ionization mode at *m/z* 216.1391 (red line), 217.1454 (black line) and 218.1487 (blue line) at 5 ppm accuracy. **f** EICs of deuterated species of M211 after HDX in the negative ionization mode at *m/z* 212.1120 (red line), 213.1187 (black line) and 214.1246 (blue line) at 5 ppm accuracy. **g** Zoom on the mass spectrum of the deuterated forms of M211 in the positive ionization mode. **h** Zoom on the mass spectrum of the deuterated forms of M211 in the negative ionization mode. **i** HRMS of M211 in the positive mode. **j** HCD spectrum of M211 in the positive mode (28% NCE). **k** HCD spectrum of the oxidized form of M211 in the positive mode (*m/z* 209.0921) at 28% NCE. **l** HCD spectrum of M211 in the negative mode (*m/z* 209.0932) at 75% NCE



and 302 nm (Fig. 1d) that are evocative of an aromatic moiety (Singh et al. 2007).

Accurate mass and elemental composition of M211 matched to only two compounds present in Metlin and KEGG databases: aprobarbital and 2,6-dihydropseudonicotine. These molecules harbor 2 and 3 mobile protons, respectively (Fig. S1). To check whether these molecules could correspond to M211, we determined its mobile proton number. This was established in positive and negative ionization modes using in-solution hydrogen/deuterium exchange (HDX) approach (Chromatographic Method 2). Six mobile protons were observed in the positive mode and four in the negative mode; i.e. five mobile protons for the neutral form of the molecule (Fig. 1e–h). This result invalidated the structures proposed by the databases and confirmed that M211 was a previously unreported metabolite.

Mass spectra recorded in the positive mode displayed M211 and a series of other ions with the same extracted ion chromatogram (EIC) shape: a major ion at *m/z* 180 and minor ones at *m/z* 166, 176, 193, 194 and 209 (Fig. 1i). This latter probably originated from in-source oxidation of M211. The similar chromatographic shape of these ions suggests they could be generated by in-source dissociation of the precursor ion at *m/z* 211 (Fig. S2a).

In the negative mode M211 was detected at *m/z* 209. Two others weakly abundant ions at *m/z* 165 and *m/z* 107 observed in the mass spectrum displayed similar



| Journal : **Large 11306** | Article No : **1508** | Pages : **13** | MS Code : **MEBO-D-18-00279** | Dispatch : **11-3-2019** |
|---|---|---|---|---|

chromatographic shapes, and originated from in-source dissociation of *m/z* 209.

Fragmentation spectra were recorded under collision induced dissociation (CID) and higher energy dissociation (HCD) conditions. Under HCD, small size fragment ions become significantly more abundant. In the positive mode (Fig. 1j), fragmentation spectra display two main product ions: a major one at *m/z* 193 (water loss) and a slightly less abundant one at *m/z* 194 ($NH_3$ loss). Less abundant *m/z* 176 and *m/z* 166 product ions are related to the consecutive losses of [$NH_3 + H_2O$] and the release of [$C_2H_7N$], respectively. The neutral [$C_2H_7N$] moiety formally corresponds to saturated alkyl amine. The *m/z* 166 product ion was also formed during consecutive fragmentation of the *m/z* 194 precursor (sequential $MS^3$ experiments), as illustrated in Fig. S2b. This ion is related to the loss of $C_2H_4$ from *m/z* 194 ion produced by the loss of ammonium from *m/z* 211 precursor ion. Therefore, [$C_2H_7N$] probably corresponds to ethylamine rather than dimethylamine. HCD fragmentation also yielded very weak signals at *m/z* 150 and *m/z* 122, which likely originated from consecutive fragmentation of *m/z* 194 by the loss of $CO_2$ and [$CO_2 + C_2H_4$], respectively. These ions were also observed in the sequential $MS^3$ experiments of the selected *m/z* 194 (Fig. S2b). All these fragment ions were present in the mass spectrum, too. Ion *m/z* 180 was not detected in the fragmentation spectrum of M211, using either CID or HCD. However, it was very abundant in the fragmentation spectrum of its oxidized form at *m/z* 209 (Fig. 1k), indicating that *m/z* 180 observed in the mass spectrum of M211 originated mainly from in-source dissociation of *m/z* 209. In addition, the number of mobile protons observed on the one side for M211 (5 for the neutral species) and on the other side for *m/z* 209 (3 for the neutral form) and *m/z* 180 (mostly 2 and in a much lesser extent 3 for the neutral form) are in agreement with the loss of $CH_2NH$ (1 mobile proton) from *m/z* 209 rather than the release of $CH_3NH_2$ (2 mobile protons) from M211 (Fig. S2c–f).

In the negative mode, HCD spectra of the selected ion *m/z* 209 yielded a main product ion at *m/z*165 ($CO_2$ loss) along with two less abundant radical ions at *m/z* 107 and *m/z* 151 (Fig. 1l). These latter were formed after the loss of the radical ion $C_3H_8N\cdot$, either from *m/z* 165 (consecutive loss, which is also observed in $MS^3$ experiments on the selected precursor *m/z* 165 ion) or from *m/z* 209, respectively.

The following assumptions can be made from the fragmentation spectra: in the negative mode, a major loss of $CO_2$ is characteristic of the presence of a carboxylic group. In addition, the formation of an odd electron product ion such as *m/z* 107 $[C_6H_5ON]^{-\cdot}$ with a degree of unsaturation of 5 (5.5 for the parent ion *m/z* 209) together with UV data (Fig. 1a, b) suggests the presence of an aromatic core. According to its elemental composition, this relatively stable radical anion should represent a phenol derivative (Afonso et al. 2010).

Generally, an abundant loss of $NH_3$ in the positive mode is due to the presence of a primary aliphatic amino group. Here, a single ammonium loss suggests the presence of a single primary aliphatic amine in the unknown structure (Levsen et al. 2007). Therefore, taking into account the total number of mobile protons (five for the neutral species), a structure that involves N-containing heterocycle should be excluded. A loss of $H_2O$ in the positive mode can be related to the presence of either a carboxylic group or an aliphatic OH group (Levsen et al. 2007). Considering on the one hand a single water loss, the presence of one carboxylic group (two O-atoms), and on the other hand a total number of three oxygen atoms and the previously discussed existence of an aromatic (phenol) core, we anticipate a phenolic group rather than an aliphatic hydroxyl group. Thus, taking into account five mobile protons for the uncharged molecule, one for the carboxylic group, two for the primary amino group and one for the phenolic hydroxyl group, the molecule should involve a secondary amino group with one more mobile proton.

In conclusion, LC–HRMS/$MS^n$ data suggest that the unknown metabolite contains an aromatic moiety, possibly a phenol ring, a carboxylic group along with primary and secondary amines, but de novo interpretation of mass spectrometry data did not allow to completely elucidate its chemical structure.

## 3.2 NMR analysis of purified M211

NMR was considered to achieve the complete structural elucidation of M211. The metabolite was purified from 1.5 l of bacterial culture in MA medium containing 20 mM quinate as the carbon source. Purification was carried out using HILIC-phase chromatography and led to 100 µg of a pale yellow solid (see "Material and methods" section). Extensive 1D- and 2D-NMR analyses ($^1H$, $^{13}C$, $^1H$-$^1H$-COSY, $^1H$-$^{13}C$-HSQC, $^1H$-$^{13}C$-HMBC and $^1H$-$^{15}N$-HMBC) were performed (see Table 1 and Dataset_S1). The $^1H$-NMR spectrum clearly confirmed the presence of an aromatic ring bearing three protons. Their multiplicity (Table 1; Fig. 2a) indicated a 1,2,4-trisubstitution pattern (Fig. 2b). The set of three aliphatic signals, integrating each for two protons, suggested a propylene moiety further confirmed by several cross-correlations peaks visible in COSY spectrum (Fig. 2c, d). Combining $^1H$ and $^{13}C$ chemical shifts of aliphatic signals allowed us to conclude to a 1,3-diaminopropyl substructure, in agreement with HRMS/$MS^n$ conclusions that predicted the presence of primary and secondary amines (Fig. 2d). Cross correlation peak between proton $H_{1'}$ and carbon $C_3$ in HMBC spectrum (Fig. 2e) unambiguously demonstrated that the secondary amine is responsible for the





| Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019 |
|---|---|---|---|---|

**Table 1** $^1$H- and $^{13}$C-NMR spectral data of m/z 211 [600/150 MHz, $H_2O/D_2O$ 90:10, δ (ppm), J (Hz)]

| Proton | Chemical shifts $d_H$ | integration | multiplicity | Coupling constants (Hz) | COSY signals | $^1$H-$^{13}$C (HSQC, $^1$J) | $^1$H-$^{13}$C (HMBC, $^3$J,$^4$J) | $^1$H-$^{15}$N (HMBC, $^3$J,$^4$J) |
|---|---|---|---|---|---|---|---|---|
| 1 | – | – | – | – | | 128.6 | – | – |
| 2 | 7.22 | 1 | d | 1.95 | H6 | 114.5 | C6<br>C4<br>C1″ | $N_\alpha$ |
| 3 | – | – | – | – | | 136.2 | – | – |
| 4 | – | – | – | – | | 148.5 | – | – |
| 5 | 6.77 | 1 | d | 8.50 | H6 | 114.5 | C1<br>C3<br>C4 | $N_\alpha$ |
| 6 | 7.20 | 1 | dd | 8.50<br>1.95 | H5<br>H2 | 121.7 | C2 or C5<br>C4<br>C1″ | – |
| 1′ | 3.20 | 2 | t | 7.40 | H2′ | 41.6 | C3′<br>C2′<br>C3 | – |
| 2′ | 1.92 | 3 | q | 7.40<br>7.40 | H1′<br>H3′ | 26.8 | C1′<br>C3′ | $N_\alpha$<br>$N_\beta$ |
| 3′ | 3.03 | 2 | t | 7.40 | H2′ | 38.4 | C1′<br>C2′ | – |
| 1″ | – | – | – | – | – | 176 | – | – |



linkage of the aliphatic moiety to the aromatic ring. Characteristic chemical shift values of 147 and 176 ppm for carbon $C_4$ and $C_{1''}$ confirmed the presence of phenolic and carboxylic moieties (see Table 1 and Dataset S1). HMBC experiments indicated that the carboxylic function is indeed branched onto the aromatic system (Fig. 2f). Expected labile protons for amine, phenol and carboxylic acid groups could not be observed in $^1$H NMR neither in organic solvents (dimethylformamide-d6, acetonitrile-d3) nor in aqueous solutions (90/10 $H_2O/D_2O$ and 90/10 $H_2O/D_2O$ + 1% trifluoroacetic acid). Another difficulty arose from the superposition of two signals in $^{13}$C NMR spectrum at 114.5 ppm (Fig. 2g). Finally, we investigated the relative position of aromatic substituents using $^1$H-$^{13}$C-HMBC and $^1$H-$^{15}$N-HMBC experiments. Unfortunately, observed long-ranged $^1$H-$^{13}$C and $^1$H-$^{15}$N correlations (Fig. 2f, h) did not allow us to draw any firm conclusion. Indeed, $^2J(^1H–^{13}C)$, $^3J(^1H–^{13}C)$ and $^4J(^1H–^{13}C)$ on the one side and $^2J(^1H–^{15}N)$, $^3J(^1H–^{15}N)$ and $^4J(^1H–^{15}N)$ on the other side are known to lead to coupling values of similar frequency in aromatic systems (Martin and Williams 2008; Martin and Hadden 2000). At that stage, the complete structure elucidation of M211 was still not achieved as six regioisomers remained plausible (Fig. 3).

### 3.3 Structure determination of M211 by comparison to the synthetic standards

To help raising the ambiguity, five out of the six possible compounds were synthesized: 5-((3-aminopropyl)amino)-2-hydroxybenzoic acid (SN0079), 4-((3-aminopropyl)amino)-2-hydroxybenzoic acid (SN0080), 2-((3-aminopropyl)amino)-5-hydroxybenzoic acid (SN0123), 4-((3-aminopropyl)amino)-3-hydroxybenzoic acid (SN0124) and 3-((3-aminopropyl)amino)-4-hydroxybenzoic acid (SN0140). $^1$H and $^{13}$C NMR analyses were then recorded in the same manner as for the purified metabolite (Dataset_S1). Data clearly showed a very high similarity between M211 and SN0140 in detriment of all other structures. The observed slight $^1$H chemical shift difference between M211 and SN0140 probably originated from the elevated number of protonable functions highly sensitive to pH and/or to concentration due to inter-/intra-Hydrogen bonding (Fig. S3). To confirm the structural assignment, the standards along with a metabolome from ADP1 quinate-grown cells were further analyzed by LC/MS (Chromatographic Method 1). According to comparison of its retention time and MS$^2$ spectrum with the reference compounds, M211 was eventually unambiguously identified as 3-((3-aminopropyl)

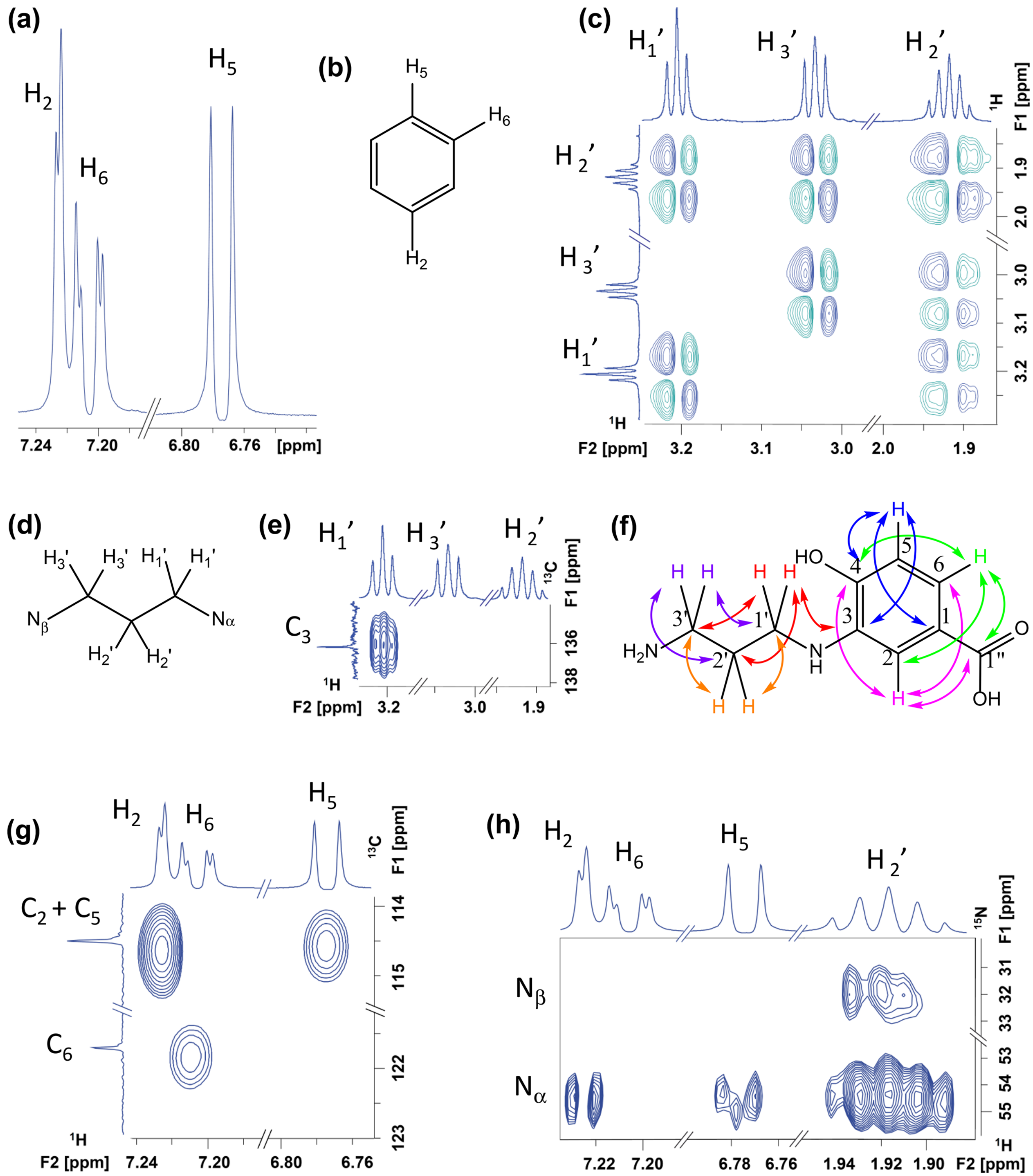


**Fig. 2** NMR analysis of purified M211. **a** Aromatic region of $^{1}$H-NMR spectrum. **b** Proposed relative positions of aromatic protons based on $^{1}$H-NMR analysis. **c** Aliphatic region of $^{1}$H–$^{1}$H COSY spectrum. **d** Proposed 1,3-diaminopropyl side chain based on $^{1}$H, $^{13}$C and $^{1}$H–$^{1}$H COSY NMR analyses. e Extract of $^{1}$H–$^{13}$C HMBC NMR spectrum showing $C_3$–$H_1{}'$ correlation. f $^{1}$H–$^{13}$C HMBC correlations placed on the proposed M211 structure. **g** Extract of $^{1}$H–$^{13}$C HSQC NMR spectrum showing the superposition of two carbons on $^{13}$C-NMR spectrum. **h** Extract of $^{1}$H–$^{15}$N HMBC NMR spectrum showing all cross correlations between $^{1}$H and $^{15}$N (natural abundance)



| Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019 |
|---|---|---|---|---|

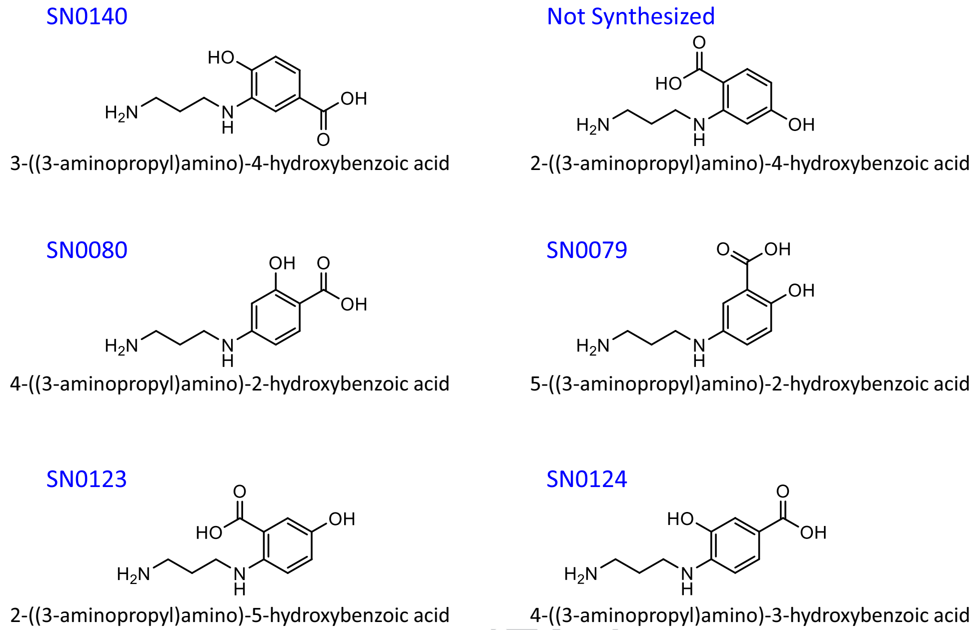


**Fig. 3** Possible structures for M211. Synthesized compounds are indicated by their name of manufacturing

amino)-4-hydroxybenzoic acid (APAH), as illustrated in Fig. S4.

## 3.4 Quantification of APAH in ADP1

In vivo quantification of APAH was conducted by LC–HRMS analysis using synthetic APAH. The calibration line (Chromatographic Method 4) was obtained by spiking metabolomes of succinate-grown cells (that are devoid of APAH) with known amounts of synthetic APAH, to overcome matrix effects. APAH was quantified according to its EIC area in metabolomes from independent quinate-grown cultures (Fig. S5). Assuming the same internal volume 2.3 $10^{-9}$ µl for ADP1 as for *E. coli* (Bazile et al. 1992), we estimated the cellular concentration of APAH to $7 \pm 4$ mM. APAH is thus present in ADP1 at a high abundance, which could affect osmolality. To draw a comparison, in *E. coli* the concentration of 70% of quantified metabolites is below the millimolar range (Bennett et al. 2009). Its abundance should be related to its role in the cell, as low metabolite concentrations are favorable for avoiding osmotic stress and disadvantageous spontaneous reactions (Bennett et al. 2009). Data in Fig. S5 indicated a wide variation of APAH in the metabolomes (between 2 and 19 mM). This unexplained variability cannot be attributed to incorrect handling of the cells during metabolomics sampling. Here, before metabolite extraction, culture centrifugation may have yield a cellular stress that caused the observed variation in APAH concentrations. However, in our previous study in which rapid cell quenching was combined with cell inactivation to 'freeze' the microbial metabolism, peak areas of APAH varied in a similar manner sevenfold (Stuani et al. 2014, Online Resource 2, Table S-6).

## 3.5 Hypotheses for APAH biosynthesis

APAH is formed in quinate-grown cells but not in succinate grown-cells (Stuani et al. 2014, Online Resource 2, Table S-6). Quinate is metabolized through the β-ketoadipate pathway (Young et al. 2005). Most aromatic compounds are mainly degraded by this pathway through the formation of catechol (catechol branch) or protocatechuate (protocatechuate branch). From catechol and protocatechuate, a parallel but separate branch converts them into succinyl-CoA and acetyl-CoA which can enter central metabolism through the TCA cycle (Fig. S6). Quinate is metabolized via the protocatechuate branch.

From a structural point of view, APAH is evocative of two metabolites: protocatechuate on the one side and 1,3-diaminopropane (DAP) on the other side. DAP is presented as the major polyamine in the genus *Acinetobacter*, with concentrations up to 2.5 µmol/g wet cell (Hamana and Matsuzaki 1992). We hypothesized that APAH may result from an enzymatic condensation of these two metabolites. Both protocatechuate and DAP are indeed present in ADP1 quinate-grown cells (Fig. 4). To gain insight in APAH biosynthesis, we explored the feasibility of a spontaneous reaction between protocatechuate and DAP. We mixed a 20 mM aqueous solution of protocatechuate with a 500 mM aqueous solution of DAP hydrochloride at a molar ratio of 1:2 (pH 7). After 30 min, we diluted twofold the solution in 80% acetonitrile and 20% 10 mM ammonium carbonate (pH 9.9) before LC/MS analysis (Chromatographic Method 1).





| Journal : Large 11306 | Article No : 1508 | Pages : 13 | MS Code : MEBO-D-18-00279 | Dispatch : 11-3-2019 |
|---|---|---|---|---|

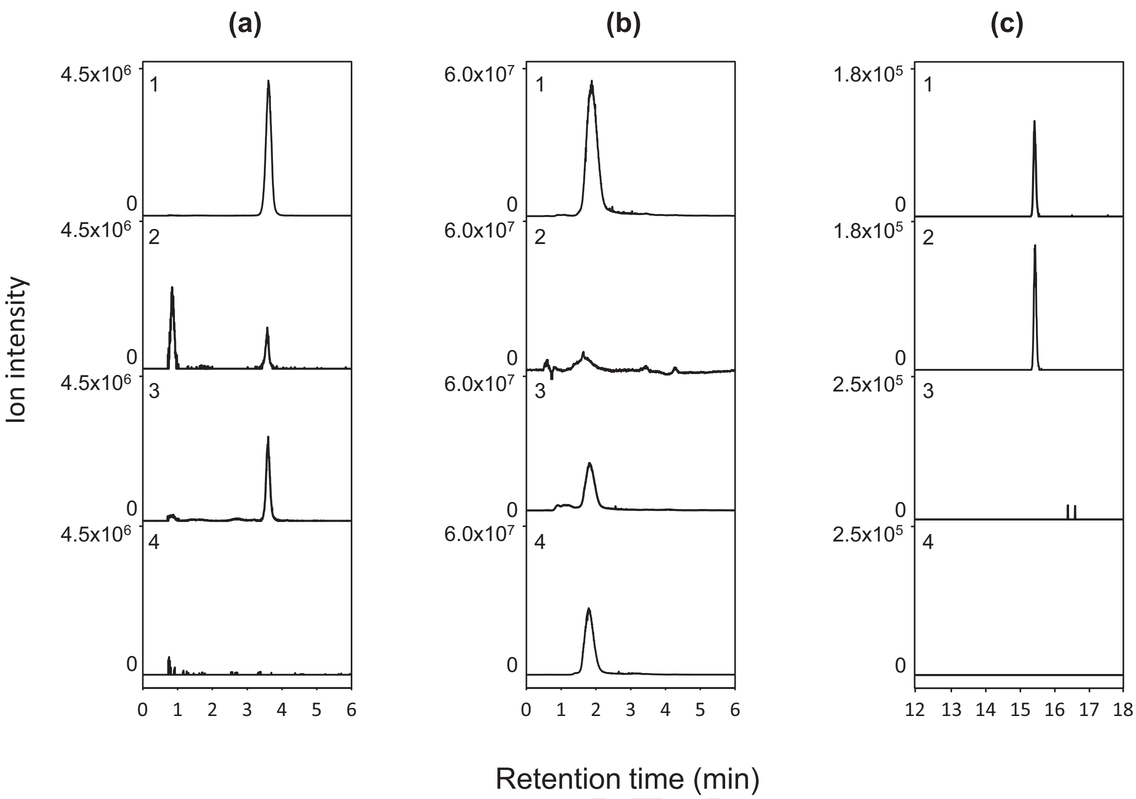


**Fig. 4** Screening of different metabolomes for the presence of M211, protocatechuate and 1,3-diaminopropane. **a** EICs correspond to the ($[M+H]^+$) of *m/z* 211.1077 in the positive mode. **b** EICs correspond to the ($[M-H]^-$) of protocatechuate (*m/z* 153.0193) in the negative mode. **c** EICs correspond to the ($[M+H]^+$) of the product of conjugation of 1,3-diaminopropane with two benzoyl groups in the positive mode (*m/z* 283.1441). All chromatograms were extracted at 5 ppm accuracy. Metabolomes were prepared from ADP1 quinate-grown cells (1), ADP1 benzyl alcohol-grown cells (2), *P. putida* KT4440 quinate grown cells (3) and *R. eutropha* H16 protocatechuate grown cells (4). Chromatograms in a2, a4 and b2 were magnified 100-fold. Chromatograms in a3, c3 and c4 were magnified tenfold



Two peaks of similar intensities were observed at *m/z* 211 (positive mode) consistent with the formation of two isomers. According to their retention times and fragmentation patterns, they corresponded to APAH and 4-((3-aminopropyl)amino)-3-hydroxybenzoic acid (SN0124), indicating that DAP can react with both hydroxyl groups of protocatechuate (Fig. S7). The reaction yield for APAH was very low (0.06%) and its concentration was estimated to be only 0.3 µM. These data thus indicate that the reaction, thermodynamically feasible, could be enzyme-catalyzed. According to this postulate, the formation of APAH should be strictly correlated to the presence of both DAP and protocatechuate in the cell.

We thus examined whether the presence of APAH was restricted to quinate metabolism solely or could be also linked to the catechol branch. To this end, we analyzed the metabolomes of ADP1 grown on benzyl alcohol, a compound whose degradation feeds into the catechol branch (Fig. S6). Data analysis revealed that APAH is formed in these conditions, albeit more weakly than in quinate-grown cells (Fig. 4a), but in quantity that can nonetheless be compared to that obtained in protocatechuate-grown cells (Fig. S8). However, although DAP was present, protocatechuate was not detected in ADP1 growing on benzyl alcohol (Fig. 4b). The weak signal detected in the metabolome of benzyl alcohol-grown cells did not match with the retention time of protocatechuate and displayed a different $MS^2$ spectrum. This suggested that APAH is not formed from protocatechuate and indicates that APAH biosynthesis is neither restricted to quinate degradation nor to the protocatechuate branch of the β-ketoadipate pathway. We next wondered about the presence of APAH in other organisms harboring the quinate degradation pathway. Genome context analyses conducted on the MicroScope platform (https://www.genoscope.cns.fr/agc/microscope/home/index.php) showed that clusters of genes orthologous to *pca* operon encoding the enzymes for protocatechuate degradation (Siehler et al. 2007) are present in the genome of organisms such as *Pseudomonas putida* KT4440 and *Rasltonia eutropha* H16. Despite the absence of genes orthologous to *qui* operon in *P. putida* for converting quinate to protocatechuate, this organism was nevertheless reported to grow with quinate as the sole carbon source (Jimenez et al. 2002). Regarding *R. eutropha* H16, its genome does not harbor the *qui* operon either and cannot grow on quinate. We thus analyzed the metabolomes of *P. putida* and *R. eutropha* grown on quinate and protocatechuate, respectively, for the presence of APAH. We detected APAH in *P. putida* but not *in R. eutropha* (Fig. 4a). This result emphasizes that APAH is not restricted to ADP1 and is produced at least in another γ-proteobacterium. On the other hand, its absence in *R. eutropha* eliminates a strict association between its synthesis and quinate/protocatechuate catabolism. Moreover, DAP was not observed in *P. putida* (Fig. 4c). Together, these results suggest that neither protocatechuate nor DAP is a biosynthetic precursor of APAH. APAH origin and function remain unknown. The metabolite is not excreted (Fig. S9) as APAH could not be detected in the medium of a culture at $OD_{600} = 0.8$, although APAH synthesis is maximal at this cell density (Fig. S10). On the contrary, protocatechuate, known to be present in the medium of quinate-degrading cells (D'Argenio et al. 1999) is indeed excreted (Fig. S9). Moreover, APAH does not behave as secondary metabolite: it is not formed at high cell density or during a decrease of the growth rate, or after exhaustion of the carbon source (Fig. S10).

## 4 Conclusion

We report here the de novo complete structural elucidation of a novel and abundant metabolite in *Acinetobacter baylyi* ADP1, APAH, a compound that is not restricted to ADP1.

To meet the requirement of 'level 1 identification' (according to the Metabolomics Standard Initiative), we conducted an extensive analytical study that combined NMR and LC/HRMS/$MS^n$ techniques. Finally, its identity was established according to comparison of its retention time, $MS^2$ and $^1H$ NMR spectra with a synthetic standard. This work illustrates the significant effort required to unambiguously identify a previously unreported metabolite, which still remains orphan of enzyme/gene.

Interpretation of experimental MS and NMR data for metabolite identification is crucial for metabolism interpretation, but cannot be scaled up to adapt with the hundreds or thousands of compounds present in a metabolome. Improvement of tandem mass spectral libraries and MS/MS scoring algorithms are needed (Kind et al. 2018) for faster compound identification and biochemical landscape covering.

To gain insight from APAH biosynthesis, we are actually conducting a LC/HRMS-based screening of the collection of 2600 single-gene deletion mutants of ADP1 (de Berardinis et al. 2008). Detecting the mutant metabolomes devoid of APAH shall inform us on the genes involved in its formation and led us to start reconstituting APAH synthetic pathway.

**Acknowledgements** We thank Jean-François Gallard from Institut de Chimie des Substances Naturelles (CNRS, Gif-sur-Yvette, France) for his expertise in NMR. The authors would like to thank Olek Maciejak and Marie-Jeanne Clément (SABNP, INSERM U 1204—Université d'Evry Val-d'Essonne, Université Paris-Saclay, France) for assistance in $^1H$ and $^{13}C$ NMR experiments. We are grateful to Cécile Robert-Ansart who was involved as part of a student internship in the project.

**Author contributions** MS and AP conceived the study. MT, LS, ED, CL and EP conducted experiments. MT, LS, ED, CL, EP, JCT, PLS and AP analyzed data. MT, ED, JCT, PLS and AP wrote the manuscript. All authors read and approved the manuscript.

**Funding** This work was supported by grants from the Commissariat à l'Energie Atomique et aux Energies Alternatives, CNRS, and Université Evry-Val-d'Essonne/Université Paris-Saclay, an by the Region Ile de France for financial support of the 600 MHz NMR spectrometer.

## Compliance with ethical standards

**Conflict of interest** All authors declare that they have no conflict of interest.

## References

Aflaki, F., Ghoulipour, V., Saemian, N., & Salahinejad, M. (2014). A simple method for benzoyl chloride derivatization of biogenic amines for high performance liquid chromatography. *Analytical Methods, 6*, 1482–1487.

Afonso, C., Cole, R. B., & Tabet, J. C. (2010) Dissociation of even-electron ions. In Wiley (Ed.), *Electrospray and MALDI mass spectrometry*. New York: Wiley

Bazile, S., Moreau, N., Bouzard, D., & Essiz, M. (1992). Relationships among antibacterial activity, inhibition of DNA gyrase, and intracellular accumulation of 11 fluoroquinolones. *Antimicrobial Agents and Chemotherapy, 36*, 2622–2627.

Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., & Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli. Nature Chemical Biology, 5*, 593–599.

Bingol, K., Bruschweiler-Li, L., Li, D., Zhang, B., Xie, M., & Bruschweiler, R. (2016) Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis, 8*, 557–573.

Chang, Y. C., Hu, Z., Rachlin, J., Anton, B. P., Kasif, S., Roberts, R. J., & Steffen, M. (2016). COMBREX-DB: An experiment centered database of protein function: knowledge, predictions and knowledge gaps. *Nucleic Acids Research, 44*, D330–D335.

D'Argenio, D. A., Segura, A., Coco, W. M., Bunz, P. V., & Ornston, L. N. (1999). The physiological contribution of *Acinetobacter* PcaK, a transport system that acts upon protocatechuate, can be masked by the overlapping specificity of VanK. *Journal of Bacteriology, 181*, 3505–3515.

de Berardinis, V., Durot, M., Weissenbach, J., & Salanoubat, M. (2009). *Acinetobacter baylyi* ADP1 as a model for metabolic system biology. *Current Opinion in Microbiology, 12*, 568–576.

de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., … Weissenbach, J. (2008). A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Molecular Systems Biology, 4*, 174.

Dias, D. A., Jones, O. A., Beale, D. J., Boughton, B. A., Benheim, D., Kouremenos, K. A., … Wishart, D. S. (2016) Current and future perspectives on the structural identification of small molecules in biological systems. *Metabolites, 6*(4), 46.

Domingo-Almenara, X., Montenegro-Burke, J. R., Benton, H. P., & Siuzdak, G. (2018). Annotation: A computational solution for streamlining metabolomics analysis. *Analytical Chemistry, 90*, 480–489.

Dunn, W., Erban, A., Weber, R. M., Creek, D., Brown, M., Breitling, R., … Viant, M. (2013a). Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics, 9*, 44–66.

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., … Viant, M. R. (2013b). Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics, 9*, 44–66.

Frainay, C., Schymanski, E., Neumann, S., Merlet, B., Salek, R., Jourdan, F., & Yanes, O. (2018). Mind the gap: Mapping mass spectral databases in genome-scale metabolic networks reveals poorly covered areas. *Metabolites, 8*, 51.

Galperin, M. Y., & Koonin, E. V. (2004). 'Conserved hypothetical' proteins: Prioritization of targets for experimental study. *Nucleic Acids Research, 32*, 5452–5463.

Hamana, K., & Matsuzaki, S. (1992). Diaminopropane occurs ubiquitously in *Acinetobacter* as the major polyamine. *The Journal of General and Applied Microbiology, 38*, 191–194.

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., … Nishioka, T. (2010). MassBank: A public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry, 45*, 703–714.

Jimenez, J. I., Minambres, B., Garcia, J. L., & Diaz, E. (2002). Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environmental Microbiology, 4*, 824–841.

Kale, N. S., Haug, K., Conesa, P., Jayseelan, K., Moreno, P., Rocca-Serra, P., … Steinbeck, C. (2016). MetaboLights: An open-access database repository for metabolomics data. *Current Protocols in Bioinformatics, 53*, 14 13 1–18.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research, 40*, D109–D114.

Kind, T., & Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics, 8*, 105.

Kind, T., & Fiehn, O. (2010). Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical Reviews, 2*, 23–60.

Kind, T., Tsugawa, H., Cajka, T., Ma, Y., Lai, Z., Mehta, S. S., … Fiehn, O. (2018). Identification of small molecules using accurate mass MS/MS search. *Mass Spectrometry Reviews, 37*, 513–532.

Levsen, K., Schiebel, H. M., Terlouw, J. K., Jobst, K. J., Elend, M., Preiss, A., … Ingendoh, A. (2007). Even-electron ions: A systematic study of the neutral species lost in the dissociation of quasi-molecular ions. *Journal of Mass Spectrometry, 42*, 1024–1044.

Martin, G., & Williams, A. (2008) *Utilizing long-range $^{1}H$-$^{15}N$ 2D NMR spectroscopy in chemical structure elucidation*. New York: Wiley.

Martin, G. E., & Hadden, C. E. (2000). Long-range 1H–15N heteronuclear shift correlation at natural abundance. *Journal of Natural Products, 63*, 543–585.

Metzgar, D., Bacher, J. M., Pezo, V., Reader, J., Döring, V., Schimmel, P., … de Crécy-Lagard, V. (2004). *Acinetobacter* sp. ADP1: An ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Research, 32*, 5780–5790.

Peironcely, J. E., Rojas-Cherto, M., Fichera, D., Reijmers, T., Coulier, L., Faulon, J. L., & Hankemeier, T. (2012). OMG: Open molecule generator. *Journal of Cheminformatics, 4*, 21.

Rojas-Cherto, M., van Vliet, M., Peironcely, J. E., van Doorn, R., Kooyman, M., te Beek, T., … Reijmers, T. (2012). MetiTree: A web application to organize and process high-resolution multi-stage mass spectrometry metabolomics data. *Bioinformatics, 28*, 2707–2709.

Roux, A., Lison, D., Junot, C., & Heilier, J. F. (2011). Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clinical Biochemistry, 44*, 119–135.

Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., … Saito, K. (2012). RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database. *Phytochemistry, 82*, 38–45.

Siehler, S. Y., Dal, S., Fischer, R., Patz, P., & Gerischer, U. (2007). Multiple-level regulation of genes for protocatechuate degradation in *Acinetobacter baylyi* includes cross-regulation. *Applied and Environmental Microbiology, 73*, 232–242.

Singh, D. P., Govindarajan, R., Khare, A., & Rawat, A. K. (2007). Optimization of a high-performance liquid chromatography method for the separation and identification of six different classes of phenolics. *Journal of Chromatographic Science, 45*, 701–705.

Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., … Siuzdak, G. (2005). METLIN: A metabolite mass spectral database. *Therapeutic Drug Monitoring, 27*, 747–751.

Stuani, L., Lechaplais, C., Salminen, A., Ségurens, B., Durot, M., Castelli, V., … Perret, A. (2014) Novel metabolic features in *Acinetobacter baylyi* ADP1 revealed by a multiomics approach. *Metabolomics, 10*(6), 1–16.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., … Viant, M. R. (2007) Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics, 3*, 211–221.

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., … Arita, M. (2016). Hydrogen rearrangement rules: Computational MS/MS fragmentation and structure elucidation using MS-FINDER Software. *Analytical Chemistry, 88*, 7946–7958.

van der Hooft, J. J., Vervoort, J., Bino, R. J., Beekwilder, J., & de Vos, R. C. (2011). Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation. *Analytical Chemistry, 83*, 409–416.

Viant, M. R., Kurland, I. J., Jones, M. R., & Dunn, W. B. (2017). How close are we to complete annotation of metabolomes? *Current Opinion in Chemical Biology, 36*, 64–69.

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., … Scalbert, A. (2013). HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Research, 41*, D801–D807.

Wolf, S., Schmidt, S., Muller-Hannemann, M., & Neumann, S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics, 11*, 148.

Young, D. M., Parke, D., & Ornston, L. N. (2005). Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annual Review of Microbiology, 59*, 519–551.

Zhu, Z. J., Schultz, A. W., Wang, J., Johnson, C. H., Yannone, S. M., Patti, G. J., & Siuzdak, G. (2013). Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nature Protocols, 8*, 451–460.







| Journal : **Large 11306** | Article No : **1508** | Pages : **13** | MS Code : **MEBO-D-18-00279** | Dispatch : **11-3-2019** |
|---|---|---|---|---|

| | |
|---|---|
| Journal: | **11306** |
| Article: | **1508** |

# Author Query Form

**Please ensure you fill out your response to the queries raised below and return this form along with your corrections**

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

| Query | Details Required | Author's Response |
|---|---|---|
| AQ1 | Author: The affiliation '3' has been split into two different affiliations '3 and 4'. Please check, and correct if necessary. | |
| AQ2 | Author: Kindly check and confirm all affiliations. | |
| AQ3 | Author: Reference de Berardinis (2009) was cited in text but not provided in the reference list. Please provide reference in the list or delete the citation. | |
| AQ4 | Author: As per the information provided by the publisher, Fig. 1 will be black and white in print; hence, please confirm whether we can add "colour figure online" to the caption. | |