

Numéro d'ordre : 42540

# **Université de Lille – Sciences et Technologies**

École Doctorale des Sciences de la Matière, du Rayonnement et de l'Environnement

## **Thèse de Doctorat**

En vue de l'obtention du grade de

### **Docteur de l'Université de Lille**

Discipline :

Optique, Lasers, Physico-Chimie et Atmosphère

par

### **Thomas SIGNOUR**

**Extraction de signatures de bactéries par  
microspectroscopie Raman et chimiométrie. Application à  
l'étude de la composition biologique des aérosols dans  
l'environnement.**

Thèse soutenue le 11 Décembre 2017

---

#### **Rapporteurs**

Dr. Patrice BOURSON      Professeur, Université de Lorraine

Dr. Gérald THOUAND      Professeur, Université de Nantes

#### **Examineurs**

Dr. Alicia MAROTO      Maître de conférences, Université de Bretagne Occidentale

Dr. Cyril RUCKEBUSCH      Professeur, Université de Lille

#### **Directeur de thèse**

Dr. Ludovic DUPONCHEL      Professeur, Université de Lille



Numéro d'ordre : 42540

# **Université de Lille – Sciences et Technologies**

École Doctorale des Sciences de la Matière, du Rayonnement et de l'Environnement

## **Thèse de Doctorat**

En vue de l'obtention du grade de

### **Docteur de l'Université de Lille**

Discipline :

Optique, Lasers, Physico-Chimie et Atmosphère

par

### **Thomas SIGNOUR**

**Extraction de signatures de bactéries par  
microspectroscopie Raman et chimométrie. Application à  
l'étude de la composition biologique des aérosols dans  
l'environnement.**

---

#### **Rapporteurs**

Dr. Patrice BOURSON      Professeur, Université de Lorraine

Dr. Gérald THOUAND      Professeur, Université de Nantes

#### **Examineurs**

Dr. Alicia MAROTO      Maître de conférences, Université de Bretagne Occidentale

Dr. Cyril RUCKEBUSCH      Professeur, Université de Lille

#### **Directeur de thèse**

Dr. Ludovic DUPONCHEL      Professeur, Université de Lille



# Remerciements

Les différents travaux de recherche présentés dans ce manuscrit ont été réalisés au sein du Laboratoire de Spectrochimie Infrarouge et Raman (LASIR, UMR CNRS 8516). Ainsi je tiens à remercier Hervé VEZIN, directeur du laboratoire LASIR.

Je remercie la Direction Générale de l'Armement pour son financement sans lequel la réalisation de ces travaux n'aurait pas été possible. Je remercie également le Centre National de la Recherche Scientifique pour leur accompagnement durant ces trois années.

Je tiens ensuite à remercier Patrice BOURSON, Professeur de l'Université de Lorraine, et Gérald THOUAND, Professeur de l'Université de Nantes, tous deux rapporteurs de cette thèse, pour leurs remarques et les discussions sur le manuscrit. Je tiens à remercier également Alicia MAROTO, Maître de conférence de l'Université de Bretagne Occidentale pour avoir accepté d'examiner cette thèse. Au-delà de son rôle dans le jury, elle a toujours su m'encourager durant mes années à Brest. Enfin, je remercie Cyril RUCKEBUSH, Professeur de l'Université de Lille, pour avoir accepté d'être examinateur de ma thèse. Très présent durant ces trois années, il a toujours garantis la bonne ambiance au sein du groupe.

Je tiens maintenant à exprimer toute ma reconnaissance à mon directeur de thèse, Professeur Ludovic DUPONCHEL. Je le remercie non seulement pour son encadrement et son accompagnement durant toute la durée des travaux, mais également pour m'avoir guidé dans mon travail tout en me laissant une grande autonomie, m'avoir partagé ses connaissances et son expérience, et m'avoir soutenu tout au long de cette thèse.

Je remercie maintenant les différents membres du projet SIBIRAM, à savoir Christophe PANNETIER (EPISTEM R&C) et Thi-Lan HA (CSTB) avec qui j'ai étroitement collaboré durant toute la période de la thèse. Je remercie également Isabelle LACAZE (CSTB) avec qui j'ai beaucoup échangé pendant les travaux.

Je souhaite remercier tout particulièrement Myriam MOREAU. Elle a toujours été présente durant ces trois années pour résoudre mes problèmes lorsque le spectromètre était capricieux ou pour égayer mes longues journées lorsque venait le temps de faire des acquisitions pendant plusieurs jours dans le noir !

Je souhaite désormais remercier l'ensemble du groupe, la « team chimométrie » du LASIR. Merci à Ludo, Cyril, Michel, Olivier. Je remercie également Siewert que l'on nomme « Truifel » dans le milieu. Ce belgo-belge qui a commencé en même temps que moi est devenu un véritable copain et un vrai support durant ces trois années. Toujours présent dans les moments de rigolade, les virées à Quick le vendredi soir et tellement d'autres moments. Je n'oublie pas mon autre ami du bureau d'en face, Rafa. Cet italien qui n'arrive toujours pas à prononcer correctement « au lait cru », mais qui malgré cela a également été un vrai compagnon durant ces années passées. Je souhaite également remercier un ami de longue date, Silvère « QuickSilvere ». Celui avec qui j'ai habité durant la période de Licence-Master et qui un jour m'a dit « Hey, tu veux tenter ta chance à Lille ? ». Toi qui m'a fait découvrir MC, le streameur de feu et tellement d'autres choses ! Trop de trucs à dire donc j'arrête là ! Je pense également à tous ceux avec qui j'ai pu passer d'excellents moments durant ces trois années : Lucas, Julia (Raman Bro), Princess Maya, Romain, Mattéo, Rémo, Amélie, Mohamed, Dario.

Bien sûr je remercie toutes les personnes avec qui j'ai passé d'excellents moments tant au laboratoire qu'en extérieur : Samantha « Jinou », Cécilia, Vincent, Matthieu, Lucie, Diksha, Gwendoline, Émilie, Julien, Jean-Pierre, Pascal, Thomas, Ismail. J'en oublie très certainement, c'est pourquoi je remercie également tous les membres du laboratoire.

Je remercie maintenant mes amis Carole, Elise, William. Depuis Brest, Will et Quick ont toujours été à mes côtés (au sens propre et figuré du terme). Leur soutien m'a énormément aidé à accomplir mes projets donc merci les gars.

Un gros merci également pour ma famille, notamment ma sœur Maëlle, ma maman Josiane et mon papa Job. Ils ont toujours été présents dans les bons comme les mauvais moments. Ils m'ont supporté dans mes choix et encouragé à poursuivre coûte que coûte dans mon parcours. Je remercie aussi mon autre famille, ceux du Penquer ! Merci pour votre aide

Enfin, comment l'oublier, j'adresse le plus gros des mercis à la conjointe, Anne-Charlotte BOZEC, sans qui rien n'aurait été possible. Malgré la distance qui nous sépareit quotidiennement, tu as toujours su être à mes côtés, me soutenir, supporter mon stress et sans toi, ça aurait quand même été vachement plus dur ! Merci =)

# Résumé

## **Extraction de signatures de bactéries par microspectroscopie Raman et chimiométrie. Application à l'étude de la composition biologique des aérosols dans l'environnement.**

Depuis plusieurs années, l'étude et le contrôle de la qualité de l'air sont au cœur de toutes les préoccupations. En 2012, la DGA (Direction Générale de l'Armement) met en place le programme ASTRID (Accompagnement Spécifique des Travaux de Recherches et d'Innovation Défense) accompagnant les travaux de recherche duale civile et militaire. Cette thèse s'inscrit dans cette démarche et propose d'étudier la faisabilité du concept de détection et d'identification rapides des microorganismes présents dans un échantillon d'air par microspectroscopie Raman, avec une résolution au niveau de l'espèce. Pour cela, nous construisons un modèle chimiométrique de classification des microorganismes représentatifs de la biodiversité naturelle en acquérant, sans *a priori*, d'une part les spectres Raman de ces microorganismes après biocollecte et étalement sur la lame d'un microspectromètre Raman, et d'autre part les séquences génomiques codant les ARN 16S de ces mêmes microorganismes.

Les travaux de recherche présentés dans cette thèse présentent donc les différentes études mises en œuvre lors du développement d'un nouveau protocole permettant l'analyse des bactéries issues d'aérosols naturels environnementaux. Nous démontrons la nécessité d'optimiser l'acquisition des spectres Raman sur les bactéries et le traitement statistique des données spectrales permettant le développement de modèles de classification présentant des taux de reconnaissance élevés.

**Mots clés :** spectroscopie Raman, chimiométrie, bactérie, classification

# Abstract

## **Extraction of bacterial signatures by Raman microspectroscopy and chemometrics. Application to the study of the biological composition of aerosols in the environment.**

For several years, the study and the control of the quality of the air are at the heart of all the concerns. In 2012, the DGA (Direction Générale de l'Armement) employs the ASTRID program (Accompagnement Spécifique des Travaux de Recherches et d'Innovation Défense), to accompany the dual civil and military research work. This thesis is part of this approach and proposes the feasibility study, by Raman microspectroscopy, of the concept of rapid detection and identification of microorganisms present in an air sample, with a resolution at the species level. For this, we construct a chemometric model for the classification of micro-organisms representative of the natural biodiversity. Such a model is built by acquiring, without *a priori* i) the Raman spectra of these microorganisms after biocollection; and ii) the genomic sequences encoding the 16S RNAs of these same microorganisms.

The research presented in this thesis therefore presents the different studies carried out during the development of a new protocol allowing the analysis of bacteria from natural environmental aerosols. We demonstrate the need to optimize the acquisition of Raman spectra on bacteria and the statistical processing of spectral data that allows the development of classification models with high recognition rates.

**Key words:** Raman spectroscopy, chemometrics, bacteria, classification



# Table des matières

<b>Résumé .....</b>	<b>i</b>
<b>Abstract .....</b>	<b>iv</b>
<b>Introduction .....</b>	<b>1</b>
<b>Chapitre 1 - Instrumentation et outils chimiométriques pour l'identification des bactéries .....</b>	<b>4</b>
<b>1. La spectroscopie Raman.....</b>	<b>4</b>
1.1. Principe de diffusion Raman.....	5
1.2. Le système de mesure Raman .....	8
1.2.1. <i>Le microspectromètre Raman .....</i>	<i>8</i>
1.2.2. <i>Les sources.....</i>	<i>10</i>
<b>2. La Microdissection à Capture Laser (LCM) .....</b>	<b>11</b>
<b>3. Les outils chimiométriques .....</b>	<b>15</b>
3.1. Le prétraitement des données spectrales.....	15
3.1.1. <i>Les techniques de normalisation .....</i>	<i>16</i>
3.1.2. <i>Les méthodes de correction de ligne de base.....</i>	<i>17</i>
3.2. Analyse en composantes principales.....	19
3.3. Les méthodes de classification .....	21
3.3.1. <i>Les méthodes de classification non-supervisées .....</i>	<i>21</i>
3.3.2. <i>Les méthodes de classification supervisées.....</i>	<i>23</i>

<b>4. Etat de l’art de l’analyse par spectroscopie Raman et application de la chimométrie pour la classification de bactéries.....</b>	<b>28</b>
<b>Chapitre 2 – Développement d’un protocole original de classification, agents bactériens et échantillonnage.....</b>	<b>32</b>
<b>1. Protocole développé pour la construction d’un modèle de classification des agents bactériologiques.....</b>	<b>32</b>
<b>2. Agents bactériologiques.....</b>	<b>34</b>
2.1. Classification des bactéries.....	34
2.2. Morphologie et composition .....	36
2.3. Caractère pathogène de certains individus.....	37
2.4. Souches bactériennes cultivées.....	39
2.4.1. <i>Staphylococcus epidermidis</i> .....	39
2.4.2. <i>Pseudomonas fluorescens</i> .....	39
2.4.3. <i>Pseudomonas syringae</i> .....	40
2.4.4. <i>Escherichia coli</i> .....	40
2.4.5. <i>Souches bactériennes isolées</i> .....	41
2.5. Conditions de culture .....	41
<b>3. Préparation des échantillons.....</b>	<b>42</b>
3.1. Méthodes de prélèvement des aérosols atmosphériques.....	42
3.1.1. <i>Prélèvements des aérosols à l’aide d’un impacteur en cascade multi-étages de type Andersen</i> .....	43
3.1.2. <i>Prélèvements des aérosols en milieu liquide à l’aide du SKC Biosampler</i> .....	44
3.2. Méthodes de dépôt .....	46
3.2.1. <i>Dépôt par goutte</i> .....	46
3.2.2. <i>Dépôt par aérosolisation/sédimentation</i> .....	46

<b>4. Méthodes d'identification moléculaire .....</b>	<b>46</b>
4.1. Identification moléculaire de bactéries cultivées.....	47
4.2. Identification moléculaire des bactéries isolées par microdissection .....	49
<b>5. Conclusion .....</b>	<b>51</b>
 <b>Chapitre 3 – Optimisation de la mesure Raman, des modèles de classification et confrontation aux aérosols naturels .....</b>	 <b>52</b>
<b>1. Acquisition des spectres Raman .....</b>	<b>53</b>
1.1. Détermination du support.....	54
1.2. Optimisation de la mesure Raman .....	58
1.2.1. Détermination de la longueur d'onde du laser pour l'acquisition.....	58
1.2.2. Choix du réseau de diffraction.....	61
1.2.3. Détermination du temps d'acquisition .....	62
1.3. Nature des contributions dans le spectre Raman des bactéries.....	63
1.4. Potentiel de la spectroscopie M-CARS et EPI-CARS .....	68
<b>2. Classification par méthode SVM .....</b>	<b>70</b>
2.1. Optimisation des prétraitements des données spectrales .....	71
2.2. Construction et validation du modèle de classification SVM.....	78
2.2.1. Estimation des paramètres SVM.....	78
2.2.2. Calcul du modèle de classification SVM .....	80
<b>3. Adaptation du protocole aux aérosols naturels.....</b>	<b>84</b>
3.1. Complexité des aérosols naturels.....	84
3.1.1. Choix de la maille de l'impacteur (étage 5 et 6).....	85
3.1.2. Morphologie des particules.....	86
3.1.3. Densité particulaire .....	87

3.2. Amélioration du protocole .....	87
3.2.1. <i>Marquage par agent fluorescent</i> .....	88
3.2.2. <i>Discrimination morphologique</i> .....	94
3.2.3. <i>Concentration en bactéries et diminution de la densité particulaire</i> .....	96
3.2.4. <i>Dépôt des particules sur CaF<sub>2</sub></i> .....	98
3.3. Étude de bactéries provenant de souches environnementales.....	99
<b>4. Conclusion .....</b>	<b>104</b>
<b>Chapitre 4 – Modules NanoGPS, microdissection laser et analyse génomique .....</b>	<b>106</b>
<b>1. Système de repositionnement absolu : le nanoGPS.....</b>	<b>106</b>
1.1. Principe général .....	107
1.2. Couplage virtuel de différents instruments .....	110
1.3. Précision du système nanoGPS .....	111
1.3.1. <i>Cas n°1 : Superposition de deux images sans réétalonnage</i> .....	112
1.3.2. <i>Cas n°2 : Utilisation de deux objectifs pour construire la cartographie</i>	113
1.3.3. <i>Cas n°3 : Superposition de deux images après rotation de la lame sans réétalonnage</i> .....	114
1.3.4. <i>Cas n°4 : Superposition de deux images avec réétalonnage entre les deux observations</i> .....	115
1.3.5. <i>Cas n°5 : Superposition entre les images acquises sur le microscope du microspectromètre Raman et sur le microdissecteur</i> .....	116
<b>2. Microdissection .....</b>	<b>117</b>
2.1. Optimisation de la taille du spot laser .....	117
2.2. Influence des modes de dépôt des bactéries sur l'extraction des cellules par microdissection laser.....	119

2.2.1. <i>Influence du dépôt par goutte</i> .....	120
2.2.2. <i>Influence du dépôt par sédimentation</i> .....	121
2.3. Difficultés liées à l'adhérence des cellules .....	121
<b>3. Analyse génomique sur cellule unique .....</b>	<b>123</b>
3.1. Analyse génomique sur cellules non marquées .....	123
3.2. Influence du marquage et du temps d'exposition laser sur l'analyse génomique	125
<b>4. Conclusion .....</b>	<b>126</b>
<b>Conclusions et perspectives .....</b>	<b>128</b>
<b>Bibliographie.....</b>	<b>132</b>

# Introduction

Depuis la découverte au XIX<sup>ème</sup> siècle des premiers microorganismes pathogènes et de leur rôle dans les maladies infectieuses, l'Homme n'a eu de cesse de développer et d'améliorer des outils pour les détecter et les identifier. D'abord centrées sur la santé humaine, les préoccupations se sont rapidement étendues à celles de la santé animale, puis plus récemment à celles de la sûreté et de la sécurité des populations et de la protection de l'environnement vis-à-vis de rejets industriels. Les principales voies de contamination bactériologiques sont les aliments, l'eau et l'air. Aussi, la présence de résidus médicamenteux dans l'environnement a permis aux agents bactériologiques de muter et de développer un caractère antibiorésistant [1]. Les chercheurs se doivent donc d'être constamment en alerte pour surveiller le développement de nouvelles espèces et le risque qu'elles présentent pour les êtres vivants. Certaines applications nécessitent des délais de réponses tellement courts qu'une détection en quasi-temps réel est nécessaire. Malheureusement, les méthodes de biologie moléculaire sont trop lentes pour fournir une alerte dans ce délai tandis que les méthodes mises en œuvre dans les détecteurs actuels sont insuffisamment sélectives et sensibles.

Dans de nombreuses applications, Il est donc nécessaire de développer des systèmes pouvant détecter en temps réel la présence de certains microorganismes dans un environnement donné. Les enjeux de cette nécessité présentent un caractère dual, c'est-à-dire qu'ils ont une importance tant dans le domaine militaire que dans le domaine civil. En effet, il est vital pour nos forces armées de pouvoir identifier les risques avant une intervention sur un théâtre d'opérations, de détecter en quelques secondes la présence dans l'air de microorganismes hautement pathogènes dispersés par un agresseur ou bien de confiner les populations atteintes par une épidémie. Dans le cadre du renseignement, une détection rapide pourrait aussi permettre de confondre des dirigeants d'un pays expérimentant des armes bactériologiques sur leurs populations. Dans le domaine civil, les enjeux sont tout aussi importants. En effet, il est nécessaire de pouvoir surveiller et contrôler la qualité de l'air de nos infrastructures publiques ou privées. Par exemple, il est important de détecter la présence d'agents pathogènes dans les hôpitaux pour diminuer au maximum le risque d'infections nosocomiales. Il faut donc développer une technologie suffisamment compacte et

transportable en tout lieu et à tout moment pour garantir la sécurité de tous. Il n'est donc pas impossible que dans plusieurs années, toutes nos habitations, bureaux et hôpitaux soient équipés d'un tel système pouvant analyser plusieurs milliers de particules à la seconde.

Depuis plusieurs années, de nombreux travaux ont porté sur l'intérêt de la mise en place de systèmes et d'outils afin d'étudier et de surveiller la qualité de l'air à travers le monde. C'est en 1970 que le terme de biosurveillance de l'air est né lorsque deux chercheurs britanniques, D. L. Hawksworth et F. Rose, ont démontré qu'en étudiant la croissance des lichens présents sur le tronc des arbres, il est possible de déterminer qualitativement la teneur en dioxyde de soufre [2]. Aux États-Unis d'Amérique, suite à des attaques à l'anthrax en 2001, le gouvernement fédéral a mis en place le programme BioWatch [3] qui permet la détection d'agents pathogènes dans l'air en cas d'attaque bioterroriste, pour un coût total de près de 40 millions de dollars. Ce sont en tout 32 des plus grandes villes du pays qui sont équipées de pompes et de filtres. Tous les jours, des échantillons sont prélevés puis une analyse génomique des bactéries est effectuée par des centres de l'Institut National de la Santé. Nous remarquons que l'analyse ne porte que sur certaines villes, sans une véritable résolution temporelle. À ce jour, par deux fois, la présence de la bactérie *Francisella tularensis* a été reportée par les capteurs mis en place. En France, ce n'est que très récemment, en 2006, que l'AFNOR (Association Française de Normalisation) a décidé de créer une commission de biosurveillance de l'air dénommée T95Air. La biosurveillance de l'eau ou de l'air est donc un sujet au cœur de l'actualité.

Le programme ASTRID (Accompagnement Spécifique des Travaux de Recherches et d'Innovation Défense) financé par la DGA (Direction Générale de l'Armement) et mis en œuvre par l'ANR (Agence Nationale de la Recherche) accompagne depuis 2012 les travaux de recherche duale civile et militaire. Ce programme a permis le financement de deux projets ASTRID parallèles, à savoir NEOSPRAM (Nouveaux procédés de spectroscopie Raman pour l'analyse d'agents biologiques) et SIBIRAM (Signature des agents biologiques par microspectroscopie Raman). Le premier projet ASTRID, NEOSPRAM, a pour objectif de développer un outil de mesure par spectroscopie Raman exalté. Cette technique offre des temps d'analyse tellement courts, que l'analyse de plusieurs milliers de particules à la seconde devient possible. Le second projet, SIBIRAM, a pour objectif principal d'apporter une preuve solide de la faisabilité de concept de détection et d'identification rapides des microorganismes

présents dans un échantillon d'air par microspectroscopie Raman, avec une résolution au niveau de l'espèce. Les travaux de cette thèse s'inscrivent donc dans cette dernière ANR. Pour cela, nous construisons un modèle chimiométrique (traitement statistique des données spectrales multivariées) de classification des microorganismes représentatifs de la biodiversité naturelle en acquérant, sans a priori, d'une part les spectres Raman de ces microorganismes après biocollecte et étalement sur la lame d'un microspectromètre Raman, et d'autre part les séquences génomiques codant les ARN 16S de ces mêmes microorganismes.

Les travaux de recherche présentés dans le manuscrit présentent donc les différentes études mises en œuvre lors du développement d'un nouveau protocole permettant l'analyse des bactéries issues d'aérosols naturels environnementaux. Le premier chapitre est consacré à la présentation des agents bactériologiques étudiés au cours de ces travaux. Nous détaillons également les méthodes de préparation mises en jeu durant toute la durée des travaux et les outils d'analyse moléculaire des bactéries, tant cultivées que microdissectées. Un second chapitre est dédié aux principes et à l'instrumentation de la microspectroscopie Raman, de la microdissection à capture laser et de la chimiométrie. D'une part, nous abordons les principes et le fonctionnement de chacun de ces deux instruments. Et d'autre part, nous présentons les outils chimiométriques mis en œuvre pour le traitement des données spectrales (correction des spectres) et les techniques de classification chimiométriques. Dans un troisième chapitre, nous détaillons toute l'optimisation et la mise en place de la mesure par microspectroscopie Raman. Nous expliquons également le développement et l'optimisation pour la construction d'un modèle de classification SVM (*Support Vector Machine*). Puis nous abordons la problématique des échantillons naturels et les dispositions prises pour améliorer le protocole de recherche et de localisation des bactéries dans les échantillons d'air. Enfin, dans le quatrième et dernier chapitre, nous présentons le module nanoGPS qui est un outil de repositionnement absolu permettant un couplage virtuel entre le microspectromètre Raman et le microdissecteur. Puis, nous détaillons la mise en place du prélèvement des bactéries par la microdissection à capture laser sur cellules uniques. Enfin, nous évaluons les résultats obtenus suite aux analyses moléculaires opérées sur les cellules bactériennes uniques prélevées par microdissection.



# Chapitre 1 - Instrumentation et outils chimométriques pour l'identification des bactéries

## 1. La spectroscopie Raman

Les techniques spectroscopiques sont des méthodes utilisées pour obtenir des informations sur la structure et la composition de la matière, grâce à un processus d'interaction rayonnement/matière. Certaines spectroscopies reposent sur l'absorption et l'émission d'un rayonnement électromagnétique par un noyau atomique possédant un spin nucléaire pour la spectroscopie RMN (Résonance Magnétique Nucléaire) ou par certains électrons présents dans des espèces chimiques radicalaires pour la spectroscopie RPE (Résonance Paramagnétique Electronique). En exposant certains composés chimiques à des radiations dans l'ultraviolet (UV) ou dans le domaine du visible, allant de 200nm à 400nm pour l'UV et de 400nm à 750nm pour le visible, les composés chimiques sont susceptibles de subir une ou plusieurs transitions électroniques se traduisant par une absorption d'énergie. Il existe également des spectroscopies vibrationnelles qui permettent d'obtenir des informations sur la composition chimique de l'échantillon. Ces techniques sondent les vibrations des liaisons chimiques lors d'une interaction entre la lumière et la matière. Les spectroscopies moyen-infrarouge (3000 – 5000 nm) et proche-infrarouge (780 – 3000 nm) étudient l'absorbance des composés chimiques dans ce domaine électromagnétique. Dans la suite des travaux, une autre spectroscopie vibrationnelle, sera utilisée : la spectroscopie Raman.

## **1.1. Principe de diffusion Raman**

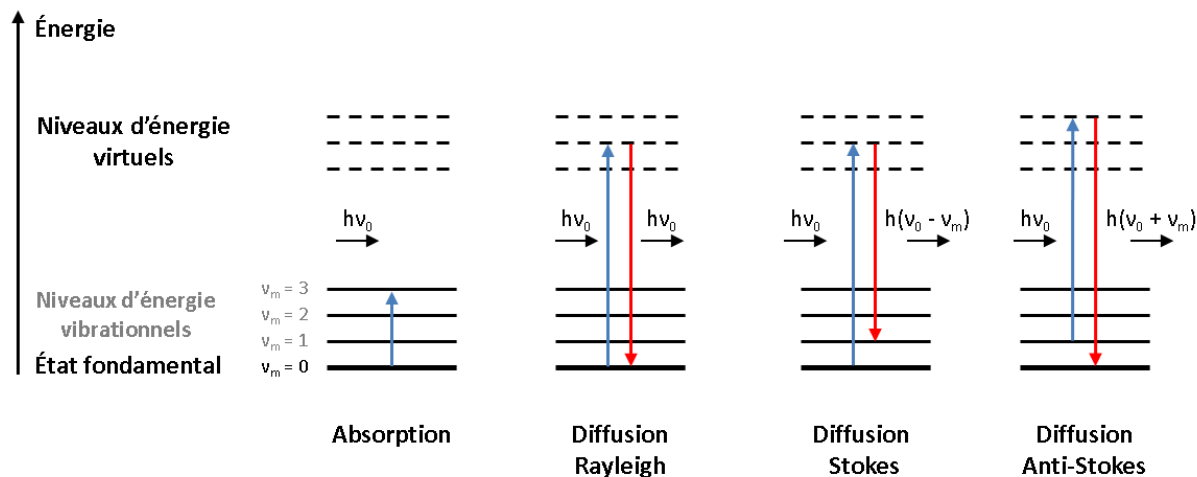
C'est en 1928 que le phénomène de diffusion Raman est rapporté pour la première fois. Sir C.V. Râman et K.S. Krishnan, deux physiciens indiens, mettent en évidence un nouveau phénomène de diffusion moléculaire de la lumière [4]. En parallèle et indépendamment de ces travaux, deux physiciens russes, G. Landsberg et L. Mandelstam font les mêmes observations [5], mais C.V. Râman sera le premier à rapporter ses observations et se verra attribuer le Prix Nobel en 1930 pour ses travaux sur la diffusion de la lumière.

Lorsque la matière, caractérisée par des niveaux d'énergie vibrationnelle, est exposée à un rayonnement lumineux, celle-ci peut présenter différents comportements (Figure 1). Si la fréquence de vibration de l'excitation lumineuse  $\nu_0$  est égale à la fréquence caractéristique du niveau d'énergie de la molécule constituant la matière observée  $\nu_m$ , alors un phénomène d'absorption est observé. C'est sur ce principe que repose notamment la spectroscopie infrarouge. En revanche, si la fréquence  $\nu$  est très supérieure à  $\nu_m$ , alors des phénomènes de diffusion sont observés. C'est sur ce phénomène de diffusion que repose la spectroscopie Raman. La spectroscopie Raman est une technique de caractérisation, non destructive, de la composition moléculaire d'un matériau.

Lorsqu'un matériau est impacté par un flux de photons d'énergie  $h\nu_0$ , ces derniers sont absorbés et réémis sous forme de diffusion. La diffusion la plus souvent rencontrée est la diffusion Rayleigh. Lors de la désexcitation du matériau, un photon est émis sans perte d'énergie. La diffusion est dite élastique. Dans le cas où l'énergie du photon émis par la diffusion est différente de l'énergie incidente, suite à un échange d'énergie avec la matière, il faudra parler de diffusion inélastique. Deux types de diffusion inélastique sont exploitées en spectroscopie Raman : la diffusion Stokes et la diffusion Anti-Stokes. La diffusion Stokes est caractérisée par une perte d'énergie et le photon émis possède une énergie  $h(\nu_0 - \nu_m) < h\nu_0$ . Inversement, le milieu peut céder de l'énergie au photon et lors de la diffusion, le photon émis aura une énergie  $h(\nu_0 + \nu_m) > h\nu_0$  se traduisant par l'apparition de raies Anti-Stokes, symétriques à celles du Stokes par rapport au Rayleigh correspondant à la longueur d'onde incidente.

Ces phénomènes de diffusion Stokes et Anti-Stokes sont des phénomènes rares, se produisant seulement une fois sur un million de photons. C'est pourquoi il est nécessaire

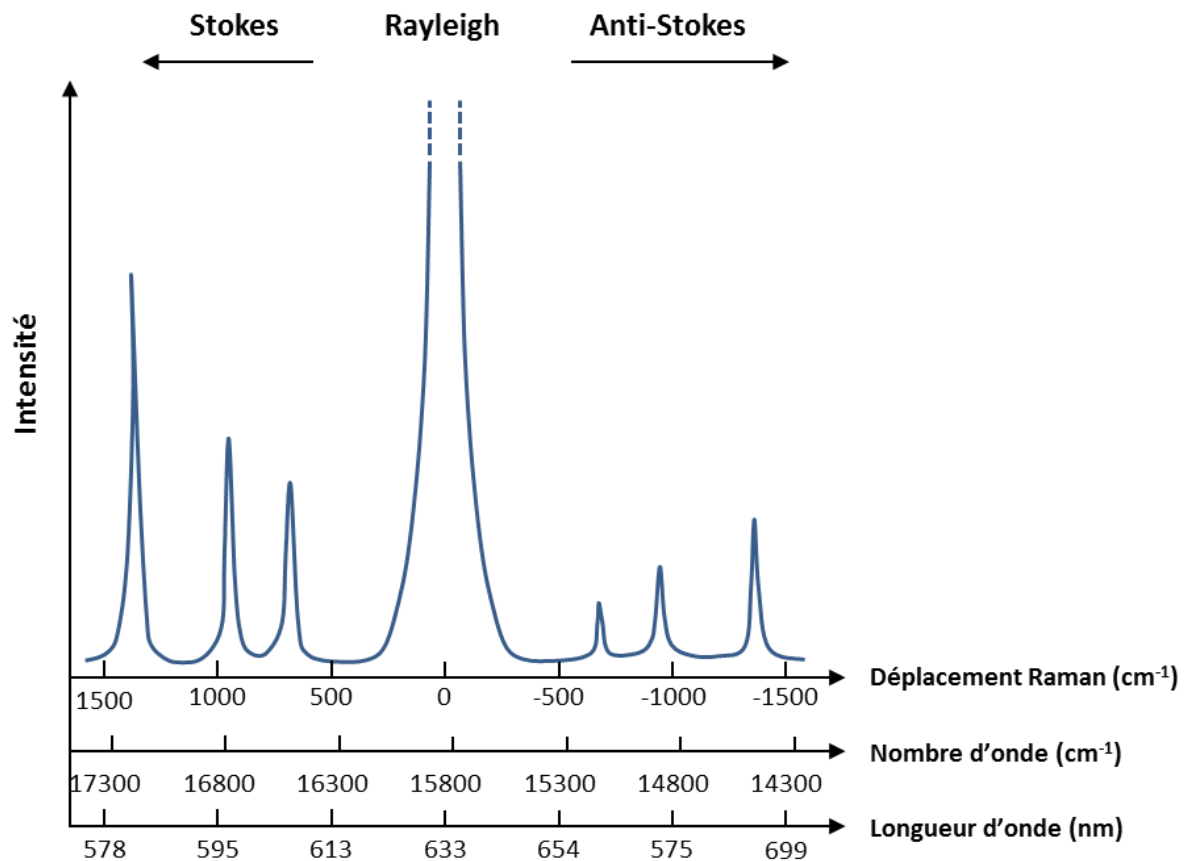
d'utiliser des sources laser pour obtenir un signal Raman suffisamment intense. La collecte de la lumière diffusée permet d'obtenir un spectre Raman offrant de nombreuses informations sur les modes de vibrations et donc sur le caractère chimique de l'échantillon.



**Figure 1 – Représentation des phénomènes engendrés suite à l'excitation d'un corps par la lumière.**

La spectroscopie Raman est une technique complémentaire à la spectroscopie infrarouge. Toutefois, l'origine du signal observé avec les deux techniques est différente. Dans le cas de la spectroscopie infrarouge, il s'agit d'un phénomène d'absorption de la lumière et les bandes obtenues sont causées par des vibrations des liaisons accompagnées d'une variation du moment dipolaire de la molécule. Ainsi, les liaisons présentant un fort moment dipolaire présenteront une forte absorption dans l'infrarouge. Inversement en spectroscopie Raman, la vibration de liaison faiblement polarisée conduira à une variation du moment dipolaire induit de la molécule. Les liaisons covalentes, a priori plus polarisables, conduiront donc à des raies Raman plus intenses. En général, une vibration très active en Raman engendrera un signal faible en infrarouge et inversement rendant ces deux techniques complémentaires. Néanmoins certaines molécules peuvent être actives à la fois dans ces deux types de spectroscopie ou au contraire dans aucune d'entre elles.

Lors de la mesure de spectres Raman, l'effet qui sera principalement observé sera la diffusion Stokes car celle-ci est plus intense que la diffusion Anti-Stokes dans la plupart des conditions expérimentales comme représenté sur la Figure 2.



**Figure 2 - Représentation d'un spectre de diffusion utilisant une source monochromatique émettant à 633nm.**

Le rapport d'intensité entre les raies Stokes et anti-Stokes peut aisément être évalué grâce à l'utilisation de la loi de distribution statistique de Maxwell-Boltzmann. Les modes vibrationnels de haute énergie étant moins peuplés, les raies Stokes seront plus intenses que les raies anti-Stokes.

$$N_1 = N_0 e^{-\left(\frac{E_1 - E_0}{k_B T}\right)}$$

(Eq.1)

$N_0$  et  $N_1$  correspondent respectivement à la récurrence du phénomène de diffusion Stokes et anti-Stokes.  $E_0$  et  $E_1$  correspondent respectivement aux niveaux énergétiques fondamental et vibrationnel auxquels peuvent être portées les molécules.  $E_0$  étant toujours inférieure à  $E_1$ , la valeur de l'exponentielle sera toujours comprise entre 0 et 1. Il en résulte

que la valeur de  $N_1$  sera toujours inférieure à la valeur  $N_0$ . C'est pourquoi en spectroscopie Raman conventionnelle, les signaux enregistrés sont focalisés sur le signal de diffusion Stokes.

## **1.2. Le système de mesure Raman**

Le nombre d'application du principe de diffusion Raman est aussi important que le nombre d'appareils en fonction dans le monde. C'est pourquoi dans ces travaux, nous nous limiterons à l'utilisation de la microspectroscopie Raman conventionnelle.

### **1.2.1. Le microspectromètre Raman**

Malgré sa bonne capacité d'analyse, la spectroscopie Raman manquait cruellement de résolution spatiale à ses débuts. Durant les années 1970, le constructeur Jobin-Yvon, eu l'idée de coupler son spectromètre Raman à un microscope optique lui permettant d'atteindre alors une résolution de l'ordre du micromètre [6, 7]. Puis avec l'amélioration continue des systèmes optiques et des microscopes, la microspectroscopie Raman s'est alors généralisée permettant aujourd'hui d'obtenir une résolution spatiale inférieure au micromètre à l'aide d'objectifs x100.

Après excitation de l'échantillon par le laser, le signal diffusé passe par le spectromètre pour finalement parvenir au détecteur pour l'acquisition du spectre Raman. La majorité des détecteurs présents dans les spectromètres sont multicanaux. Ils peuvent faire l'acquisition simultanée de toutes les fréquences du signal diffusé. Les détecteurs CCD (*Charge Coupled Device*) en sont un très bon exemple. Le couplage d'un réseau de diffraction au détecteur CCD est nécessaire.

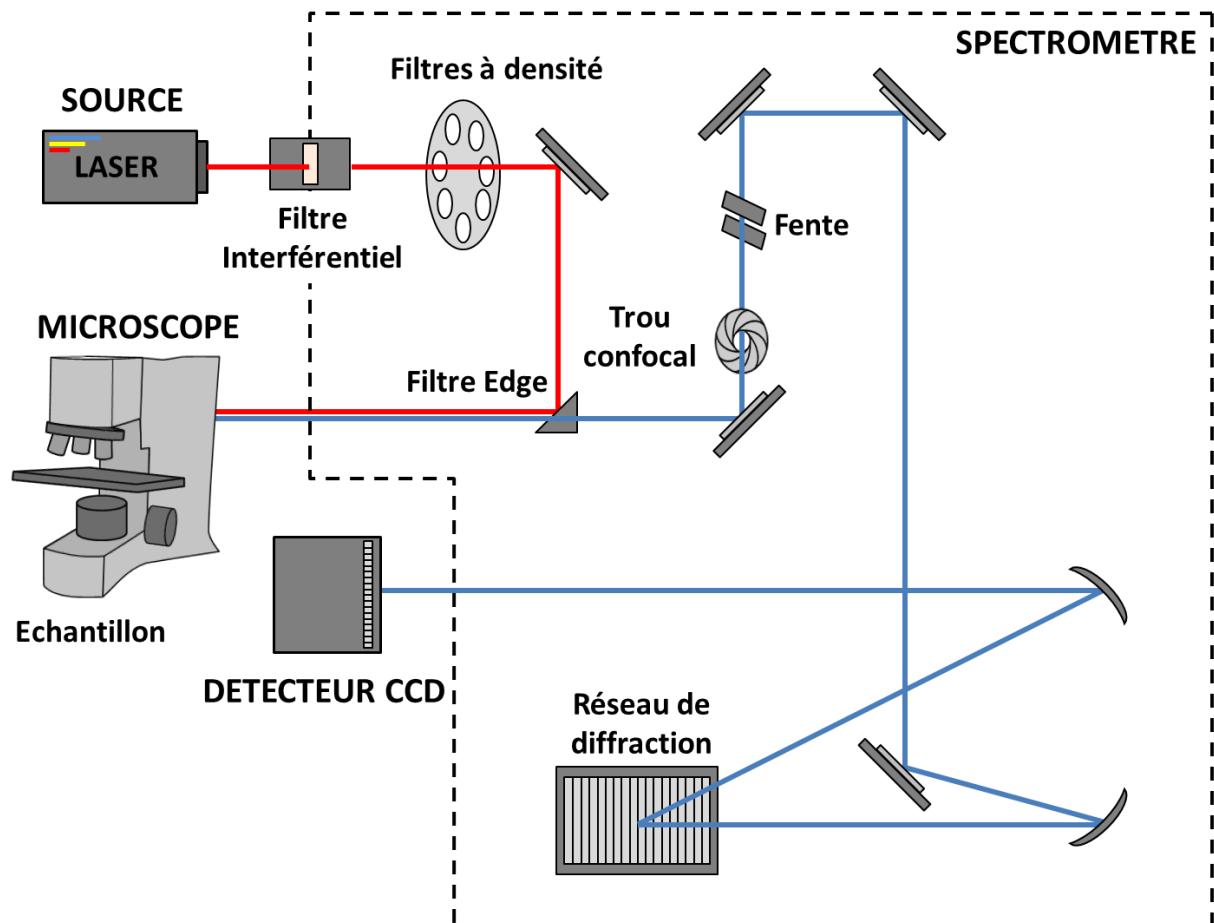


Figure 3 - Schéma du microspectromètre Raman LabRam HR 800 (Horiba Jobin Yvon).

Les analyses effectuées lors de ces travaux ont été réalisées avec un microspectromètre Raman LabRam HR 800 (HORIBA Jobin-Yvon) représenté sur la Figure 3. En sortie de source, le faisceau laser passe par un filtre interférentiel adapté à la longueur d'onde utilisée qui a pour but de filtrer la longueur d'onde puis la puissance du faisceau laser est contrôlée grâce à une roue sur laquelle sont fixés des filtres à densité. Le faisceau est ensuite conduit à l'échantillon à travers le microscope. Suite à l'excitation de l'échantillon, les rayons diffusés sont orientés en premier lieu vers un filtre Edge ou Notch permettant d'éliminer notamment les signaux de diffusion Rayleigh. Le signal passe ensuite à travers un trou confocal permettant de limiter le volume sondé au niveau de l'échantillon. Selon le nombre de traits que comporte le réseau, la résolution spectrale sera plus ou moins élevée, autrement dit le spectre obtenu sera plus ou moins détaillé. Les signaux sont ensuite dirigés vers le détecteur CCD qui assure la conversion d'un signal lumineux en un signal électrique.

Un traitement informatique des données récoltées permet de représenter le spectre Raman sous forme d'intensité des signaux en fonction des fréquences relatives à la source excitatrice.

Le spectromètre peut être équipé de différents réseaux de diffraction (300, 600 et 1800 traits par mm). Le choix du réseau à utiliser est important. Plus le nombre de traits sera important, plus la résolution spectrale sera bonne mais au détriment de l'efficacité de diffraction et donc du rapport signal sur bruit (ou SNR pour *Signal to Noise Ratio*). Dans le cadre de ces travaux, un des objectifs est de mettre en place un protocole permettant l'analyse Raman des bactéries dans un temps suffisamment restreint. Plusieurs paramètres sont donc optimisables et seront présentés dans le Chapitre 3 - 1.2. Optimisation de la mesure Raman.

### **1.2.2. Les sources**

L'utilisation de sources lumineuses suffisamment intenses est indispensable pour réaliser l'acquisition d'un spectre Raman. Le laser, source lumineuse monochromatique, répond à cette demande devenant ainsi la source principale d'excitation au sein des dispositifs actuels. Depuis les années 1960, les lasers n'ont cessé d'évoluer et de se perfectionner. Il devient alors possible d'analyser un échantillon aussi bien dans l'ultra-violet que dans le proche infrarouge en passant par le visible.

Une des sources lasers les plus communes sont des sources à gaz permettant d'accéder à un large nombre de longueur d'onde. Les sources à gaz les plus connues sont les lasers à hélium-néon (He-Ne) émettant à 632.80 nm. On retrouve également les lasers à argon, krypton ou xénon ionisé, permettant de travailler à une multitude de longueurs d'onde selon le gaz utilisé. Cependant ces derniers restent tout de même plus contraignants à cause de la nécessité d'un refroidissement de la source. Les sources solides occupent également une place importante sur le marché de la spectroscopie malgré le fait qu'elles n'offrent pas une gamme de longueur d'onde aussi large que les sources à gaz. Une des sources communément rencontrée est le laser Nd-YAG (*neodymium-doped yttrium aluminium garnet*), qui permet l'émission d'un faisceau à 1064 nm, dans le proche l'infrarouge. En utilisant un doubleur de fréquence, il est possible d'accéder à une longueur d'onde d'excitation de 532 nm. Les diodes lasers, aujourd'hui majoritairement utilisées dans l'industrie présentent également un intérêt

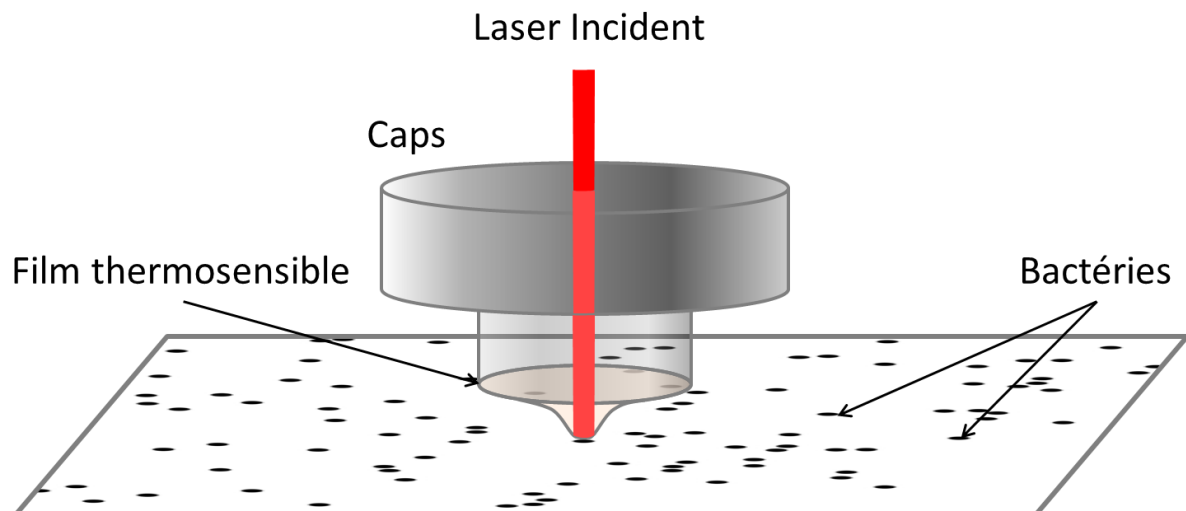
pour la spectroscopie. Il est aussi possible d'obtenir d'excellentes qualités de faisceaux balayant le spectre électromagnétique. Les diodes peuvent également être utilisées pour pomper un laser (DPPS : *Diode-pumped solid-state*) permettant ainsi d'accéder à une gamme de longueur d'onde très large. Au cours de ces travaux de recherche, plusieurs lasers ont été utilisés afin de déterminer lequel permet d'obtenir une signature Raman avec un rapport signal sur bruit suffisamment élevé dans un temps le plus court possible sans détériorer le matériel biologique et génétique.

Le microspectromètre Raman est un système «multi-excitations». Il est couplé à trois sources lasers monochromatiques : un laser interne He-Ne irradiant à 633 nm et deux lasers solides externes fonctionnant à 532 nm et à 473 nm. Il est donc possible d'analyser une même particule à différentes longueurs d'ondes. De plus, la principale interférence de la spectroscopie Raman est la fluorescence. En considérant que la fluorescence est dépendante de la longueur d'onde d'excitation et que la diffusion Raman ne l'est pas, il est possible de trouver rapidement un bon compromis en utilisant l'un ou l'autre des lasers équipés.

## 2. La Microdissection à Capture Laser (LCM)

La microdissection à capture laser (LCM de l'anglais *Laser Capture Microdissection*) est une technique d'isolement de cellules. Grâce à cette technique, il est possible d'isoler une cellule ou un groupe de cellules à partir d'un échantillon hétérogène permettant d'améliorer l'efficacité et la précision des méthodes d'analyse génomique et protéomique. Cette technique datant du milieu des années 1990, a été initialement conçue pour pallier les inconvénients que présentent les méthodes de microdissection tissulaires existantes à cette époque, à savoir la pureté des échantillons, la nécessité d'opérateurs manuels qualifiés, le grand risque de contamination et le temps d'extraction [8]. A l'origine conçu pour isoler des tissus cellulaires cancéreux [9], l'instrument est aujourd'hui utilisé dans bien des domaines, aussi bien pour l'extraction de cellules cancéreuses que pour des enquêtes de la police scientifique pour la recherche de traces ADN.

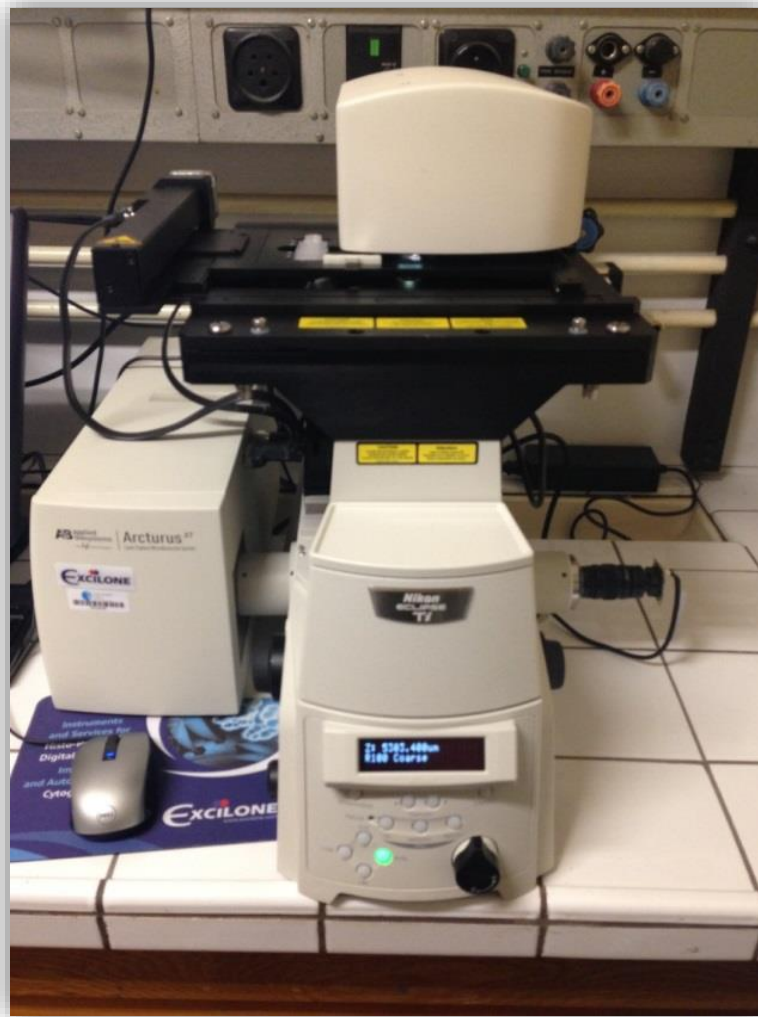




**Figure 4 - Schéma de principe de la microdissection laser.**

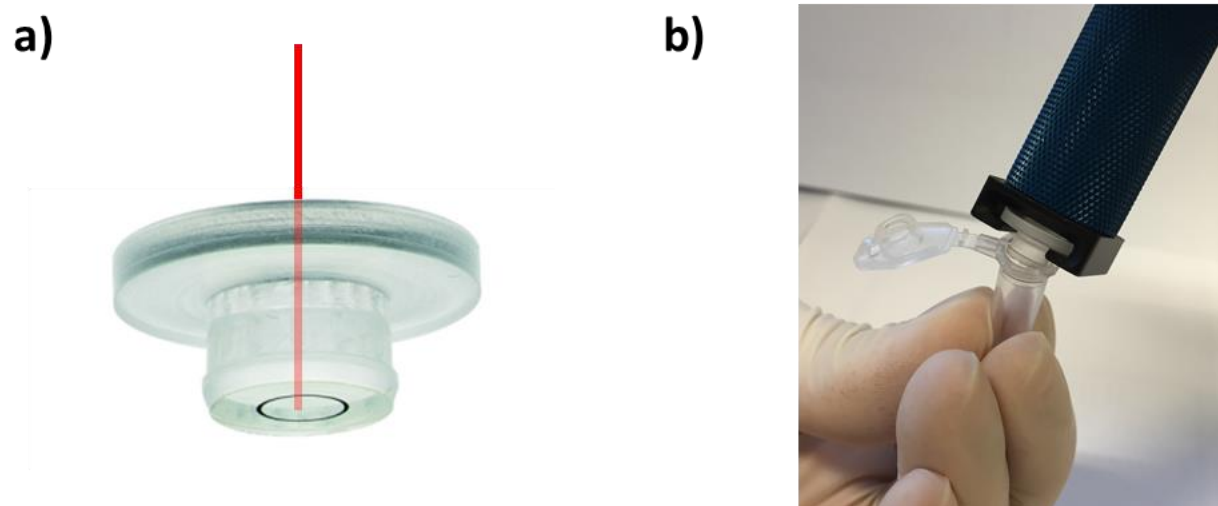
La méthode d'extraction, schématisée sur la Figure 4, consiste à placer un fin film de polymère thermosensible au-dessus de la zone d'intérêt, autrement dit où sont situées les cellules à prélever. Ensuite un faisceau laser infrarouge à faible puissance est dirigé à travers le film thermoplastique. Une impulsion laser va faire fondre le polymère et celui-ci va couler sur la zone d'intérêt. L'impulsion laser doit être suffisamment courte et intense pour que la zone recouverte soit d'un diamètre le plus petit possible. Il est possible à l'aide des outils présents sur le marché de créer des impacts de l'ordre d'une dizaine de micromètres de diamètre. Pour une extraction de tissus cellulaire de grande surface (plusieurs dizaine voire centaine de  $\mu\text{m}^2$ ), l'opération d'impulsion laser à travers le polymère est répétée autant de fois que nécessaire afin de couvrir la totalité de la zone à extraire.

L'instrument utilisé lors de ces travaux est un ArcturusXT™ LCM System (Figure 5). Il est composé de deux parties principales. La première est un microscope Nikon ECLYPSE Ti qui constitue toute la partie inférieure de l'instrument. Ce microscope est équipé d'une platine amovible et de plusieurs objectifs (x2, x10, x40, x60). L'observation des échantillons se fait à l'aide d'une caméra reliée au poste de commande de l'appareil. La seconde partie du microdissecteur est constituée d'un laser solide infrarouge à faible puissance et d'un système muni d'un bras mécanique permettant de disposer les capsules de prélèvement sur l'échantillon.



**Figure 5 - Instrument de microdissection à capture laser (LCM) ArcturusXT™ LCM System.**

Pour réaliser les prélèvements par microdissection laser, il faut utiliser des capsules de prélèvement, dites «caps» (Figure 6a). Ces «caps» sont constituées d'un plastique transparent dans la masse, et recouvertes d'un fin film de polymère thermosensible sur la face inférieure. Lorsque des particules doivent être prélevées, la «caps» de prélèvement est déposée par le bras mécanique au-dessus de la zone d'intérêt. La «caps» est équipée de rails sur la face inférieure, ce qui permet d'éviter tout contact entre l'échantillon et le polymère. Pour réaliser un prélèvement, il faut ensuite effectuer une impulsion laser au travers de la «caps», ce qui va faire fondre le polymère thermosensible sur les particules d'intérêt. Enfin, quand la «caps» est retirée par le bras mécanique, les particules sont ainsi piégées dans le polymère.



**Figure 6 - «Caps» de prélèvement par microdissection (Arcturus® capsure® HS LCM Caps).**

Lorsque les prélèvements par microdissection sont réalisés, les «caps» peuvent ensuite être conditionnées à l'aide de bague d'extraction dans lesquelles est injecté un tampon de lyse. Ou bien, comme dans notre cas, nous protégeons directement le matériel biologique récolté à l'aide de capsules Eppendorf vierges de tout ADN, DNase, RNase et inhibiteurs de PCR comme présenté Figure 6b. Les échantillons sont ensuite transférés à un laboratoire de microbiologie pour analyse génomique.

La technique de microdissection laser n'est pas commune pour l'isolation de bactérie unique et est difficile à mettre en place. Jusqu'à ce jour, un seul rapport fait état de prélèvement de cellules uniques par microdissection laser [10]. À la différence de nos expériences, ils utilisent comme surface de dépôt un tissu clinique sur lequel ils viennent déposer les bactéries. Les bactéries et leur support sont ensuite décollés et catapultés par impaction laser pour atterrir dans un tampon de lyse. Dans notre cas, les bactéries sont déposées par sédimentation sur le support en  $\text{CaF}_2$  (transparent dans la zone spectrale observée par microspectroscopie Raman) puis ces agents microbiologiques sont piégés dans la «caps» d'extraction.

Une présentation plus détaillée sur l'optimisation du prélèvement par microdissection laser sera donnée plus tard avec les résultats expérimentaux.

### 3. Les outils chimiométriques

Historiquement, la chimiométrie est née de la nécessité d'expliquer de manière simple et efficace un phénomène complexe en chimie analytique. À cause de la complexité des jeux de données et de leur taille très conséquente, les informations utiles sont bien souvent difficiles à extraire. Les chimiométriciens ont donc décidé de mettre à profit leurs connaissances pour trouver un moyen simple de représenter les informations extraites de ces jeux de données. Ainsi, la chimiométrie est définie comme une discipline de la chimie faisant appel aux mathématiques, aux statistiques et à la logique formelle pour modéliser et sélectionner de manière optimale les procédures expérimentales. Le but étant d'extraire un maximum d'informations chimiques pertinentes en analysant les données chimiques pour ainsi mieux connaître et comprendre les systèmes chimiques étudiés [11].

Ainsi depuis le début des années 1970, la chimiométrie s'est fortement développée notamment grâce à l'avènement des ordinateurs et aux progrès faits en terme de vitesse de calcul. Aujourd'hui, de plus en plus d'industriels et d'autorités, préconisent l'utilisation de ces outils afin d'accéder à une meilleure maîtrise de leurs procédés et obtenir un meilleur contrôle qualité des produits finis ou en cours de fabrication. Grâce à ces outils, il est possible d'obtenir des résultats en temps réel, permettant ainsi une prise de décision en conséquence.

Dans cette partie il n'est pas question de détailler tous les outils chimiométriques aujourd'hui disponibles mais simplement de présenter les méthodes qui ont été utilisées pour l'analyse des spectres Raman dont les résultats seront donnés dans les chapitres suivants. Nous allons donc voir les différents prétraitements applicables aux spectres Raman, une présentation de l'analyse en composantes principales qui est un outil d'exploration simple mais puissant, ainsi qu'une présentation des outils de classification qui serviront à la construction d'un modèle prédictif.

#### **3.1. Le prétraitement des données spectrales**

Le prétraitement des données spectrales est une étape à ne pas sous-estimer. C'est à ce moment précis que commence l'analyse du jeu de données acquis. L'application d'un mauvais prétraitement ou même son absence peut masquer une information présente dans

le jeu de données, rendant les conclusions incomplètes ou erronées. Il existe de très nombreuses façons de prétraiter un jeu de données spectrales. Cela va dépendre du type de données à traiter, du domaine spectral étudié et de ce que l'on veut obtenir. Un prétraitement optimal permet de corriger les données sans dénaturer le spectre. Certaines informations indésirables peuvent être ôtées des spectres et certaines informations « cachées » peuvent être révélées.

Le spectre Raman de bactérie unique peut être affecté par la taille des particules ou tout simplement la fluorescence de celles-ci, ce qui peut engendrer des interférences sur le spectre. C'est pourquoi l'utilisation de prétraitements spectraux est préférable pour l'analyse des spectres et une potentielle discrimination entre espèces.

### **3.1.1. Les techniques de normalisation**

Le but des méthodes de normalisation est la réduction des variations entre les spectres liées à des phénomènes non pas chimiques ou biologiques mais physiques comme la diffusion ou encore les différences de volumes sondés. Deux méthodes sont fréquemment rencontrées et utilisées pour l'analyse spectroscopique : les normalisations *Standard Normal Variate (SNV)* et *Multiplicative Scatter Correction (MSC)*.

La correction *Standard Normal Variate* va avoir pour effet de supprimer les interférences liées à la dispersion et à la taille des particules, soit les informations physiques, de l'échantillon. Il sera alors possible de diminuer l'effet lié à la variation incontrôlée de l'intensité des spectres. Cette méthode est basée sur le calcul de l'écart-type et de la moyenne de chacun des spectres. Chacun des spectres est corrigé selon l'équation 2.

$$x_{i,corr} = \frac{x_i - \bar{x}}{s}$$

(Eq.2)

où  $\bar{x}$  est la moyenne des intensités Raman du spectre considéré,  $x_i$  l'intensité Raman à corriger à la longueur d'onde  $i$ ,  $s$  l'écart type à la moyenne des intensités Raman du spectre et  $x_{i,corr}$  l'intensité Raman corrigée.

Avec la technique *Multivariate Scatter Correction* comme pour la correction *SNV*, le but va être de corriger l'effet des différences de diffusion liées à la variation en taille des particules. Pour pouvoir appliquer cette correction, il sera nécessaire de posséder un spectre de référence ou à défaut le spectre moyen du jeu de données à disposition. Le principe de la correction *MSC* est la régression de chaque spectre  $\mathbf{x}_i$  sur le spectre de référence  $\mathbf{x}_{Ref}$ . Les coefficients  $\mathbf{b}_0$  et  $b_1$  de cette régression linéaire, dans l'équation 3 nous permettent d'obtenir les facteurs additifs et multiplicatifs de la diffusion

$$\mathbf{x}_i = b_1 \mathbf{x}_{Ref} + \mathbf{b}_0 + \mathbf{e}_i \quad (\text{Eq.3})$$

où  $\mathbf{b}_0$  et  $b_1$  sont les coefficients de la régression calculés,  $\mathbf{x}_{Ref}$  est le spectre de référence,  $\mathbf{x}_i$  le spectre Raman à corriger. Ensuite les spectres sont normalisés comme décrit dans l'équation 4.

$$\mathbf{x}_{i,corr} = \frac{\mathbf{x}_i - \mathbf{b}_0}{b_1} \quad (\text{Eq.4})$$

Au cours de ces travaux, les deux méthodes seront comparées afin de déterminer laquelle des deux est la plus performante pour la construction d'un modèle de classification chimiométrique.

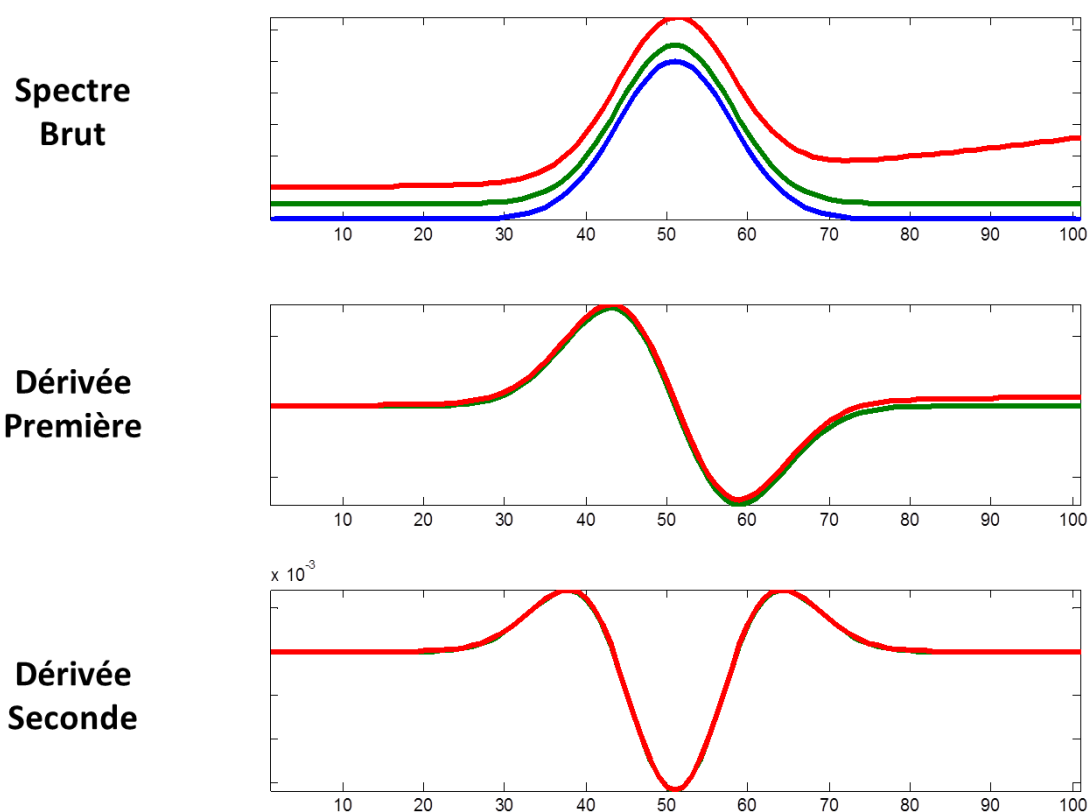
### **3.1.2. Les méthodes de correction de ligne de base**

La correction de ligne de base est utilisée lorsque les spectres présentent des déviations de la ligne de base. En spectroscopie Raman, cet effet peut être dû à la présence d'un fond de fluorescence.

En règle générale, la plupart des techniques de correction de ligne de base se basent sur l'ajustement d'un polynôme de degré  $k$ , puis ce polynôme est ensuite soustrait au spectre à corriger. Un des problèmes majeur avec ces méthodes est la détermination du degré  $k$  du polynôme à soustraire, c'est pourquoi les techniques de dérivations sont préférentiellement utilisées pour la correction de ligne de base.

La méthode la plus couramment utilisée pour réaliser une dérivation des spectres est l'algorithme de Savitzky-Golay [12]. L'algorithme calcule la dérivée pour chacune des variables  $i$  du spectre. Il va dans un premier temps ajuster un polynôme de degré  $K$  sur une fenêtre de  $N$  points autour du point  $i$ , puis le polynôme est dérivé au point  $i$ . Le spectre sera balayé en faisant glisser la fenêtre jusqu'à obtenir le spectre dérivé complet.

Sur la Figure 7 sont représentés les différents effets qu'entraîne l'application d'une dérivée première ou seconde sur des spectres présentant une augmentation de la ligne de base. Cette augmentation peut se traduire par deux effets possibles : un effet additif représenté par le spectre vert et un effet multiplicatif représenté par le spectre rouge, voire quelques fois les deux.



**Figure 7 - Représentation des effets de la dérivation sur les spectres. En bleu un spectre ne présentant pas de déviation de ligne de base, en vert un spectre présentant une augmentation de la ligne de base (effet additif), en rouge un spectre présentant une double augmentation de ligne de base (effet additif et multiplicatif).**

Généralement, lorsque qu'une méthode de dérivation est appliquée sur des spectres, une diminution du rapport signal sur bruit est observée et la qualité spectrale est affectée, ce qui impacte directement les performances des modèles. L'avantage de l'algorithme de Savitzky-Golay réside dans le lissage réalisé lors du balayage du spectre réduisant ainsi le bruit apporté lors de la dérivation. On peut également utiliser l'algorithme pour faire un simple lissage du spectre en fixant l'ordre de la dérivée comme nul.

### 3.2. *Analyse en composantes principales*

L'Analyse en Composantes Principales (ACP ou PCA de l'anglais *Principal Component Analysis*), est une technique reconnue comme étant la base de l'analyse multivariée et de la chimiométrie. À partir d'un jeu de données complexe de grande dimension, il est possible de condenser l'information et d'extraire ainsi quelques composantes, facilitant ainsi l'interprétation et la représentation de ces données sans perte d'information [13]. Le concept d'analyse en composantes principales fut initialement formulé au début du XX<sup>e</sup> siècle par un mathématicien britannique, Karl Pearson [14]. Mais ce n'est que plusieurs décennies plus tard que l'ACP telle que nous la connaissons aujourd'hui a été achevée. Dans les années 1930, les travaux de développement mathématiques de H. Hotelling [15], et dans les années 1970, les travaux de S. Wold [16] sur l'algorithme d'estimations non-linéaires itératives par les moindres carrés partiels, ou NIPALS (*Non-linear Iterative Partial Least Squares*) auront ainsi permis de finaliser l'analyse en composantes principales. L'ACP se classe dans les méthodes exploratoires et n'a pour but que de permettre à l'analyste d'appréhender au mieux le jeu de données sans caractère quantitatif.

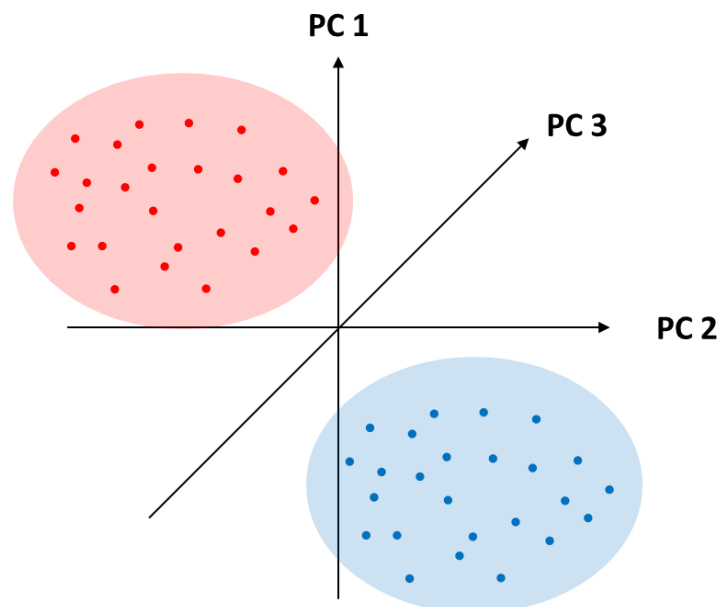
Considérons un jeu de données multidimensionnelles  $\mathbf{X}$  composé de  $m$  spectres (considérés comme  $m$  individus) à  $n$  variables (longueur d'onde ou déplacement Raman). Ce jeu de données est représenté sous la forme matricielle dont chaque ligne représente un spectre et chaque colonne représente l'intensité du signal mesuré à chacune des variables spectrales. La matrice  $\mathbf{X}$  peut alors être décomposée comme présenté dans l'équation 5.

$$\mathbf{X}_{(m \times n)} = \mathbf{T}_{(m \times k)} \mathbf{P}^t_{(k \times n)} + \mathbf{E}_{(m \times n)}$$

(Eq.5)



Grâce à cette décomposition, il est désormais possible de représenter les  $m$  individus dans un nouvel espace à  $k$  dimensions et non plus à  $n$  dimensions. Il est important de noter que  $\mathbf{P}$  est la matrice contenant l'ensemble des  $k$  vecteurs (les  $k$  composantes principales) à  $n$  dimensions constituant la nouvelle base orthonormée, aussi appelés « *loadings* ». La matrice  $\mathbf{T}$  contient les coordonnées factorielles des  $m$  individus dans le nouveau repère à  $k$  composantes, aussi appelées « *scores* ». Enfin, la matrice  $\mathbf{E}$ , de mêmes dimensions que  $\mathbf{X}$ , représente les résidus, c'est-à-dire l'information restante qui n'a pas été expliquée par les  $k$  composantes principales calculées. Chacune des  $k$  composantes calculées est orthogonale aux autres, ainsi la variance expliquée par une composante ne peut pas être expliquée par une autre. Dans un repère en 3 dimensions, donc à 3 composantes, la représentation des individus se ferait comme représenté ci-dessous Figure 8.



**Figure 8 – Graphique des *scores* : projection des individus dans le nouveau repère définis par les nouvelles variables, les composantes principales**

Lors d'une analyse en composantes principales, la première composante (PC1) explique le maximum de la variance présente dans le jeu de données initial. Ensuite, la deuxième composante principale (PC2) va expliquer le maximum de la variance restante, c'est-à-dire la variance qui n'a pas encore été expliquée par PC1. Il est alors possible de calculer par itération autant de composantes principales que souhaité tout en considérant que

le nombre de composantes principales doit être inférieur au nombre de variables initiales dans le jeu de données ( $k < n$ ) ou au nombre de spectres.

En pratique, le calcul des composantes principales est réalisé à l'aide de l'algorithme NIPALS (de l'anglais *Nonlinear Iterative Partial Least Square*) et il est donc possible de représenter le jeu de données de départ  $\mathbf{X}$  comme la combinaison des différentes composantes principales calculées comme décrit dans l'équation 6 :

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^t + \mathbf{t}_2 \mathbf{p}_2^t + \dots + \mathbf{t}_k \mathbf{p}_k^t + \mathbf{E} \quad (\text{Eq.6})$$

$\mathbf{t}_1$  et  $\mathbf{p}_1$  représentent respectivement les vecteurs des *scores* et des *loadings* pour PC1,  $\mathbf{t}_2$  et  $\mathbf{p}_2$  quant à eux représentent respectivement les vecteurs des *scores* et des *loadings* pour PC2. On remarque qu'il restera toujours un résidu  $\mathbf{E}$  qui ne sera pas expliqué pour le nombre de composantes fixées. Après avoir calculé un nombre suffisant de composantes, ce résidu ne représentera plus que du bruit et n'aura aucune utilité.

### **3.3. Les méthodes de classification**

Les méthodes de classification sont aussi nombreuses que variées mais elles ont toujours pour objectif de répartir les individus présents dans un jeu de données en différentes catégories ou classes. Pour regrouper ces individus en classe il faut une stratégie de classification et donc déterminer des critères. En microbiologie, on retrouve ce concept avec la taxonomie. En chimiométrie, les méthodes de classification reposent sur deux grands principes opposés : les méthodes non supervisées et les méthodes supervisées.

#### **3.3.1. Les méthodes de classification non-supervisées**

Les méthodes de classification non supervisées (ou *clustering*) ont pour but de regrouper ou classer les individus en catégorie sans à priori préalable sur le jeu de données. L'opérateur va pouvoir utiliser plusieurs critères statistiques et commencer à optimiser sa classification.

Certaines méthodes comme la classification ascendante hiérarchique (CAH) se basent sur le calcul d'un tableau de distance entre les individus. Les deux individus les plus proches sont considérés de la même classe et forme désormais un nouvel individu moyen qui sera le barycentre des deux individus précédents. Les deux individus suivants les plus proches vont être couplés de la même façon pour former de nouveau un individu moyen et ainsi de suite jusqu'à obtenir un arbre binaire ou encore appelé dendrogramme comme représenté sur la Figure 9.

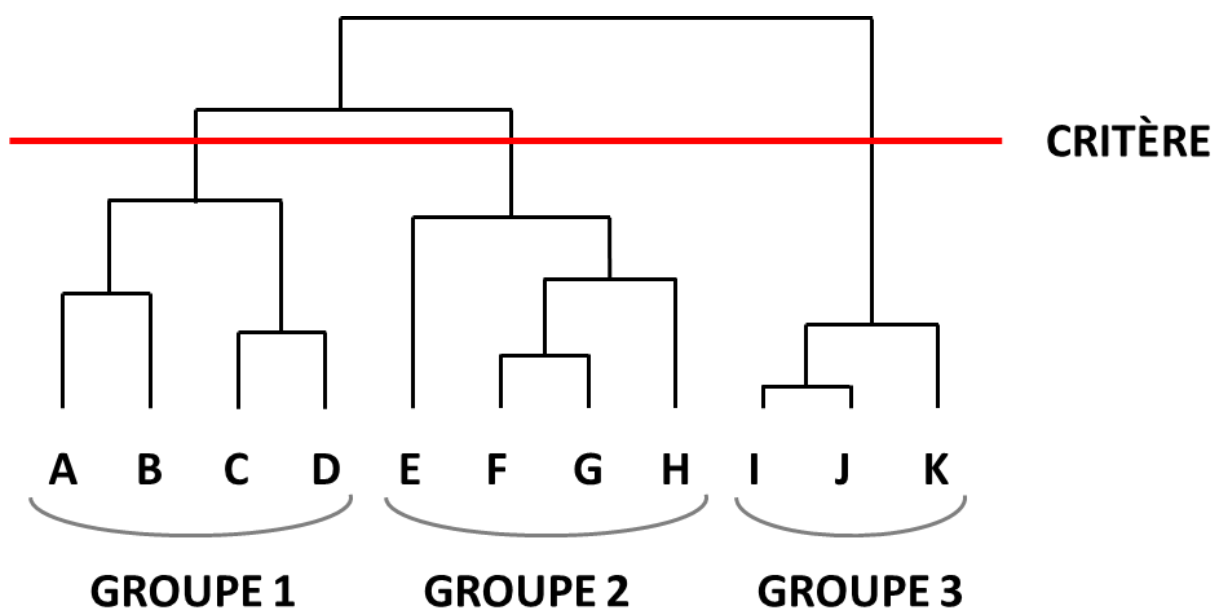


Figure 9 - Schéma d'une représentation en dendrogramme

Pour définir la proximité de deux individus, un critère de distance peut être utilisé comme par exemple la distance euclidienne dans un espace à  $k$  composantes. Au départ il y a autant de classe que d'individus puis le dendrogramme se restreint pour ne finalement proposer qu'une seule classe. En fixant un critère, comme le saut de Ward ou encore la statistique du « *gap* », il est possible de déterminer un nombre final de classes représentatives du jeu de données et de déterminer les individus composants chacune de ces classes.

Une autre méthode très connue dans le domaine des méthodes non supervisées est la méthode des *K-means*. Cet algorithme va chercher à minimiser l'inertie interclasse jusqu'à la convergence. Initialement on suppose un nombre de classe, puis, aléatoirement des points

sont choisis comme centre de la classe. Les individus à classer sont regroupés avec le centre de classe le plus proche. De nouveaux centres sont définis comme étant les barycentres des nouvelles classes créées. Le processus est répété jusqu'à convergence. L'inconvénient majeur de cette technique est que nous n'atteignons pas forcément le minimum global. Il est donc possible de déterminer différentes classes pour un individu selon le choix initial des centres de classe.

### **3.3.2. Les méthodes de classification supervisées**

Les méthodes de classification supervisées ont pour objectif de construire un modèle de classification à partir de données dont on connaît la classe au préalable. Le but est d'entraîner le modèle pour qu'il soit le plus complet et robuste possible et qu'il puisse associer un échantillon inconnu à l'une des classes prédéfinies. Il existe plusieurs méthodes de classification supervisée connues comme SIMCA, PLS-DA, KNN ou encore SVM.

#### **La méthode SIMCA (*Soft Independent Modelling of Class Analogy*).**

Avec SIMCA, une ACP est effectuée sur les individus correspondant à chaque classe dans le jeu de données. Un nombre suffisant de composantes principales est retenu, permettant d'expliquer la majeure partie de la variance intra-classe. Le nombre de composantes principales peut être différent pour expliquer chaque classe. De nouveaux échantillons peuvent être projetés sur chacun des modèles ainsi développés, et l'appartenance à une des classes est évaluée en fonction de la distance à ces modèles. Une mesure de cette distance peut être basée sur la statistique Hotelling de l'erreur.

#### **La méthode PLS-DA (*Partial Least Square Discriminant Analysis*).**

La régression PLS est une méthode de régression multiple alliant la décomposition en composantes principales de l'ACP aux principes de régression multilinéaire. Avec la méthode PLS-DA, le but n'est pas d'expliquer ou de modéliser une variable quantitative mais plutôt une variable qualitative qui est la classe de l'individu.

### **La méthode k-NN (*k-Nearest Neighbors*)**

La méthode k-NN, ou *k* plus proches voisins en français, utilise un algorithme intuitif et simple d'utilisation. Pour chaque nouvel individu à identifier, il faut commencer par identifier l'ensemble des *k* plus proches voisins, autrement dit, il faut paramétrer le rayon dans lequel l'analyse va être menée. Puis la classe affectée au nouvel individu est la classe majoritairement présente de cet ensemble des *k* plus proches voisins.

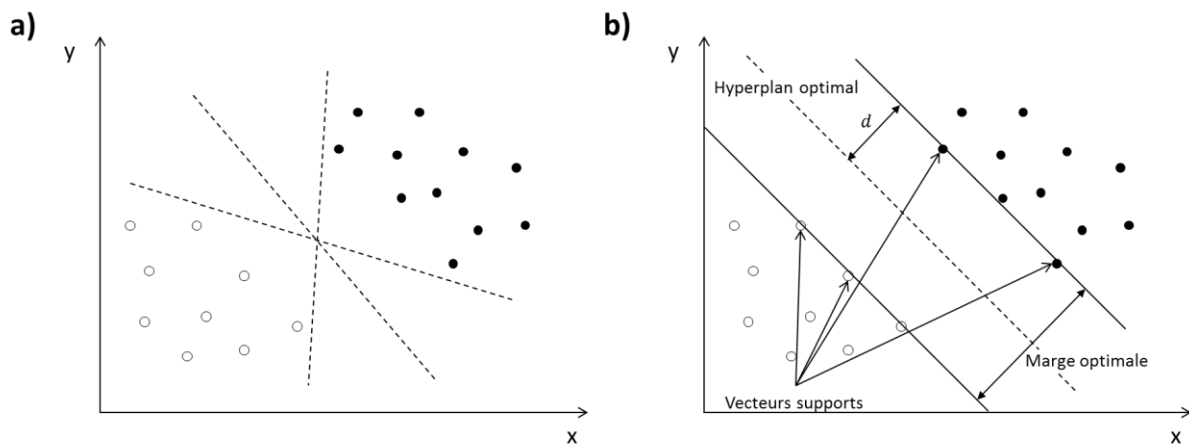
Ici ce ne sont que quelques exemples de classifications supervisées qui ont été résumés succinctement mais il en existe bien d'autre comme les arbres de décision, la classification naïve bayésienne, les réseaux de neurones, les modèles gaussiens ou encore les machines à vecteurs de support. Cette dernière méthode est celle qui a retenu notre attention lors du projet, qui a été utilisée sur les spectres Raman des bactéries et mérite donc d'être présentée de manière plus détaillée.

### **La classification SVM**

Les machines à vecteurs de support ou séparateurs à vaste marge (de l'anglais *Support Vector Machine*) noté SVM, est une technique relativement récente datant des années 1990, qui a émergée suite aux travaux de V. Vapnik [17]. Cette méthode peut être utilisée pour résoudre des problèmes de régression ou de classification. Comme pour la plupart des méthodes de classification supervisée, il faut utiliser un ensemble d'étalonnage, un *training-set*, basé sur des échantillons dont la classe est connue. Dans notre cas, nous travaillons sur des échantillons de bactéries qui sont cultivées dans les conditions parfaitement maîtrisées, enlevant ainsi tout problème de contamination des souches par un autre individu. Le modèle doit être ensuite validé par un ensemble test, un *test-set*, c'est-à-dire un jeu de spectres indépendant, dont la classe est également connue. On va ainsi pouvoir juger les performances du modèle pour qu'il puisse être utilisé par la suite. L'intérêt pour certains chercheurs, n'est pas seulement d'assigner une classe ou une autre à des échantillons inconnus, mais également d'identifier des échantillons qui ont pu être mal identifiés par une autre méthode d'analyse.

Les SVM reposent sur deux points qui permettent de traiter des problèmes de discrimination linéaire et non linéaire. Ces deux points sont la notion de marge maximale et la notion de fonction de noyau.

Dans le cas d'une discrimination linéaire, supposons qu'il soit possible de séparer deux classes par un classifieur linéaire, encore appelé hyperplan. Or, en réalité, il n'existe pas qu'un seul classifieur linéaire possible mais plutôt une infinité (Figure 10a). C'est pourquoi la notion de marge maximale a été introduite. La marge est la plus courte distance entre l'hyperplan et les plus proches individus de chacune des classes. Ces individus sont appelés les vecteurs de support. Grâce à l'optimisation de la marge  $m$ , c'est-à-dire en la maximisant, il est possible de déterminer la marge optimale et donc l'hyperplan optimal pour séparer les deux classes représentées (Figure 10b).



**Figure 10 – a) Représentation des classifieurs possible pour séparer les deux groupes d'individus et b) représentation de l'optimisation de la marge maximale pour le choix du classifieur optimal en se basant sur les vecteurs de support**

Pour la SVM linéaire, la règle de classification est qu'il faut que tous les individus du *training set* se situent du bon côté de la bordure définie par la marge optimale comme écrit dans l'équation 7.

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) \geq 1$$

(Eq.7)

où  $y_i$  est la classe de l'individu  $i$ ,  $\mathbf{w}$  est le vecteur normal à l'hyperplan optimal et  $b$  le biais. La distance entre les points en bordure de marge et l'hyperplan optimal est notée  $d$ . Elle est définie comme l'inverse de la norme de  $\mathbf{w}$ . Ainsi dans l'équation 8, la marge est égale à :

$$d = \frac{1}{\|\mathbf{w}\|} \quad \text{et} \quad m = \frac{2}{\|\mathbf{w}\|}$$

(Eq.8)

L'optimisation de la largeur de marge est obtenue en résolvant le problème d'optimisation quadratique contraint présenté dans l'équation 9.

$$\begin{aligned} & \text{minimisation de } \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{soumis à } y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) - 1 \geq 0 \end{aligned}$$

(Eq.9)

Selon les théories d'optimisation [18] la forme duale du Lagrangien peut être utilisée pour résoudre ces problèmes.

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x}_i + b = \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i \cdot \mathbf{x} + b \quad \text{avec } 0 \leq \alpha_i$$

(Eq.10)

Lorsque  $\alpha_i > 0$ , l'individu  $i$  n'intervient pas dans le calcul du modèle, lorsque  $\alpha_i = 0$ , l'individu  $i$  est un vecteur support. Dans le cas présenté précédemment, il s'agissait d'une séparation de deux classes par un classifieur linéaire. Dans certains cas, les données ne sont pas linéairement séparables. Pour résoudre ces problèmes, Cortes et Vapnik proposent d'utiliser une technique dite de marge souple [17]. Pour cela il faut tolérer de mauvais classements et introduire des variables d'écart afin de modéliser l'erreur  $\xi_i$  de chacun des individus. Si  $\xi_i = 0$  alors l'individu est correctement assigné, si ce n'est pas le cas, l'erreur est prise en compte dans le modèle. L'équation 10 précédente peut donc s'écrire de la manière suivante :

$$\begin{aligned} & \text{minimisation de } \left( \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right) \\ & \text{soumis à } \xi_i + y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad \text{et} \quad \xi_i \geq 0 \end{aligned}$$

(Eq.11)

Un nouveau facteur  $C$  appelé méta-paramètre de régularisation pour la pénalisation des erreurs est donc introduit.

Pour traiter des cas où les données ne sont pas linéairement séparables, la deuxième idée clé des SVM est de transformer l'espace initial de représentation des données en un espace de plus grande dimension dans lequel une séparation linéaire est possible. Cette transformation va se faire par le biais des fonctions noyau (*Kernel functions* en anglais).

En classification non-linéaire, le Lagrangien s'écrit selon l'équation 12 :

$$f(\phi\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x}_i + b = \sum_{i=1}^n y_i \alpha_i \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) + b \quad \text{avec } 0 \leq \alpha_i \leq C$$

(Eq.12)

Le produit scalaire  $\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x})$  est la fonction noyau  $K$ . Les fonctions RBF (*Radial Basis Functions* en anglais) en particulier les fonctions Gaussiennes sont largement utilisées et présentent généralement de bons résultats. La fonction noyau RBF gaussienne est donnée dans l'équation 13, où  $G$  est le paramètre contrôlant la largeur du noyau gaussien centré sur chaque vecteur de support.

$$K(\mathbf{x}_i, \mathbf{x}_{i+1}) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_{i+1}) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_{i+1}\|}{2\sigma^2}\right) = \exp(-G\|\mathbf{x}_i - \mathbf{x}_{i+1}\|)$$

(Eq.13)

Un des inconvénients majeurs des SVM est que lors du calcul du modèle il faut optimiser les méta-paramètres de la fonction de kernel ( $C$  et  $G$ ) dans le but de contrôler le risque de surentrainement du modèle et la complexité des frontières. Si la valeur de  $C$  est trop faible, la marge sera trop importante et la discrimination ne sera pas possible, en revanche, si la valeur de  $C$  est trop importante l'erreur est diminuée mais la marge sera très faible et la limite plus complexe. Quand  $G$  est petit, le classifieur a une tendance plus ou moins linéaire, alors que quand il est grand, il devient tout de suite plus complexe. On notera également que plus  $G$  est grand, plus le nombre de vecteurs supports augmente, et l'influence de  $C$  est très faible. C'est pourquoi il faut trouver le bon compromis pour  $C$ ,  $G$  et le nombre de vecteurs



supports. L'ajustement se fait par essais successifs sur une grille d'optimisation avec comme critère la minimisation de l'erreur en validation croisée [19].

## **4. Etat de l'art de l'analyse par spectroscopie Raman et application de la chimiométrie pour la classification de bactéries**

Depuis plusieurs années, la spectroscopie et la chimiométrie sont utilisées afin d'identifier et de classer les bactéries. La nécessité d'une détection rapide des agents microbiologiques dans l'environnement a permis le développement des techniques utilisées de nos jours. A l'origine, la détection d'agent microbiologique s'est révélée nécessaire dans le domaine de l'agroalimentaire. Les industriels se devaient de garantir qu'un produit distribué n'est pas contaminé et qu'il soit propre à la consommation. C'est aussi le cas dans le milieu médical, à cause des risques d'infections nosocomiales et pour le diagnostic de certaines contaminations par agent bactériologique.

Les premiers travaux menés sur l'analyse des bactéries par spectroscopie datent des années 1950 [20]. A cette époque, c'est la spectroscopie infrarouge qui est largement utilisée. Des chercheurs ont montré qu'il est possible de trouver des différences spectrales entre deux bactéries de la même souche. Durant leurs analyses, les scientifiques n'analysaient pas directement les cellules, mais procédaient à des extractions des composants présents dans les bactéries [21]–[23]. Les premiers résultats sont prometteurs, car déjà, des différences spectrales apparaissent entre les espèces. Depuis de nombreuses techniques ont été testées, toujours dans le but de trouver des différences entre les espèces bactériennes. On a pu voir des analyses faites par spectrométrie de masse [24] ou de fluorescence [25]. Mais les deux techniques les plus employées à ce jour sont les spectroscopies infrarouges et Raman.

La spectroscopie infrarouge a démontré son potentiel pour l'analyse des bactéries. Le caractère non destructif de la méthode et l'utilisation de cellules entières et intactes sont mis en avant [26]. L'analyse des bactéries par spectroscopie infrarouge n'est possible que sur un ensemble d'individus et non pas sur cellule unique pour des raisons évidentes de résolution spatiales. On parle d'analyse en *bulk* de bactéries cultivées, souvent sous forme de suspension

bactérienne [27]. L'analyse *bulk* présente deux avantages. Le premier est directement lié à la quantité de cellules analysées. Plus le nombre de cellules analysées est important, plus le signal observé sera intense. L'autre avantage de ce type d'analyse réside dans le fait que le signal observé est un signal moyen de l'échantillon. Donc si une légère contamination a lieu lors de la manipulation de l'échantillon, celle-ci sera négligeable par rapport à la quantité de cellules de la même espèce présente dans l'ensemble de cet échantillon.

Plus tard, grâce au développement constant des systèmes optiques, des lasers et des microspectromètres, l'analyse de cellules uniques devient alors possible. En 1990, Puppels et al. développent un microscope Raman confocal permettant d'acquérir des spectres Raman de cellules uniques humaines [28]. Les premières études de bactéries à l'aide de systèmes Raman apparaissent dans les années 1990, avec l'étude de bactéries par spectroscopie Raman résonante Ultraviolet (UVR), mais les données sont toujours acquises sur des suspensions bactériennes. Il faudra attendre plusieurs années avant que les microscopes confocaux se généralisent dans les laboratoires de recherche. Puis, au fur et à mesure des années, les systèmes d'acquisition se sont généralisés et, aux vues du nombre conséquent des publications sur la thématique, la microspectroscopie Raman s'est majoritairement imposée laissant la spectroscopie infrarouge quelque peu en retrait. Des techniques comme le SERS (*Surface Enhanced Raman Spectroscopy*) ont également été investiguées [29] sur les bactéries, des spores ou encore des cellules fongiques et sans améliorations des modèles.

Les techniques de préparation des échantillons se sont améliorées et l'optimisation constante des protocoles d'acquisition permettent désormais une analyse par microspectroscopie Raman fiable, rapide et répétable. L'analyse des spectres a également évolué. Avec le développement constant des ordinateurs et des techniques d'analyses, on est passé de la simple attribution de bandes Raman à l'analyse multivariée des données spectroscopiques.

Depuis de nombreuses années, le nombre de publications parues, rapportant les résultats obtenus sur l'analyse par spectroscopie Raman et les tentatives de classification, est très important [30]–[32]. Les résultats présentés jusqu'à ce jour sont convainquant avec des modèles de classification présentant un bon pourcentage de classification [33]–[36]. Néanmoins, le problème majeur de tous ces travaux est le manque de considération pour

l'influence des facteurs environnementaux sur les organismes bactériologiques. Les travaux reportés proposent l'analyse de bactéries ayant subis des conditions de culture parfaitement maîtrisées et n'ayant subis aucun stress causé par des variations climatiques comme peuvent subir des bactéries extérieures. En 2005, Harz et al. [37] propose d'étudier l'influence de variations de conditions de culture sur les spectres Raman des bactéries et l'influence sur les résultats obtenus par classification. Les résultats prouvent que les conditions de cultures ont une influence sur les spectres, que selon le modèle de classification choisi les variations ont une influence sur le pourcentage de bonne classification. Le modèle CAH ne permet plus l'identification des bactéries, alors qu'avec la méthode SVM, il est toujours possible de classer les bactéries.

Les méthodes chimiométriques pour la classification des bactéries ont été aussi largement étudiées que les techniques spectroscopiques elles-mêmes. Au début, les premières études ont été faites grâce à de l'analyse de données non supervisée. Les CAH et l'ACP ont été énormément utilisées mais force est de constater que ces méthodes ne sont pas suffisantes pour classer l'ensemble des individus d'un jeu de données. De nouvelles méthodes ont donc été développées et étudiées. Suite à la démultiplication des méthodes d'analyse, Schmid et al. [38] mènent une étude en 2009 et propose de comparer plusieurs modèles de classification sur un même jeu de données. Le but étant d'estimer lequel des algorithmes permet d'obtenir les meilleurs pourcentages de reconnaissance selon le prétraitement appliqué sur les spectres. Les résultats obtenus avec les méthodes QDA (Quadratic Discriminant Analysis), MDA (Multiple Discriminant Analysis) et kNN (k-Nearest Neighbors) présentent des pourcentages de classifications avoisinant les 80% de bonne reconnaissance, mais en moyenne, la méthode des SVM propose des taux de reconnaissance supérieurs aux autres méthodes, avec un pourcentage de reconnaissance de l'ordre de 87%.

Les bactéries de culture ont donc été largement étudiées par spectroscopie et les modèles de classification optimisés au fur et à mesure des études. Les avis de la communauté scientifique convergent dans le même sens : l'utilisation de la microspectroscopie Raman et les modèles de classification SVM sont parmi les outils les plus utiles pour caractériser et identifier de façon rapide et efficace les bactéries. Tous les travaux présentés à ce jour décrivent des résultats obtenus à partir de bactéries de culture. Or l'impact environnemental comme une exposition à des composés chimiques, aux rayonnements ultraviolets, aux

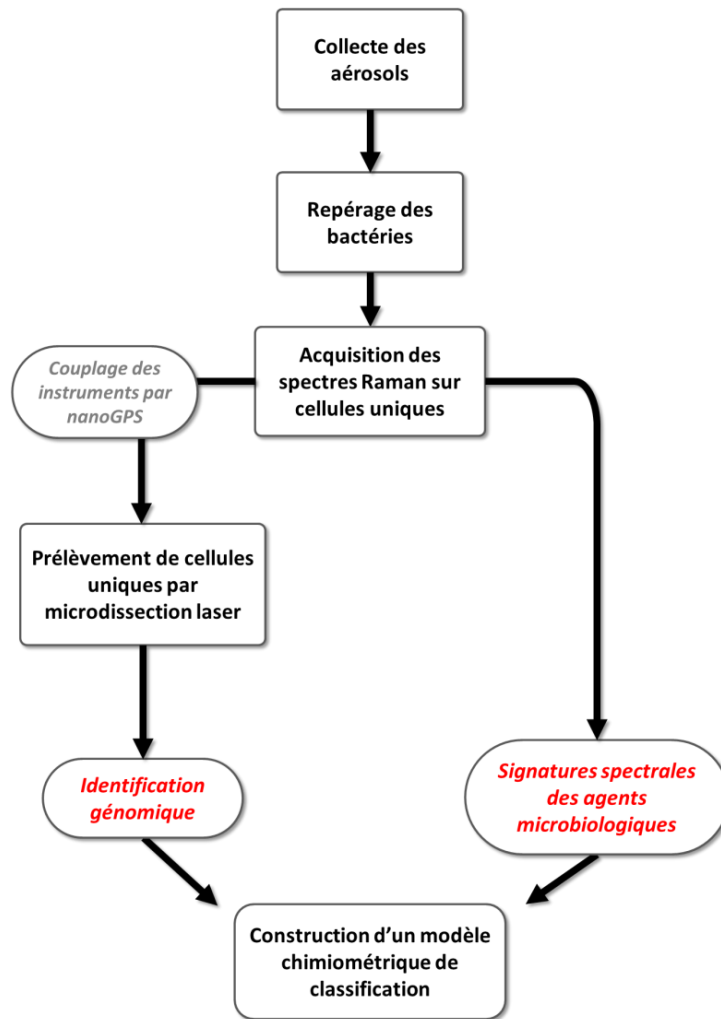
variations de température et d'hygrométrie vont avoir une influence sur les bactéries et donc le spectre, les modèles construits ne peuvent être utilisés pour identifier les micro-organismes naturels. C'est pourquoi, au cours de ces travaux, nous souhaitons mettre en place un protocole global qui permettra de construire un modèle de classification le plus complet possible à partir de bactéries naturelles issues de l'environnement.

# **Chapitre 2 – Développement d'un protocole original de classification, agents bactériens et échantillonnage**

Dans ce premier chapitre, nous présentons tout d'abord le protocole original que nous avons développé pour permettre une étude des agents bactériologiques présents dans les aérosols naturels. Ensuite, nous détaillons le matériel microbiologique étudié tout au long de ces travaux, puis, comment les échantillons de cultures analysés ont été préparés et comment nous avons réalisé les prélèvements d'aérosols pour isoler des bactéries naturelles. Enfin, nous nous intéresserons à l'analyse d'identification moléculaire, étape clé pour la construction d'un modèle de classification.

## **1. Protocole développé pour la construction d'un modèle de classification des agents bactériologiques**

Pour atteindre les objectifs du projet SIBIRAM, nous avons développé un nouveau protocole (Figure 11) permettant d'acquérir la signature spectrale Raman des microorganismes, suivi d'une extraction individuelle par microdissection laser des cellules analysées dans le but de réaliser une identification par analyse génomique sur cellule unique. L'étude des données spectroscopiques combinées aux résultats obtenus par l'analyse des séquences génomiques codant les ARN 16S, permettra la construction d'un modèle chimio-métrique de classification des microorganismes présents naturellement dans les aérosols de l'environnement.



**Figure 11 – Schéma de principe du protocole global mis en place pour répondre aux enjeux de la biosurveillance de l'air.**

La première étape du protocole correspond au développement d'une méthode de collecte et de dépôt des aérosols. Il est primordial que la génération des échantillons n'engendre pas de perturbations sur la suite des analyses effectuées sur les particules. La deuxième étape, tout aussi importante, correspond à la mise en place d'une méthodologie permettant un repérage efficace et systématique des agents microbiologiques présents dans les prélèvements d'aérosols. De par la nature complexe des aérosols naturels et la population variée des particules, des outils au-delà de la simple observation au microscope seront nécessaires pour le repérage.

Lorsque les cellules d'intérêt sont repérées, il faut procéder à l'optimisation de la mesure par microspectroscopie Raman, en veillant à ne pas dégrader le matériel biologique.

Aussi, des temps d'exposition trop longs ou des puissances de laser trop importantes peuvent entraîner la carbonisation des cellules ciblées. Les signatures spectrales des agents microbiologiques sont conservées durant l'attente des résultats issus de l'identification génomique. Chacune des cellules analysées, subit ensuite une étape de prélèvement individuel grâce à l'outil de microdissection laser. Pour assurer un retour fiable sur les cellules analysées, nous utiliserons le système nanoGPS, qui permet un couplage virtuel des différents instruments. Ce dispositif original a fait l'objet d'une collaboration entre le Laboratoire de Spectrochimie infrarouge et Raman (LASIR) et le département de Recherche et Développement de la société HORIBA Jobin Yvon. L'ADN des cellules ainsi microdissectées est ensuite extrait, amplifié et les fragments d'ARN 16S sont séquencés. Les résultats obtenus après séquençage sont assignés aux signatures spectrales correspondantes et le modèle chimioométrique de classification peut être construit.

## 2. Agents bactériologiques

Le terme « bactérie », aujourd'hui couramment employé, est un nom désignant un type d'organismes vivants microscopiques et procaryotes présents dans l'environnement. La structure cellulaire des procaryotes ne comporte pas de noyau et ces êtres microbiologiques sont unicellulaires. Depuis leur découverte et jusqu'à aujourd'hui, l'Homme n'a cessé d'étudier leur développement, leur comportement et leur pathogénie. Les premières observations de bactéries ont été faites lors de la seconde moitié du XVII<sup>ème</sup> siècle, mais ce n'est que deux siècles plus tard, grâce aux travaux menés par Louis Pasteur, que la bactériologie connaîtra une révolution. Récemment, dans les années 1990, la découverte d'un nouveau type de cellules procaryotes a mené les scientifiques à revoir les termes employés. Désormais, les êtres vivants sont classés en trois grands groupes : les Eubactéries, les Archéobactéries et les eucaryotes. Durant toute la suite des travaux de thèse, le terme « bactérie » au sens général sera utilisé.

### 2.1. *Classification des bactéries*

Les organismes procaryotes regroupent les organismes unicellulaires ne présentant pas de noyau individualisé, c'est à dire les bactéries et les archaebactéries.

En classification microbiologique traditionnelle, il existe différents niveaux de hiérarchie pour classer les bactéries. Au premier niveau de classification, nous allons retrouver le règne. Vient ensuite le domaine, le phylum, la classe, l'ordre, la famille, le genre et l'espèce. Cette dernière constitue l'unité de classification. Les bactéries sont désignées selon une nomenclature binomiale. Toutefois, il est souvent nécessaire de subdiviser une espèce en différentes sous-espèces. Dans le cas d'*Escherichia coli*, certaines sous-espèces peuvent présenter un caractère pathogène (5% des sous-espèces) alors que d'autres sont vitales (95% des sous-espèces). Pour cela, les scientifiques utilisent le nom du genre et le nom d'espèce (exemple : *Escherichia coli*).

En taxonomie biologique, la représentation phylogénétique (Figure 12) schématise cette façon de classer les êtres vivants en se basant sur des données moléculaires, comme la comparaison des séquences d'ARN. De par la diversité des êtres vivants rencontrés, ce classement est très difficile à effectuer. Celui-ci n'est pas arrêté et évolue constamment, au fur et à mesure que de nouvelles études et de nouvelles découvertes sont réalisées.

Selon le domaine de recherche, la classification va différer. À titre d'exemple, en bactériologie médicale, les bactéries peuvent être classées selon un caractère pathogène, c'est-à-dire, selon le type de maladie qu'elles vont induire, selon le mécanisme pathogène mis en jeu, ou selon un caractère clinique, c'est-à-dire les syndromes entraînés par la présence d'une bactérie.

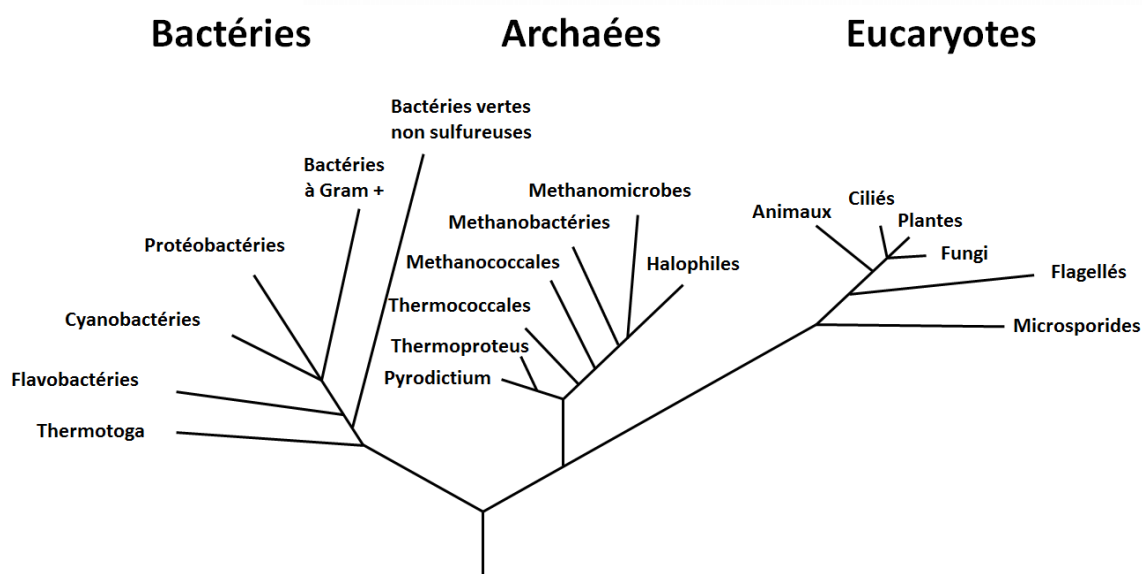


Figure 12 – Arbre phylogénétique simplifié du monde vivant selon C. Woese [39].



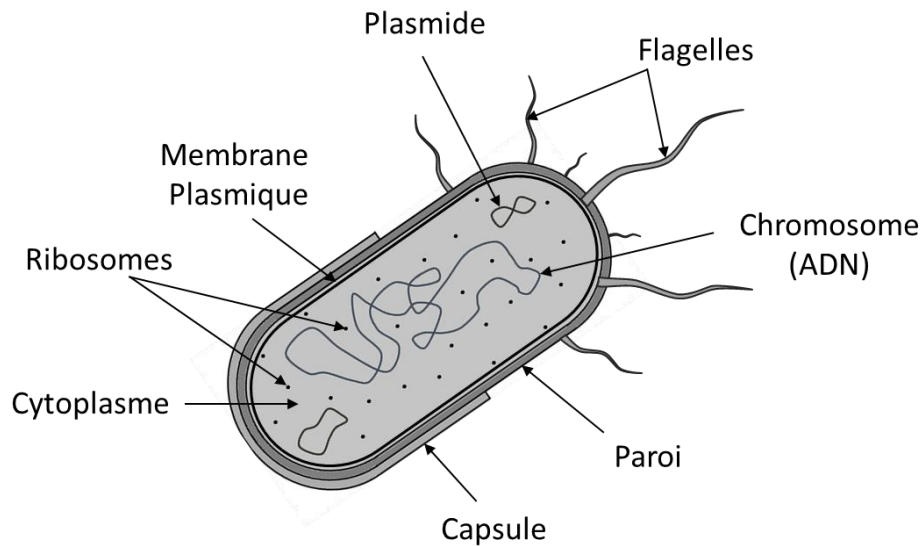
## **2.2. Morphologie et composition**

Les cellules bactériennes présentent une grande diversité de taille et de forme. La majorité des bactéries sont soit sphériques, on parle alors de coques ; soit en forme de bâtonnet, on parle alors de bacilles. La taille de ces micro-organismes est généralement comprise entre 0,5 et 5  $\mu\text{m}$  de longueur. Malgré leur taille réduite et leur caractéristique unicellulaire, une bactérie reste un micro-organisme complexe.

Les bactéries à Gram négatif possèdent deux membranes plasmiques (interne et externe) contrairement aux bactéries à Gram positif qui ne possèdent qu'une seule membrane plasmique interne. Pour différencier ces deux groupes de bactéries, il existe un test empirique appelé test de coloration de Gram, connu depuis la fin du XIX<sup>ème</sup> siècle. Ce test repose sur la réactivité de la paroi constituée d'une substance, la muréine qui est un peptidoglycane. La paroi est recouverte par une membrane externe chez les bactéries Gram négatif, tandis que les bactéries à Gram positif en sont dépourvues. À l'aide de ce critère ainsi qu'avec la reconnaissance de la morphologie, il est possible de renseigner l'ordre dont fait partie une bactérie rencontrée. Ce type de classification n'est pas sans conséquence dans le domaine médical. En effet, la résistance des bactéries et l'efficacité d'antibiotiques dépendent du type de bactérie.

Une bactérie possède un chromosome généralement unique portant la majorité des gènes. Généralement, lors d'une amplification en chaîne par polymérase, plus couramment appelé PCR (de l'anglais *Polymerase Chain Reaction*), ce sont ces gènes qui seront ciblés pour déterminer l'identité de la bactérie étudiée. Certains gènes ayant des fonctions particulières (résistance à un antibiotique, adaptation physiologique au milieu, etc.) sont cependant localisés sur des petites sections d'ADN circulaires libres appelées plasmides.

Tous ces éléments sont situés dans le cytoplasme qui peut être défini comme le contenu de la cellule. Composé à environ 70% d'eau, le cytoplasme des cellules bactériennes est un hydrogel colloïdal comprenant une phase dispersante (sels minéraux et composés solubles de nature lipoprotéique) et une phase dispersée formée de nucléoprotéines et de lipides. Le cytoplasme contient des ribosomes, des acides ribonucléiques et des substances



**Figure 13 - Schéma d'une cellule bactérienne.**

de réserves. Les ribosomes sont composés de protéines et d'ARN, dont le rôle est de synthétiser les protéines en décodant l'information contenue dans l'ARNm. Les plasmides sont de petits fragments d'ADN qui confèrent aux bactéries des résistances aux antibiotiques et aux métaux lourds. Les cils ou flagelles confèrent à la bactérie une certaine mobilité qui peut alors se déplacer sur la surface d'une gélose ou dans un milieu liquide.

### **2.3. Caractère pathogène de certains individus**

Le but d'un outil de détection en temps réel est de prévenir une contamination, de déployer tous les moyens nécessaires, et cela le plus rapidement possible pour stopper la pandémie, protéger les personnes qui ne sont pas encore touchées par l'agent infectieux et prendre en charge les personnes contaminées. Toutes ces mesures citées doivent être prises dans le cas d'une détection d'un agent bactériologique pathogène dans l'air. En France, dans le cadre du plan Biotox lancé en 2001, des stocks d'antibiotiques ont été constitués, avec différents plans de distribution de médicaments. Ce stock national permettrait d'assurer le traitement d'un million de personnes pendant une durée de 8 semaines, et le traitement devrait être remis à l'ensemble de la population exposée en moins de 24 heures [40].

Le caractère pathogène d'un individu bactériologique est donc le critère d'alerte principal pour une mise en quarantaine et une intervention directe des autorités sanitaires

sur un lieu contaminé. La pathogénicité sera dépendante du nombre de bactéries, des conditions d'ingestion ou bien encore de l'état de santé du patient. Certaines bactéries, même à très faibles doses chez des individus en bonne santé lors de l'infection, resteront pathogènes. Ainsi pour des souches possédant un même pouvoir pathogène, certaines espèces peuvent être plus ou moins virulentes. Pour des agents très virulents seulement quelques bactéries suffisent pour développer une infection alors que plusieurs milliers sont nécessaires pour des bactéries peu virulentes mais toujours pathogènes.

Dans certains cas, comme lors d'opérations en milieu hospitalier, le personnel doit s'assurer en permanence qu'aucune contamination du matériel qui va être utilisé n'a eu lieu afin d'éviter tout risque d'infection nosocomiale. Un rapport co-écrit par la Direction Générale de la Santé et par la Direction de l'Hospitalisation et de l'Organisation des Soins [41], montre qu'il reste énormément de travail en termes de compréhension quant aux infections nosocomiales impliquant des bactéries environnementales. Notre approche pourrait ainsi apporter une réponse intéressante à cette problématique. Voici quelques exemples de bactéries pathogènes pour l'Homme.

**Table 1 – Liste de bactéries présentant un caractère pathogène pour l'Homme [42]**

Maladie	Espèce Pathogène	Année de découverte
Charbon	<i>Bacillus anthracis</i>	1876
Fièvre typhoïde	<i>Salmonella Typhi</i>	1879
Tuberculose	<i>Mycobacterium tuberculosis</i>	1882
Choléra	<i>Vibrio cholerae</i>	1883
Diphthérie	<i>Corynebacterium diphtheriae</i>	1883
Tétanos	<i>Clostridium tetani</i>	1885
Diarrhée	<i>Escherichia coli</i>	1885
Pneumonie	<i>Streptococcus pneumoniae</i>	1886
Méningite	<i>Neisseria meningitidis</i>	1887
Peste	<i>Yersinia pestis</i>	1894
Syphilis	<i>Treponema pallidum</i>	1905
Coqueluche	<i>Bordetella pertussis</i>	1906

## **2.4. Souches bactériennes cultivées**

Au cours de ces travaux de thèse, différentes bactéries sont étudiées, et comme dit précédemment, aucune des souches analysées n'est pathogène. Durant toute la phase de développement et d'optimisation du protocole d'analyse des cellules par microspectroscopie Raman, des agents microbiologiques modèles ont été utilisés. Les cellules choisies ont été sélectionnées pour représenter au maximum la biodiversité naturelle au sein des bactéries tout en tenant compte des bactéries à disposition chez notre partenaire, le CSTB (Centre Scientifique et Technique du Bâtiment), lors de ce travail.

### **2.4.1. *Staphylococcus epidermidis***

Les staphylocoques sont des bactéries présentes naturellement dans l'environnement et peuvent être retrouvées chez l'Homme ou chez les animaux. Plus d'une vingtaine d'espèces sont dénombrées à ce jour et la plus pathogène de la famille est le *Staphylococcus aureus* (staphylocoque doré) bien connu pour être la cause de nombreuses infections. Cependant, la majorité des espèces ne sont pas pathogènes. Au sein de cette famille, on trouve également l'espèce *Staphylococcus epidermidis*, espèce très connue et probablement la plus commune des bactéries. Elles sont fréquemment retrouvées sur la peau et les muqueuses humaines et animales. L'espèce *Staphylococcus epidermidis* est généralement non pathogène mais peut, dans certains cas, causer des infections (sinusite, infection cutanée, etc.) chez des personnes dont le système immunitaire est déficient. Morphologiquement, le *Staphylococcus epidermidis* est de forme sphérique d'un diamètre moyen de 1 à 2  $\mu\text{m}$ . C'est une bactérie à Gram positif souvent identifiée comme modèle du genre *Staphylococcus*. C'est pourquoi lors du développement du protocole d'acquisition des spectres Raman, cette espèce a été sélectionnée.

### **2.4.2. *Pseudomonas fluorescens***

Le genre *Pseudomonas* présente une soixantaine d'espèces différentes et sont souvent retrouvées dans les eaux. Mais comme toutes les bactéries, de par leurs petites tailles, elles peuvent facilement être transportées dans l'air et ainsi se déposer sur les végétaux, l'épiderme et infiltrer les bronches. Au cours de ces travaux, l'espèce *Pseudomonas*

*fluorescens*, a été étudiée. L'espèce *Pseudomonas fluorescens* est souvent retrouvée dans les sols. Elle est non-pathogène pour l'Homme et même bénéfique pour les plantes. Malgré cela, il est possible de retrouver des contaminations par *Pseudomonas fluorescens* dans le domaine de l'agroalimentaire comme lors de l'affaire des mozzarellas bleues en 2010. À l'inverse du *Staphylococcus epidermidis*, le *Pseudomonas fluorescens* est une bactérie à Gram négatif se présentant sous la forme de bacille dit « en bâtonnet » et pouvant atteindre jusqu'à 5  $\mu\text{m}$  de long. L'espèce présente des caractéristiques intéressantes pour notre étude, c'est pourquoi elle a été désignée dans le panel des espèces à étudier.

#### **2.4.3. *Pseudomonas syringae***

L'espèce *Pseudomonas syringae*, comme sa cousine *Pseudomonas fluorescens*, est une bactérie à Gram négatif en forme de bâtonnet. Peu pathogène pour l'Homme, elle l'est beaucoup plus pour les plantes. De nombreux cas d'arbres malades ont été recensés entre 2005 et 2006 au Royaume-Unis, en Belgique et aux Pays-Bas, avec à chaque fois la présence de la bactérie *Pseudomonas syringae*. Cette bactérie est d'autant plus intéressante pour notre étude qu'elle fait partie du même genre que *Pseudomonas fluorescens* et présente des caractéristiques physiques et biologiques similaires. Il est donc intéressant d'introduire cette espèce dans le panel d'individus qui servira au développement de la méthode d'analyse par spectroscopie Raman et à l'optimisation du traitement chimiométrique des données acquises.

#### **2.4.4. *Escherichia coli***

*Escherichia coli* est certainement la bactérie la plus connue et la plus étudiée à ce jour. Cet individu se présentant sous la forme d'un bacille, pouvant mesurer de 1 à 3  $\mu\text{m}$ , à Gram négatif, tapisse la majeure partie de la flore intestinale de l'être humain. Elle est nécessaire pour la vie et donc loin d'être la bactérie la plus pathogène pour l'Homme. Néanmoins, certaines mutations ont conférée à certaines souches un caractère pathogène pouvant entraîner des infections inflammatoires du système digestif et des infections urinaires. Reconnue comme étant une bactérie modèle, cette espèce présente un bon potentiel pour la suite des travaux.

#### **2.4.5. Souches bactériennes isolées**

La majeure partie du développement et de l'optimisation du protocole d'acquisition des spectres Raman a été menée à l'aide des quatre souches vues précédemment. Cependant, lors des campagnes de prélèvements d'agents présents naturellement dans l'air, d'autres souches ont pu être isolées, mises en culture et déposées par sédimentation sur des lames de CaF<sub>2</sub>. Ces souches, alors inconnues, ont été analysées par identification moléculaire afin de déterminer leur identité. L'analyse a révélé la présence de 8 espèces différentes dans ces prélèvements d'aérosols naturels.

Parmi ces 8 souches isolées, nous retrouvons *Staphylococcus epidermidis*, une des souches initialement analysées lors de la mise en place du protocole d'acquisition des spectres. Nous retrouvons également une autre souche de staphylocoque, *Staphylococcus capitis*. Comme *Staphylococcus epidermidis*, cette bactérie est souvent retrouvée sur la peau humaine. D'autres bactéries d'origine végétale sont aussi isolées comme *Microbacterium testaceum*, *Arthrobacter agilis*, *Micrococcus yunnanensis* et d'autres plutôt originaires des sols comme *Kocuria palustris*, *Sphingomonas aerolata*, *Rhodococcus globerulus*. Toutes ces espèces seront analysées par microspectroscopie Raman et feront l'objet d'une étude chimio-métrique pour vérifier si la construction d'un modèle de classification basé sur des espèces environnementales est possible.

### **2.5. Conditions de culture**

Les espèces bactériennes de référence utilisées lors de la préparation des premiers échantillons sont toutes issues de la collection du Centre Scientifique et Technique du Bâtiment (CSTB). Les cellules sont cultivées sur milieu TSA (gélose Trypto-caséine soja) (Réf. Merck 1.05458) avant de procéder au dépôt sur lame de CaF<sub>2</sub>. Les températures d'incubation dépendent de la température optimale de croissance de chacune des espèces bactériennes. Pour les quatre souches de référence utilisées (Tableau 1), les températures et les durées d'incubation sont respectivement comprises entre 28°C et 37°C et varient de 18 h à 24 h.

Pour les espèces environnementales, les agents microbiologiques sont collectés sur gélose TSA ou gélose R2A (*Reasoner's 2A*) (Réf. Merck 1.00416). Les espèces sont incubées sur le milieu de collecte, puis isolées et transplantées deux fois sur un milieu de culture neuf, avant

la génération d'un aérosol pour dépôt sur lame de CaF<sub>2</sub>. Comme la nature des souches isolées est inconnue, les températures et durées d'incubation sont adaptées au cas par cas selon le développement de chacune des espèces.

**Tableau 1 – Liste des espèces bactériennes de référence pour l'optimisation des divers protocoles.**

Désignation	Référence	Origine de la collection
<i>Staphylococcus epidermidis</i>	CIP 81.55 T	France
<i>Escherichia coli</i> - NCIB 8134	ATCC®10799™	U.S.A.
<i>Pseudomonas fluorescens</i>	CIP 56.90	France
<i>Pseudomonas syringae</i>	DSMZ 21482	Allemagne

### 3. Préparation des échantillons

#### 3.1. Méthodes de prélèvement des aérosols atmosphériques

Plusieurs méthodes de prélèvement d'air sont disponibles pour l'étude de la faune bactérienne et des moisissures selon l'importance des concentrations de microorganismes attendues dans l'environnement de travail. Pour les environnements à concentrations élevées, c'est-à-dire à plus de 10 000 UFC/m<sup>3</sup> (unité formatrice de colonies), un prélèvement sur filtre de polycarbonate est recommandé. Pour les autres situations, la méthode standard, basée sur l'utilisation d'un impacteur en cascade de type Andersen, peut être employée. Dans certains cas, l'impaction en cascade ne permet pas d'obtenir des échantillons suffisamment dispersés comme lorsque l'échantillonnage est effectué directement sur des lames. C'est pourquoi, des instruments de collecte en milieu liquide sont également employés. Lors d'analyses de routine, les prélèvements réalisés lors de ces collectes ne sont pas analysés directement sur place mais doivent ensuite être acheminés dans un laboratoire d'analyses microbiologiques. Les microorganismes ainsi collectés nécessitent une période d'incubation

spécifique et des analyses d'identification moléculaire sont effectuées pour permettre une identification.

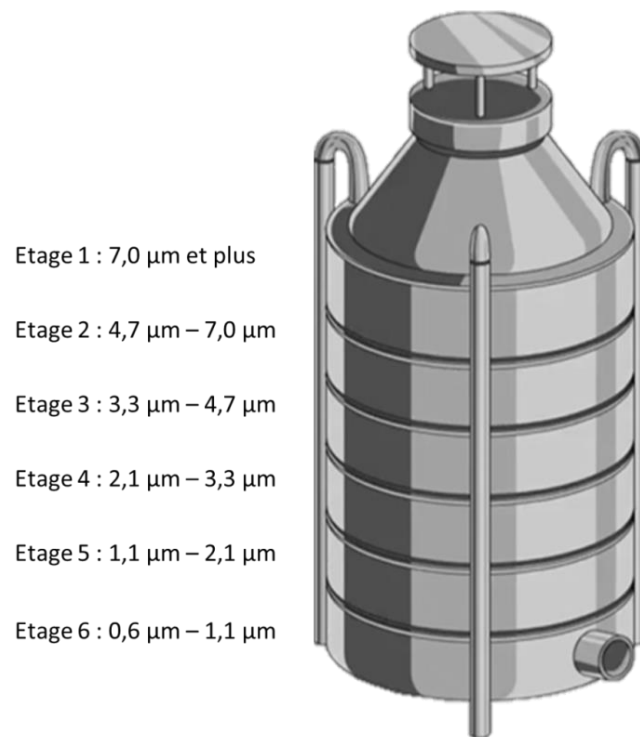
Le prélèvement des aérosols est donc une étape clé durant la mise en place du protocole d'identification des agents bactériologiques naturels. Il est important de verrouiller rapidement le mode de collecte des particules atmosphériques pour obtenir des échantillons exploitables en spectroscopie Raman, mais également adaptés pour la microdissection laser.

### ***3.1.1. Prélèvements des aérosols à l'aide d'un impacteur en cascade multi-étages de type Andersen***

Pour effectuer le prélèvement d'aérosols naturels, nous avons dans un premier temps utilisé un impacteur en cascade multi-étages de type Andersen [43]. Ce dispositif est généralement utilisé pour des prélèvements de courte durée et dans un environnement faiblement chargé en particules [44]. Dans ce dispositif, l'air est aspiré dans l'impacteur et une séparation des particules est opérée par inertie. Les particules ayant une inertie suffisante viennent s'impacter sur la surface de collecte tandis que les autres poursuivent leur descente dans les étages de l'impacteur. Au fur et à mesure de la progression du flux d'air dans les étages, la taille des particules impactées va décroître grâce à la séparation inertielle et à des filtres présents entre chacun des étages. De ce fait, les particules les plus grosses vont être récoltées dans les étages supérieurs de l'instrument et les plus petites dans les étages inférieurs. Un schéma (Figure 14) présente les intervalles de tailles des particules récoltées selon l'étage auquel la récolte est effectuée.

Généralement, le prélèvement des microorganismes est réalisé en utilisant un milieu collecteur capable de maintenir les agents bactériologiques en vie. L'impaction des microorganismes s'effectue alors sur un milieu constitué d'une gélose nutritive. Il est ainsi possible d'obtenir des échantillons d'aérosols naturels pour des analyses de comptage et d'identification moléculaire des agents bactériologiques. Contrairement aux pratiques usuelles, nous avons directement placé des lames de  $\text{CaF}_2$  dans l'impacteur pour générer des échantillons présentant des particules n'ayant subies aucun traitement spécifique ou phase de croissance entre l'étape de prélèvement et l'analyse spectroscopique. Les particules ainsi





**Figure 14 – Schéma d'un impacteur en cascade Andersen [45] et tailles des particules récoltées à chacun des six différents étages.**

collectées sont donc représentatives de la composition des milieux naturels intérieurs et extérieurs.

Les lames utilisées pour la récolte des échantillons ont été placées aux étages 5 et 6 de l'impacteur en cascade pour diminuer le risque d'un prélèvement de particules de grandes tailles. Les particules récoltées présentent donc théoriquement un diamètre compris entre 0,6  $\mu\text{m}$  et 2,1  $\mu\text{m}$ . Cependant, les échantillons générés à l'aide de l'impacteur en cascade multi-étages nous ont confronté à des difficultés nous obligeant à mettre en place un protocole de collecte différent de celui-ci.

### **3.1.2. Prélèvements des aérosols en milieu liquide à l'aide du SKC Biosampler**

Pour remédier aux difficultés rencontrées lors de la collecte des aérosols à l'aide de l'impacteur en cascade, nous avons opté pour un prélèvement en milieu liquide. Pour effectuer les prélèvements, nous utilisons l'échantillonneur SKC BioSampler (Figure 15).



**Figure 15 - Image de l'échantillonneur SKC BioSampler équipé d'un filtre particulaire (Source CSTB).**

Depuis sa conception, cet échantillonneur a été largement utilisé [46]–[48] et a démontré son potentiel pour la collecte d'aérosols. Le dispositif que nous utilisons est équipé d'un filtre particulaire en entrée avec un diamètre de maille ne laissant pénétrer dans le dispositif que les particules dont la taille est inférieure à 5  $\mu\text{m}$ . Le temps total de la collecte est d'environ 4 h, ce qui correspond à un volume d'air prélevé équivalent à 3  $\text{m}^3$ .

Les prélèvements en milieu liquide présentent l'avantage de limiter le stress rencontré par les particules à cause du choc mécanique induit par l'impaction sur milieu solide. Ce dispositif permet également de réaliser des prélèvements de plus longue durée et minimise les la ré-aérosolisation des particules [49]. Mais les deux avantages majeurs que présente ce mode de collecte pour notre étude sont, premièrement, la possibilité d'effectuer un dopage à l'aide d'une culture d'une souche connue et deuxièmement, la possibilité de procéder à un marquage par agent fluorescent au sein même du biocollecteur. Ainsi, toutes les cellules microbiologiques présentes, qu'elles soient naturelles ou issues d'une culture, peuvent être marquées.

### **3.2. Méthodes de dépôt**

Au même titre que les méthodes de prélèvements, les méthodes de dépôt sont multiples. Les principaux modes de dépôts utilisés pour la génération des échantillons au cours de cette thèse sont le dépôt par impaction (détaillé dans la partie précédente), le dépôt par goutte et le dépôt par aérosolisation/sédimentation.

#### **3.2.1. Dépôt par goutte**

Le mode de dépôt par goutte est le plus simple et le plus utilisé en microbiologie. Les cellules peuvent être cultivées puis transférées dans un milieu liquide, ou bien être directement collectées en milieu liquide à l'aide d'un échantillonneur d'aérosols comme le *SKC BioSampler*. Pour réaliser les échantillons, 20 à 50 microlitres d'une suspension bactérienne sont déposés sur une lame de  $\text{CaF}_2$ . La lame nouvellement contaminée est ensuite placée sous PSM (Poste de Sécurité Microbiologique) pour un séchage lent. Nous verrons dans la suite de la thèse, que ce mode de dépôt n'est pas le plus adapté à l'étude des bactéries par microspectroscopie Raman et pour la microdissection. C'est pourquoi un autre mode de dépôt est également exploité.

#### **3.2.2. Dépôt par aérosolisation/sédimentation**

Le dépôt par sédimentation est également une pratique courante en microbiologie. Cette méthode consiste à aérosoliser un échantillon qui initialement se trouve en phase solide ou liquide. Pour réaliser le dépôt par sédimentation, une colonie de bactéries est mise en suspension dans un eau distillée stérile et un aérosol est généré à l'aide d'un nébuliseur de type Collison. Les particules en suspension dans l'air vont lentement se déposer sur la surface de la lame de  $\text{CaF}_2$  par effet de sédimentation. L'ensemble de ces étapes sont effectuées sous PSM.

## **4. Méthodes d'identification moléculaire**

Lors de campagnes de collecte dans le cadre de la surveillance de la qualité de l'air, un des objectifs principaux est de déterminer la présence ou non d'un agent infectieux. Le but de

ces campagnes est la prévention et le déploiement de mesures nécessaires permettant la sécurité et la prise en charge de la population en cas de contamination bactériologique.

Durant les prélèvements de routine, les particules sont récoltées sur des géloses nutritives et sont ensuite placées en incubation selon le protocole défini à la page 41. Si des bactéries viables et cultivables sont présentes, elles vont subir une phase de croissance permettant de mettre en évidence les différentes colonies. L'étape suivante correspond à l'isolement des colonies. Chacune des colonies est ciblée et des cellules sont extraites pour procéder à un repiquage en culture pure. Lorsque ces nouvelles cultures arrivent à maturité, une identification de l'agent microbiologique est réalisée à partir des caractéristiques macroscopiques et microscopiques, du profil en acides gras ou encore grâce aux techniques d'identification moléculaire. L'identification moléculaire d'agents bactériens est une technique fiable et pratiquée dans la quasi-totalité des laboratoires d'analyses microbiologiques.

#### **4.1. Identification moléculaire de bactéries cultivées**

Au cours de ces travaux de thèse, une des campagnes de prélèvement d'agents bactériologiques dans des aérosols naturels nous a permis d'isoler des cellules viables. Ces microorganismes ont été placés en culture dans un incubateur, puis une portion des cellules a été utilisée pour générer un aérosol et effectuer un dépôt sur lame de CaF<sub>2</sub>. Une autre portion a été utilisée pour déterminer l'identité de ces espèces isolées par une analyse d'identification moléculaire.

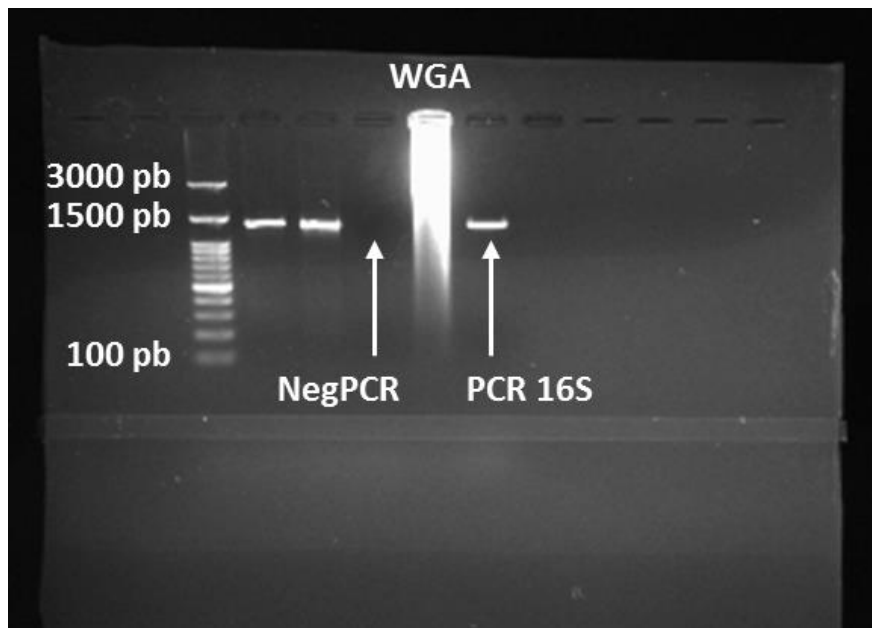
La première étape pour réaliser l'identification moléculaire d'une des colonies consiste à extraire l'ADN des bactéries à l'aide du kit QIAamp DNA Micro (Qiagen). Pour cela nous plaçons les bactéries dans un microtube de 1.5 mL contenant 10 µL de protéinase K, 100 µL de tampon ATL et 100 µL de tampon AL. La quantité utilisée de bactéries pour effectuer cette analyse est estimée entre 2 et 1000 cellules en suspension dans la solution. Elles sont ensuite soumises à une phase d'incubation de 10 minutes à une température de 56 °C, puis centrifugées brièvement au terme de l'incubation. 50 µL d'éthanol sont ensuite ajoutés, puis la solution est agitée au vortex en plusieurs fois pendant 15 secondes. Une nouvelle phase

d'incubation de 3 minutes à température ambiante est ensuite réalisée avant de centrifuger brièvement de nouveau. À cette solution, nous ajoutons 500 µL de tampon AW1, puis nous réalisons une centrifugation à 8000 rpm durant 1 minute. Après avoir changé de tube collecteur, ce sont 500 µL de tampon AW2 qui sont ensuite ajoutés, puis la solution est à nouveau centrifugée pendant 1 minute à 8000 rpm. À cette solution nous ajoutons entre 20 et 100 µL d'eau stérile de qualité biologie moléculaire avant d'effectuer une dernière phase de centrifugation de 1 minute à 13 000 rpm. Toutes ces étapes permettent ainsi d'extraire un maximum d'ADN. Au terme de ces différentes étapes, l'ADN extrait présent en quantité suffisante, subit une étape d'amplification génomique de l'ADNr 16S. L'ADNr 16S correspond aux gènes codant l'ARN ribosomique 16S et son étude permet d'identifier l'espèce du microorganisme isolé. Deux amorces sont utilisées pour réaliser l'amplification génomique par PCR (réaction en chaîne par polymérase) et les séquences de ces amorces sont renseignées dans le Tableau 2.

**Tableau 2 - Désignation, séquence et position des amorces utilisées pour l'identification des bactéries viables et cultivables provenant de prélèvements d'air [50].**

Nom	Séquence	Description	Position
<b>B26</b>	CAGGCCTAACACATGCA-AGTC	Bacteria ADNr 16S	63F
<b>B27</b>	GGGCGGWGTGTACAAGGC	Bacteria ADNr 16S	1387R

La PCR, qui est une suite de réactions enzymatiques permettant de dupliquer en grande quantité un fragment d'ADN spécifique, est effectuée jusqu'à obtenir suffisamment de fragments d'une longueur d'environ 1300 pb. Avant de procéder à l'étape de séquençage, il est nécessaire de s'assurer du bon déroulement de la PCR en réalisant une observation des produits sur gel d'électrophorèse (Figure 16).



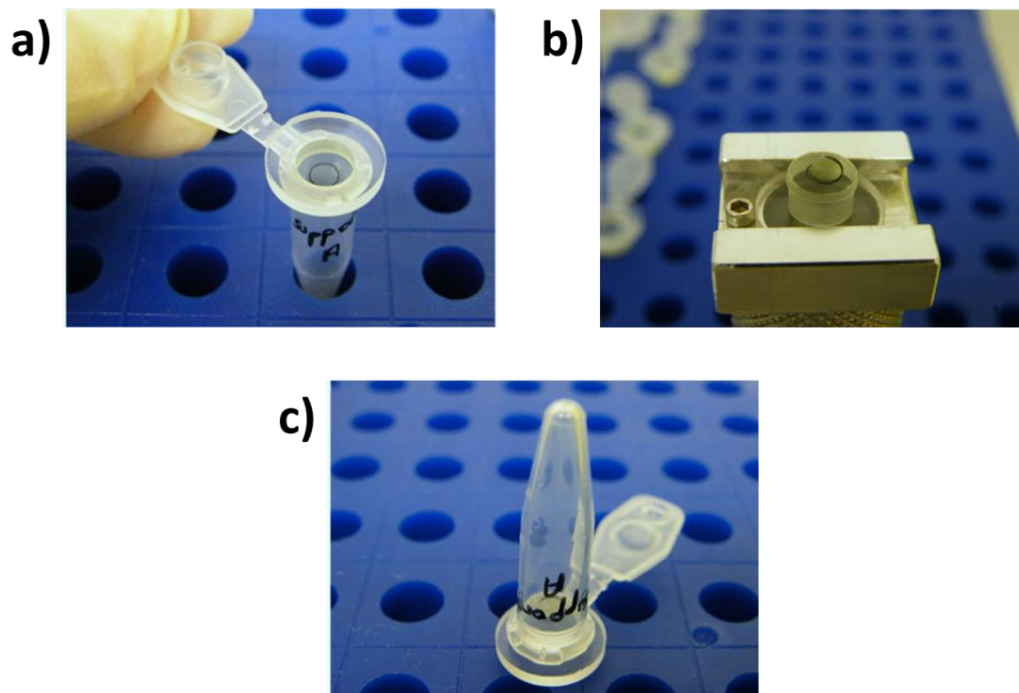
**Figure 16 – Détection sur gel d'électrophorèse des produits issus de l'amplification du génome entier (WGA) et du fragment à 1300pb issue de la PCR (PCR 16S).**

Un témoin négatif de PCR (NegPCR) est préparé systématiquement afin de s'assurer qu'aucune contamination n'a lieu lors de la manipulation du matériel génétique. La méthode de détection sur gel d'électrophorèse permet de conclure du bon déroulement de l'extraction du génome entier (WGA) et de l'amplification de l'ARN 16S (PCR 16S). Finalement, les produits PCR purifiés de la bactérie isolée sont soumis à une étape de séquençage SANGER sens et anti-sens, permettant d'obtenir les séquences des bactéries isolées. Ces séquences sont ensuite épurées et comparées avec les banques de données en ligne du *National Center for Biotechnology Information* (NCBI) [51]. Ce protocole a ainsi permis d'identifier les diverses souches isolées des prélèvements d'aérosols effectués durant cette campagne.

#### **4.2. Identification moléculaire des bactéries isolées par microdissection**

Un des objectifs du projet dans lequel s'inscrivent ces travaux de thèse est d'adapter le protocole d'identification moléculaire de bactéries cultivées, à une cellule unique prélevée par microdissection. Lorsque la cellule microdissectée parvient au CSTB (Figure 17), elle est piégée dans le film polymère de la « caps » de prélèvement et recouverte d'un tube Eppendorf

vierge de tout ADN afin d'éviter une contamination extérieure comme présenté sur la Figure 17 a).



**Figure 17 – Étapes d'extraction de l'ADN de la cellule isolée par microdissection (source CSTB).**

Il est donc nécessaire en amont de l'amplification génomique de réaliser l'extraction de la particule biologique présente sur le polymère. La particule microdissectée, située dans le cercle noir visible sur Figure 17 b), est recouverte d'un dépôt de 7  $\mu$ L de solution d'extraction provenant du kit REPLI-g® Single Cell (QIAGEN). La solution d'extraction contenant l'ADN de la bactérie est prélevée, et comme pour la méthode classique d'identification moléculaire, l'ADN est solubilisé dans un mélange de tampon et de polymérase provenant du kit REPLI-g® Single Cell (QIAGEN) pour être amplifié. Du fait que la quantité d'ADN est trop faible, une étape d'amplification du génome entier est effectuée, puis une amplification de l'ADNr 16S par PCR est réalisée à l'aide du Kit Pfu Turbo (Stratagene). Les séquences des amorces utilisées pour l'amplification sont précisées dans le Tableau 3.

**Tableau 3 - Désignation, séquence et position des amorces utilisées pour l'identification des bactéries prélevées par microdissection [52].**

Nom	Séquence	Description	Position
<b>B01</b>	TCCTACGGGAGGCAGCAGT	Bacteria ADNr 16S	331-349F
<b>B02</b>	GGACTACCAGGGTATCTAATCCTGTT	Bacteria ADNr 16S	772-797R

Comme pour le protocole d'identification moléculaire sur les bactéries issues d'une culture, une vérification du bon déroulement de la PCR est effectuée sur gel d'électrophorèse puis le séquençage SANGER est réalisé sur les produits PCR. Enfin, les séquences obtenues sont comparées aux bases de données en ligne de NCBI.

Il est important de noter que compte tenu de la faible quantité de matière et donc de matériel génétique présent initialement sur les « caps », l'identification génomique de cellules uniques prélevées par microdissection laser est une tâche difficile qui ne présente pas toujours des résultats positifs.

## 5. Conclusion

La biosurveillance de l'air présente des enjeux impactant les domaines civil et militaire. Il est donc nécessaire de déployer des moyens efficaces pour prévenir des risques d'une épidémie à cause de la libre circulation dans l'air d'agents bactériens. Pour cela nous proposons un nouveau protocole permettant d'isoler, de repérer, d'analyser et d'identifier des souches bactériennes dans les aérosols. Les cellules bactériennes sur lesquelles nous avons travaillé sont soit issues de souches de laboratoire soit isolées à partir de prélèvements d'aérosols. Nous avons détaillé les différents modes de collecte et de dépôt des cellules et nous privilégierons une collecte en milieu liquide à l'aide du *SKC BioSampler* et un dépôt par sédimentation. Dans une dernière partie, nous avons expliqué comment sont réalisées les analyses d'identification moléculaire aussi bien effectuées sur les bactéries de cultures que sur les souches isolées d'un aérosol naturel.



# Chapitre 3 – Optimisation de la mesure Raman, des modèles de classification et confrontation aux aérosols naturels

Dans ce chapitre, il sera question de développer, d'optimiser et de standardiser toute la première moitié du protocole d'analyse des bactéries dans l'air (Figure 18) à savoir la collecte des aérosols, le repérage des particules sur les lames de prélèvement, leur analyse Raman et la construction d'un modèle de classification SVM.

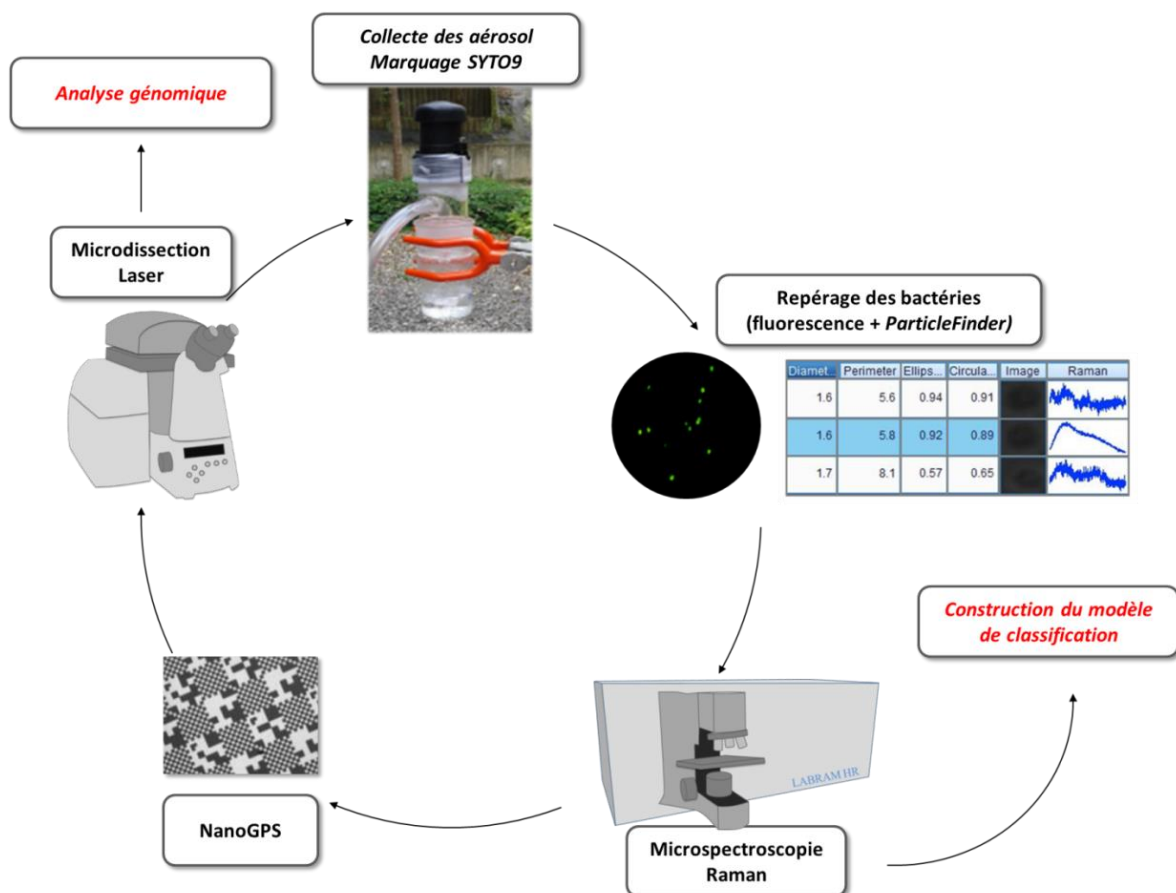


Figure 18 - Schéma du protocole pour l'analyse des bactéries dans l'air.

Dans une première partie, nous aborderons la thématique d'optimisation du protocole d'acquisition des spectres Raman sur les bactéries. Il est essentiel de déterminer sur quel support les bactéries vont être déposées et la longueur d'onde d'excitation à laquelle nous devons travailler. Des paramètres liés au spectromètre lui-même comme le choix du réseau de diffraction doivent également être étudiés. Au-delà de l'analyse Raman conventionnel, nous aborderons l'intérêt et le potentiel qu'offre un autre type de spectroscopie Raman, le CARS, pour diminuer les temps d'analyse.

Ensuite, nous évaluerons le potentiel et la puissance des modèles de classification SVM pour identifier les bactéries. Il faudra également déterminer le ou les prétraitements mathématiques à appliquer aux spectres afin d'en extraire un maximum d'informations pour améliorer les modèles de classification.

Face à la complexité des aérosols naturels, il est nécessaire de modifier et d'améliorer le protocole de collecte et d'analyse Raman des bactéries. Pour cela, nous verrons comment le protocole de collecte a été adapté pour réduire la densité particulaire des échantillons. Il sera également détaillé comment, à l'aide de sondes fluorescentes et d'un logiciel d'analyse d'images, nous parvenons à améliorer le protocole en vue de repérer les bactéries au sein du dépôt d'un aérosol.

Enfin, dans une dernière partie, nous examinerons les résultats d'une étude portée sur des bactéries cultivées en laboratoire mais issues de souches environnementales. Ces souches ont été extraites et isolées d'un prélèvement d'air, remises en culture et déposées sur la lame de  $\text{CaF}_2$  par sédimentation.

## **1. Acquisition des spectres Raman**

Au cours de ces travaux, nous développons une méthode d'analyse de bactéries par microspectroscopie Raman. La mise en place de ce protocole nécessite de perfectionner et d'optimiser l'ensemble des paramètres influençant la mesure pour obtenir des spectres Raman d'une qualité suffisante. Dans cette partie, nous déterminerons d'abord quel support utiliser pour le dépôt des bactéries, nous détaillerons ensuite l'optimisation des différents

paramètres pour l'acquisition spectrale et nous verrons enfin comment la nature chimique des bactéries permet une discrimination des espèces.

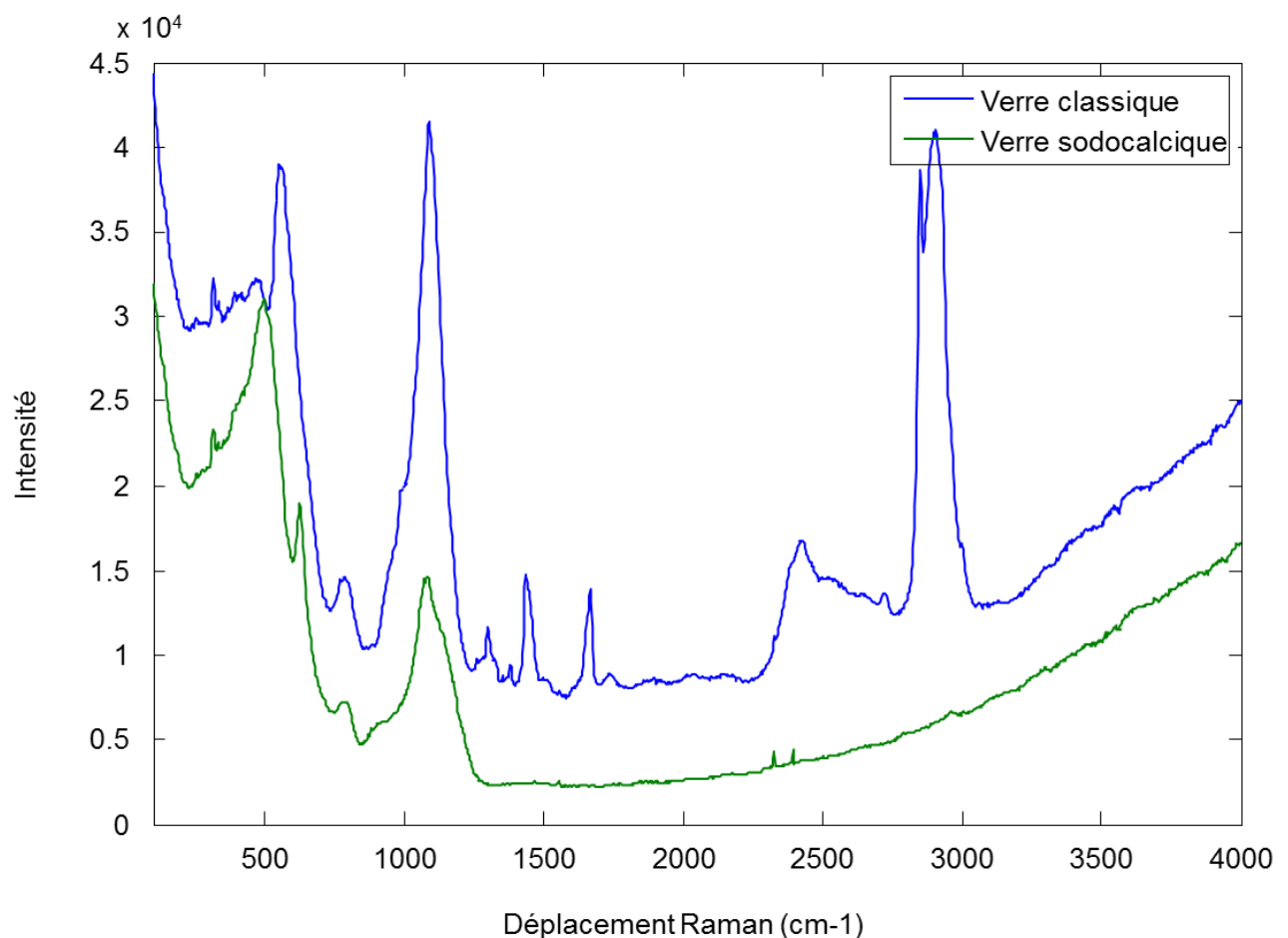
### **1.1. Détermination du support**

Dans cette partie, les bactéries sont cultivées puis déposées par sédimentation sur un support pour effectuer l'analyse par microspectroscopie Raman. La nature chimique du support peut se révéler problématique pour la suite des analyses. Par exemple l'utilisation d'un support plastique ou polymère est impossible du fait de la présence de liaisons carbone-hydrogène, ce qui entrainerait des recouvrements de bandes entre le signal lié à la bactérie et le signal lié au support. Le choix du support doit donc être fait judicieusement en prenant en compte l'entièreté du protocole mis en place pendant ces travaux et pas seulement l'aspect spectroscopique du problème. Nous entendons ici que le support doit être adapté pour la spectroscopie Raman mais également pour le transfert des échantillons sur le microdissecteur. Il a donc fallu étudier plusieurs types de supports afin de déterminer lequel sera optimal pour l'ensemble du protocole.

Pour déterminer quel est le support le plus approprié à notre étude, plusieurs critères devront être respectés. Le premier de ces critères est directement lié au caractère chimique du support. Celui-ci devra être transparent, c'est-à-dire sans contributions spectrales, idéalement sur l'intégralité du spectre ou *a minima* dans la région empreinte du spectre (*fingerprint*) de la bactérie afin d'éviter tout recouvrement de bandes qui pourrait se révéler critique lors de la séparation des espèces par les méthodes chimiométriques. Le support devra également être transparent dans le domaine du visible. Certaines étapes du protocole global nécessitent une observation de l'échantillon par transmission. En effet, lors de l'étape de microdissection laser des particules, la lame est placée sur la platine de façon à ce que les particules soient sur la face supérieure. La tourelle d'objectifs se trouve alors sous l'échantillon (observation par-dessous) et le prélèvement à l'aide des caps se fait depuis la face supérieure. Avec une lame trop opaque, l'observation des particules et la microdissection seraient impossibles. Dans la littérature, on retrouve plusieurs types de supports pour l'analyse des bactéries par microspectroscopie Raman. En SERS, des supports métalliques, comme des nano-tiges d'argent sont utilisés pour amplifier le signal [53]. Un tel type de support n'est pas

envisageable pour notre étude de par son opacité. Certains travaux rapportent également l'utilisation de lames en quartz [54] ou en  $\text{CaF}_2$  [55].

Différents supports vont être analysés et comparés afin de déterminer lequel présente les caractéristiques optimales pour une analyse des bactéries par spectroscopie Raman. Dans cette étude, quatre supports ont été sélectionnés : deux types de verre de pureté différente et deux types de lames en fluorure de calcium ( $\text{CaF}_2$ ) également de pureté différente. Pour le choix des verres, nous avons utilisé une lame de verre dite « classique » couramment exploitée en biologie et une lame de verre de qualité supérieure, un verre sodocalcique. Ce type de verre est étudié car il compose les lames sur lesquelles sont imprimés les motifs NanoGPS permettant le couplage des différents instruments utilisés (Cf.



**Figure 19 - Spectres Raman de deux lames de verre. En bleu : une lame de verre "classique" communément exploitée en biologie et en vert : une lame de verre sodocalcique de plus grande pureté.**

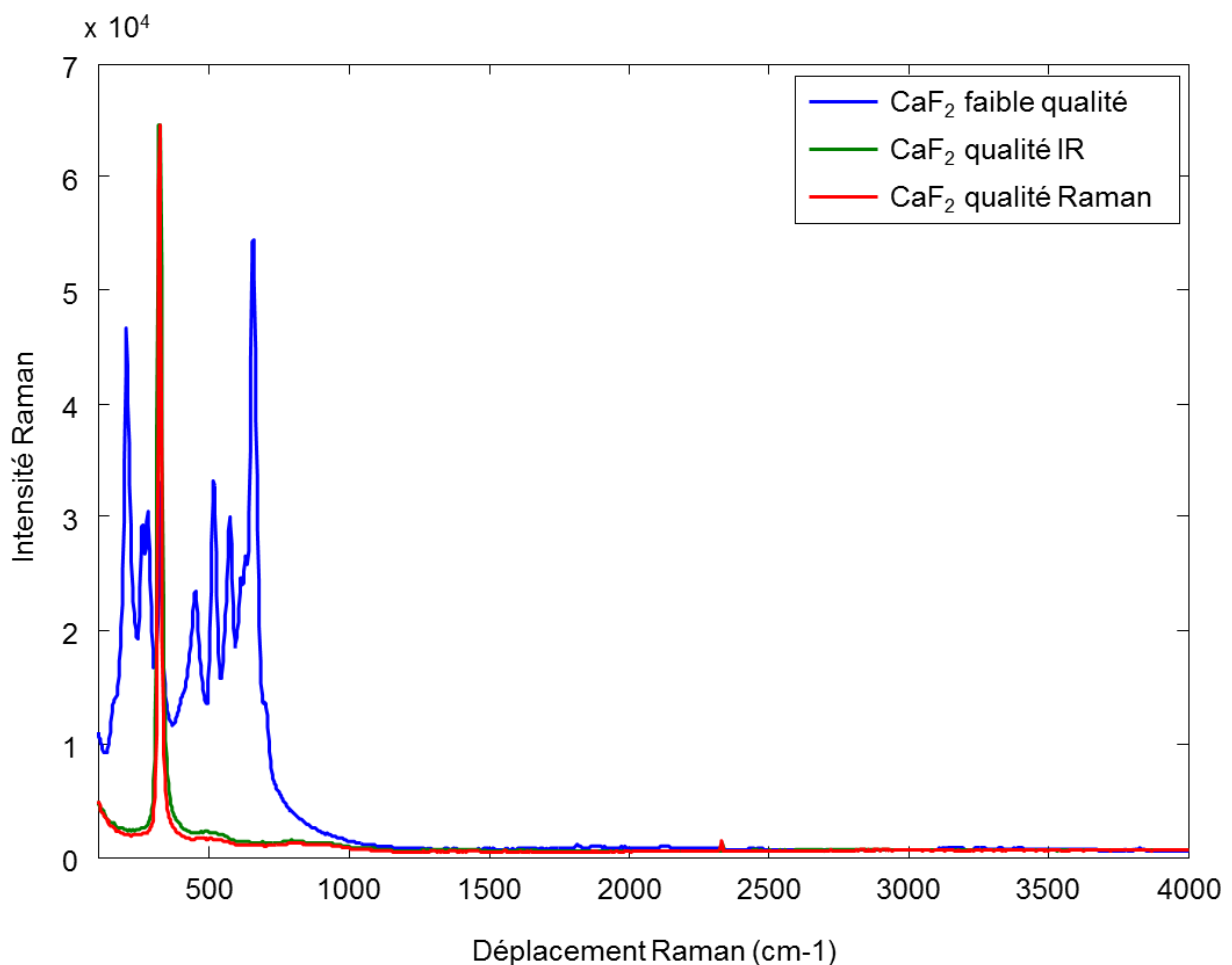
Chapitre 4 1. Système de repositionnement absolu : le nanoGPS). L'intérêt est donc de savoir s'il nous est possible de travailler directement sur ces lames ou bien s'il faut trouver un autre support de substitution. Des spectres ont été acquis sur ces deux types de lames et sont présentés dans la Figure 19.

Les spectres bleu et vert correspondent respectivement à un verre dit « classique » et à un verre sodocalcique. Le spectre du verre classique présente un grand nombre de bandes sur tout le domaine spectral et un fond de fluorescence non négligeable ce qui le rend inadapté pour la suite des analyses. Concernant le spectre du verre sodocalcique, aucune contribution Raman importante n'est observée dans la zone spectrale allant de  $1400\text{ cm}^{-1}$  à  $4000\text{ cm}^{-1}$ , en revanche un fond de fluorescence est présent. Malheureusement, dans la zone spectrale d'intérêt pour l'analyse Raman des particules bactériologiques, la *fingerprint*, qui s'étend de  $500\text{ cm}^{-1}$  à  $1800\text{ cm}^{-1}$ , le verre sodocalcique présente des contributions Raman importantes rendant également ce support inexploitable pour le dépôt des particules pour analyse par microspectroscopie Raman. Ces deux types de verres ne sont donc pas appropriés pour l'analyse Raman des bactéries.

L'utilisation de lames de  $\text{CaF}_2$  est une pratique courante [55]–[57] lorsqu'il s'agit d'analyser des particules biologiques comme les bactéries ou les moisissures par spectroscopie Raman. Mais comme pour le verre, il existe plusieurs niveaux de pureté dans les lames de  $\text{CaF}_2$ , selon le fournisseur voire chez un même fournisseur. Pour cette étude, trois lames de  $\text{CaF}_2$  de qualité différente, ont été étudiées. Les spectres présentés dans la Figure 20 ont été acquis dans les mêmes conditions expérimentales que pour les lames de verre.

Le spectre bleu correspond à une lame de faible qualité, c'est-à-dire une lame qui présente un nombre d'imperfections cristallines et d'impuretés plus important. Il présente ainsi un nombre conséquent de bandes de diffusion dans la région spectrale inférieure à  $1000\text{ cm}^{-1}$ . Ce cristal n'a pas été conçu pour un usage en spectroscopie Raman. Les deux autres spectres, vert et rouge, sont semblables. Ils présentent tous deux une bande Raman intense centrée à  $322\text{ cm}^{-1}$  et une bande plus faible centrée à  $2330\text{ cm}^{-1}$ , caractéristiques du  $\text{CaF}_2$ . Toutefois, il existe une différence, peu visible à l'œil nu, dans l'intensité de la ligne de base des deux spectres. Suite à ces comparaisons, il est évident que les lames de verre, quelle qu'elles soient, ne seront pas utilisées pour la spectroscopie Raman. La lame de  $\text{CaF}_2$  de faible qualité

ne sera également pas retenue à cause du nombre trop important de contributions spectrales qu'elle présente dans la *fingerprint*. Le choix d'une lame en  $\text{CaF}_2$  de qualité supérieure sera préférable du fait que ces dernières sont silencieuses sur l'ensemble du domaine spectral qui sera conservé, à savoir la région allant de  $500 \text{ cm}^{-1}$  à  $3500 \text{ cm}^{-1}$ . À compter de ce moment, tous les dépôts de particules et d'agents microbiologiques se feront sur les lames de  $\text{CaF}_2$  de qualité spectroscopique.



**Figure 20 - Spectres Raman de trois lames de  $\text{CaF}_2$ . En bleu : une lame de  $\text{CaF}_2$  de faible qualité, en vert : une lame de  $\text{CaF}_2$  de qualité IR et en rouge : une lame de  $\text{CaF}_2$  de qualité Raman.**

## **1.2. Optimisation de la mesure Raman**

L'objectif des travaux détaillés dans cette partie est d'obtenir des spectres de cellules uniques d'une qualité suffisamment bonne sans endommager les cellules. Pour cela, il faut trouver un compromis entre le laser utilisé, le temps d'exposition, les filtres de densité à exploiter et le réseau de diffraction. C'est donc pas-à-pas que nous allons procéder à la détermination de tous les paramètres d'acquisition pour construire le protocole d'analyse par microspectroscopie Raman. Tous les travaux présentés dans cette partie sont réalisés à l'aide du spectromètre HR 800 équipé d'un objectif x100 et une ouverture numérique de 0.90.

### **1.2.1. Détermination de la longueur d'onde du laser pour l'acquisition**

Au cours de ces travaux, différents lasers ont été utilisés pour faire l'acquisition des spectres Raman. Nous avons utilisé dans cette thèse trois lasers allant du bleu au rouge en passant par le vert (473nm, 532nm, 633nm). Un laser UV a aussi été étudié en espérant exalter le signal de diffusion, mais les photons étant trop énergétiques ils carbonisent trop rapidement les bactéries. Dans la littérature, l'utilisation de laser à 473nm (bleu) est rencontrée mais peu courante. En revanche, l'utilisation de laser émettant à 532 nm (vert) et à 633 nm (rouge) est plus commune. Il est évident que le choix du laser aura un impact sur l'intégrité de la cellule bactériologique après l'analyse. L'énergie délivrée à la bactérie est inversement proportionnelle à la longueur d'onde d'excitation. Plus la longueur d'onde est faible, plus le risque d'endommager la cellule est important. L'analyse se doit d'être non destructive et ne doit engendrer aucune modification afin de réaliser des prélèvements par microdissection et finalement une analyse génomique dans de bonnes conditions. Comme expliqué plus tôt, les bactéries ne sont pas toutes identiques et réagissent donc différemment lors de l'exposition au laser pour l'acquisition du spectre Raman.

L'acquisition des spectres à l'aide du laser émettant à 473 nm s'est rapidement révélée problématique à cause du nombre trop important de bactéries détruites, même pour des faibles puissances. Les bactéries étant déposées sur des lames de  $\text{CaF}_2$ , il est difficile de dissiper l'énergie impactée sur une cellule unique pour que celle-ci ne brûle pas. Lorsque les bactéries sont dans le milieu liquide, celui permet effectivement une meilleure dissipation de l'énergie. Des filtres de densité ont été utilisés dans le but de diminuer la puissance incidente.

Sans filtre de densité, le laser délivre une puissance moyenne d'environ 30mW à l'échantillon. L'utilisation des filtres à densité permet de diminuer la puissance à l'échantillon jusqu'à 1 mW. À ces niveaux de puissance, des spectres ont pu être acquis mais seulement sur un nombre restreint de cellules ; les autres ayant brûlées. En utilisant des filtres plus denses, pour atténuer d'avantage la puissance, l'acquisition de spectres présentant une qualité suffisante devient alors impossible pour une discrimination à l'aide des outils chimiométriques. Même si des spectres ont pu être acquis, il faut garder à l'esprit que le protocole doit être adapté à des micro-organismes provenant d'un échantillon d'air. Nous ne pouvons donc pas nous permettre de brûler un trop grand nombre de cellules. Le laser à 473 nm ne sera donc pas utilisé pour l'analyse Raman des bactéries. En revanche, il sera d'une grande utilité lors de l'étape de repérage des bactéries dans le Chapitre 4, détaillée plus tard.

Contrairement au précédent laser, les lasers émettant à 532 nm et à 633 nm, sont des sources plus couramment utilisées dans la littérature. Elles délivrent moins d'énergie à la particule analysée, diminuant le risque de dégradation des bactéries. Le premier critère que nous avons suivi est donc le nombre de bactéries détruites. Il est apparu, lors des analyses, que les bactéries brûlent moins vite. Avec le laser à 473 nm à 1 mW de puissance, les cellules brûlent en quelques dizaines de secondes, alors qu'avec le laser à 532 nm à 18 mW ou 9 mW de puissance, le nombre de cellules ayant brûlées est beaucoup moins important, pour un temps d'intégration de l'ordre de la minute. Lorsque le laser à 633 nm à 10 mW de puissance est utilisé, plus aucune cellule ne brûle. La Figure 21 schématise la résistance des bactéries lorsqu'elles sont soumises à une excitation laser.

Pour cette étude, des cellules de *Staphylococcus epidermidis* sont utilisées et exposées à un rayonnement laser à différentes longueurs d'onde et à différentes puissances. Les spectres sont enregistrés toutes les 30 s pendant 10 min. Lorsqu'une bactérie brûle, cela se traduit par une saturation des spectres, puis par une décroissance lente du signal en présentant deux bandes caractéristiques du carbone aux alentours de  $1350\text{ cm}^{-1}$  et de  $1600\text{ cm}^{-1}$ . Lorsqu'une bactérie est exposée au laser 532 nm avec une puissance de 18 mW (Figure 21 a)), elle brûle en moyenne après 4 min à 5 min d'exposition. On observe parfaitement la



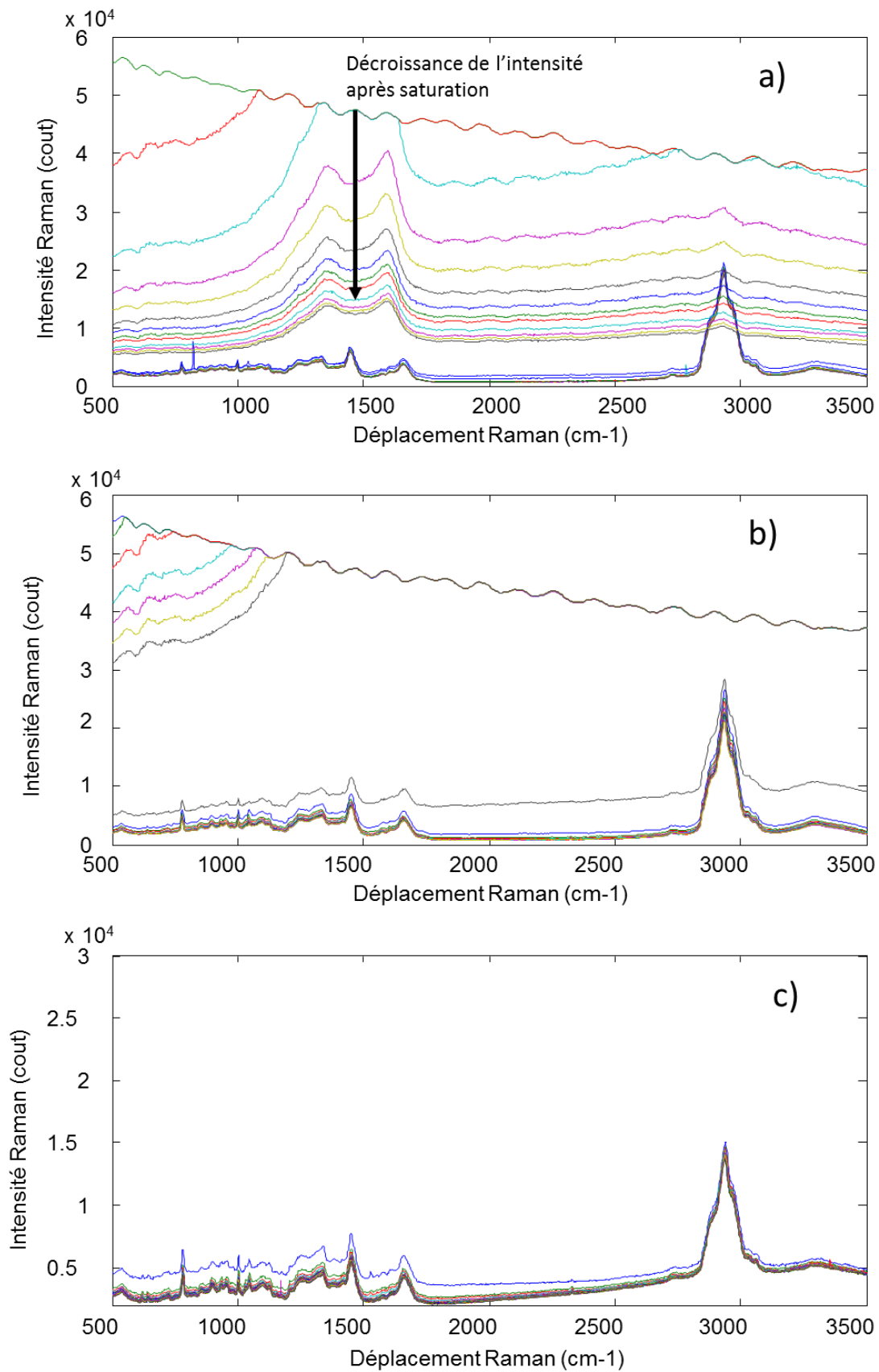


Figure 21 - Évaluation de la résistance des bactéries au laser a) 532 nm (puissance : 18 mW), b) 532 nm (puissance : 9 mW), c) 633 nm (puissance : 10 mW)

saturation du détecteur et la décroissance du spectre du carbone sur les acquisitions faites au-delà de 5 min d'exposition. Lorsqu'une bactérie est exposée au même laser mais avec une puissance deux fois plus faible c'est-à-dire de 9 mW (Figure 21 b)) elle brule en moyenne après 7 min à 8 min d'exposition. Comme précédemment on observe la saturation du spectre et on commence à observer la décroissance du signal. Lorsqu'une bactérie est exposée cette fois ci au laser 633 nm avec une puissance de 10 mW (Figure 21 c)) elle ne brule pas, même après 10 min d'exposition.

Cette étude démontre qu'à puissance égale, les bactéries supportent mieux l'exposition à la radiation lorsque les longueurs d'onde sont élevées. C'est pourquoi, dans la suite des travaux nous privilégierons l'utilisation du laser rouge à 633 nm pour effectuer les acquisitions des spectres Raman des bactéries. On remarquera qu'à cette longueur d'onde, le fond de fluorescence n'est pas très important.

### **1.2.2. Choix du réseau de diffraction**

Le réseau de diffraction est l'élément central du filtrage en spectroscopie Raman. L'utilisation de ce réseau est nécessaire pour décomposer le signal diffusé car la fente d'entrée dans le spectromètre est fixe. Un réseau présentant un nombre important de traits permettra d'obtenir des spectres bien résolus, mais au détriment du rapport signal sur bruit. À l'inverse, un réseau avec un faible nombre de fentes permettra l'acquisition d'un spectre avec un meilleur rapport signal sur bruit, au détriment de la résolution spectrale. Ainsi pour obtenir des spectres d'intensité comparable, le temps d'acquisition sera d'autant plus long que le réseau comporte un nombre important de fentes. Il faudra alors compter deux fois plus de temps pour réaliser une acquisition avec un réseau 600 traits/mm qu'avec un réseau 300 traits/mm pour avoir un rapport signal sur bruit équivalent.

Le spectromètre Raman HR 800 peut être équipé de trois réseaux de diffraction différents à savoir un réseau 300, 600 ou 1800 traits/mm. Pour déterminer quel réseau est le plus adapté à notre protocole, nous avons décidé de privilégier un temps d'acquisition faible au détriment de la résolution spectrale. L'analyse Raman devra être relativement courte afin d'éviter tout risque de modification physique ou chimique de la bactérie entraîné par une

exposition prolongée sous le laser. Ainsi, l'utilisation des réseaux 1800 traits/mm et 600 traits/mm ont rapidement été écartés.

En choisissant un réseau 300 traits, la résolution spectrale est diminuée, mais en contrepartie nous avons un gain en efficacité de diffraction et donc en rapport signal sur bruit ainsi qu'en temps d'analyse globale. Toutefois, à l'aide du réseau 300 traits / mm, la résolution spectrale est de  $2 \text{ cm}^{-1}$ , ce qui est inférieur à la largeur des contributions présentes dans le spectre Raman des bactéries, et donc suffisante pour notre étude.

### **1.2.3. Détermination du temps d'acquisition**

Le temps d'acquisition est un paramètre majeur pour la qualité des spectres et donc pour les performances des modèles de classification. En réalisant une acquisition durant un temps élevé, nous augmentons le rapport signal sur bruit et ainsi la chance d'observer de faibles contributions est augmentée. Cependant, il faut modérer ce temps d'exposition de la cellule sous le laser car le risque de brûler la bactérie est d'autant plus important que le temps d'exposition est long. Un des risques que nous encourons également en utilisant un temps d'acquisition trop élevé est la saturation des fortes contributions. Nous devons donc trouver un compromis. Le spectre des bactéries peut être décomposé en régions distinctes : la région *fingerprint* allant de  $500 \text{ cm}^{-1}$  à  $1800 \text{ cm}^{-1}$  et la région caractéristique des liaisons -C-H allant de  $2600 \text{ cm}^{-1}$  à  $3400 \text{ cm}^{-1}$ . Cette dernière présente généralement une intensité au moins deux fois supérieure à la *fingerprint* et pourrait saturer en cas d'exposition prolongée. Elle correspond principalement aux contributions des groupes C-H des molécules.

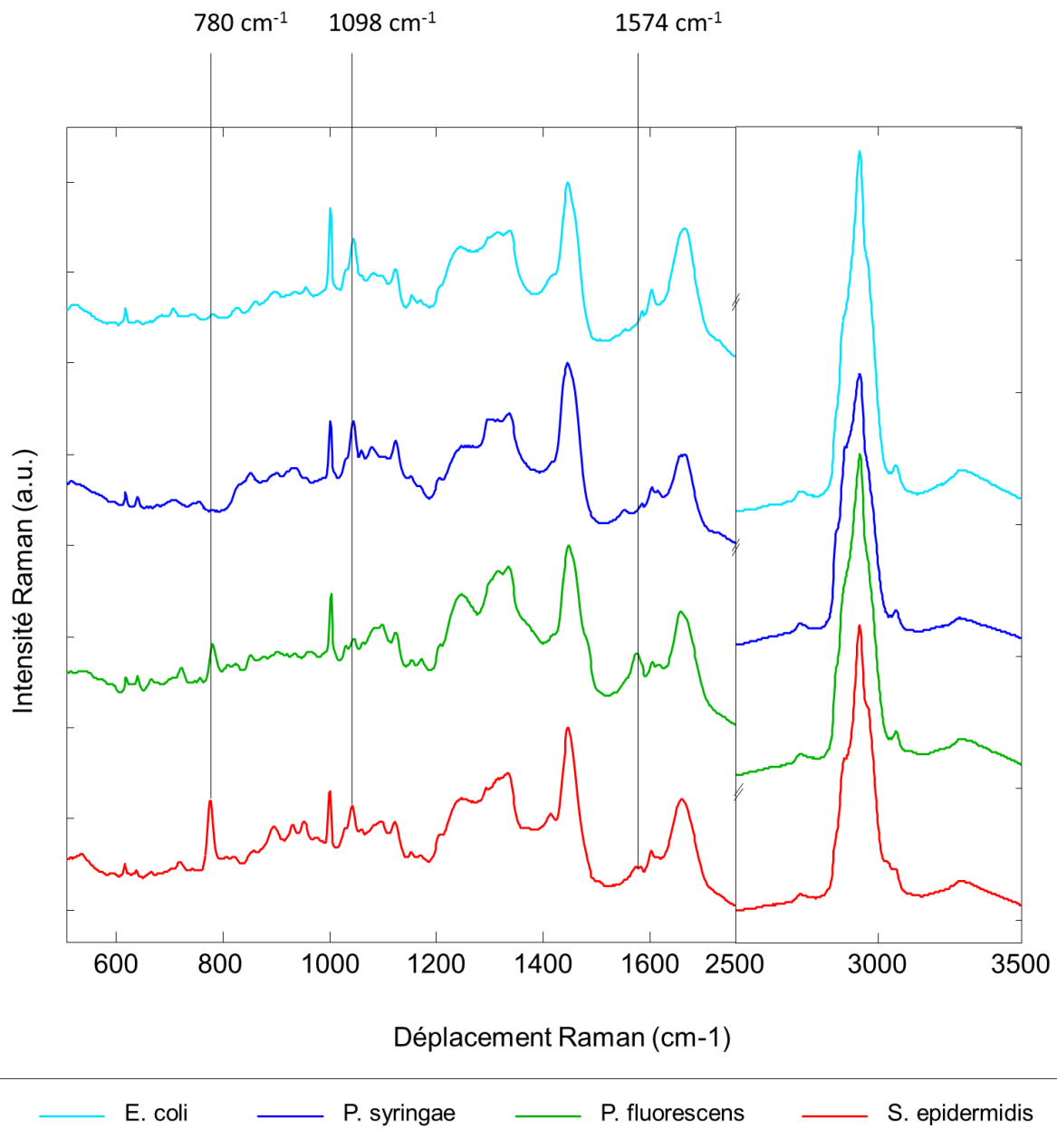
Il a fallu fixer très rapidement et dès le début de la thèse un temps d'intégration alors que nous n'avions à disposition qu'une base exhaustive de bactéries. Il n'était donc pas possible de trouver le temps d'intégration optimal permettant la construction du meilleur modèle de classification, ni de déterminer le temps d'acquisition minimum permettant une discrimination. C'est donc à l'aide d'une approche essai-erreur que nous avons testé différents temps d'acquisition afin de faire varier l'intensité du signal et le rapport signal sur bruit. De façon empirique, un temps d'exposition de 60s permet d'enregistrer des spectres présentant un bon rapport signal sur bruit pour les quatre espèces analysées. De plus, les modèles de classifications construits pour des temps d'intégration de 60s présentent d'excellents taux de

classification. Ce temps est similaire aux données dans la littérature [58]–[60]. Il est à noter que même si le réseau de diffraction choisi possède un nombre de traits relativement faibles, il ne permet pas d’acquérir un spectre sur tout le domaine en une seule acquisition. Deux fenêtres spectrales sont donc nécessaires pour générer un spectre, ce qui impose donc un temps d’intégration de 2 x 60 s par bactérie.

### **1.3. Nature des contributions dans le spectre Raman des bactéries**

Comme nous avons pu le voir précédemment, que ce soit morphologiquement, biologiquement ou au niveau de leur comportement, les bactéries sont toutes plus ou moins différentes les unes des autres. Ce sont ces différences qui permettent aux microbiologistes de classer les bactéries dans différents groupes. Via les méthodes spectroscopiques, il va être possible d’obtenir des informations sur la composition chimique de la bactérie. Selon l’espèce que nous rencontrons, les protéines et leur taux peuvent différer laissant alors la possibilité d’observer des différences spectrales. Pour illustrer ces propos sont représentés dans la Figure 22 les spectres moyens des quatre espèces étudiées pour la construction d’un modèle de classification, à savoir *S. epidermidis*, *P. fluorescens*, *P. syringae* et *E. coli*. Ces espèces présentent des spectres moyens très semblables or en regardant de plus près, il est possible de déceler de légères différences dans les contributions spectrales. La similarité des spectres s’explique par le fait que nous étudions des objets biologiques très proches. On comprend alors le challenge de cette thèse qui sera de mettre en lumière ces petites différences spectrales pour la classification finale.

Sur la Figure 22, sont pointées trois zones du spectre (780, 1098 et 1574  $\text{cm}^{-1}$ ) où des différences sont observées d’un spectre à un autre. Par exemple, on observe clairement la présence d’une contribution à 780  $\text{cm}^{-1}$  pour *S. epidermidis* et *P. fluorescens* contrairement aux deux autres spectres où la contribution est absente. Sont également présentes des variations d’intensité dans les bandes, plus difficilement repérables mais tout aussi importantes lorsqu’il s’agira de construire un modèle de classification.



**Figure 22 - Spectres moyens des quatre bactéries utilisées pour construire et optimiser le modèle de classification. Zoom sur la fingerprint.**

L'ensemble des bandes présentes dans la *fingerprint* des spectres Raman des quatre espèces de bactéries représente le contenu de la cellule à savoir des protéines, des lipides et des acides nucléiques. Les bactéries possèdent également une contribution très intense aux alentours de 2930 cm<sup>-1</sup> qui est la bande caractéristique des elongations C-H. Il est également possible d'observer un épaulement au-delà de 3100 cm<sup>-1</sup> qui est dû à la présence d'eau. L'origine des principales bandes est listée dans le Tableau 4.

**Tableau 4 – Liste des contributions Raman observées dans les spectres des bactéries à 633 nm selon la littérature [58], [59].**

Déplacement Raman	Source de la contribution	Intensité
540 cm <sup>-1</sup>	Déformation COC glycosidique	faible
620 cm <sup>-1</sup>	Phénylalanine	faible
642 cm <sup>-1</sup>	Tyrosine	faible
668 cm <sup>-1</sup>	Guanine	faible
724 cm <sup>-1</sup>	Adénine	faible
780 cm <sup>-1</sup>	Phénylalanine, Cytosine, Uracile	forte
830 cm <sup>-1</sup>	Tyrosine	faible
896 cm <sup>-1</sup>	Elongation COC	moyenne
1002 cm <sup>-1</sup>	Phénylalanine	forte
1098 cm <sup>-1</sup>	Phosphate	faible
1130 cm <sup>-1</sup>	=C-C= (Acides Gras insaturés)	faible
1156 cm <sup>-1</sup>	Elongation C-C	faible
1172 cm <sup>-1</sup>	Tyrosine, Phénylalanine	faible
1208 cm <sup>-1</sup>	Tyrosine, Phénylalanine, Protéine, Amide III	faible
1254 cm <sup>-1</sup>	Adénine, Amide III	moyenne
1336 cm <sup>-1</sup>	Adénine, Guanine, Tyrosine, Tryptophane	moyenne
1450 cm <sup>-1</sup>	Protéine	forte
1574 cm <sup>-1</sup>	C=C, def. N-H, élongation C-N (Amide II)	moyenne
1606 cm <sup>-1</sup>	Phénylalanine	moyenne
1658 cm <sup>-1</sup>	Lipides insaturés, Amide I	forte
2930 cm <sup>-1</sup>	Elongation C-H	forte
3060 cm <sup>-1</sup>	Elongation des aromatiques C=C-H	

La différence en termes de composition chimique des bactéries permet de créer une discrimination lorsqu'une ACP est effectuée sur les spectres. À l'aide d'un des jeux de données acquis pendant cette thèse, il est possible d'observer une séparation par espèces en se basant sur les différentes variables qui composent le spectre.

En réalisant une ACP sur les spectres, une visualisation du graphique des scores en trois dimensions des quatre espèces est représentée Figure 23. Les quatre espèces sont correctement séparées avec seulement trois composantes représentant 65% de la variance totale du jeu de données global. Sont ensuite représentés les scores de chaque individu sur chacune des trois composantes. La première composante PC1 tend à séparer les spectres de l'espèce *S. epidermidis* (rouge) des spectres *P. syringae* (bleu foncé). Au niveau de la deuxième composante, on observe plutôt la séparation de *P. fluorescens* (vert) du reste des espèces. Puis finalement, à l'aide de la troisième composante, les spectres de *E. coli* (bleu clair) sont discriminés et écartés des autres spectres.

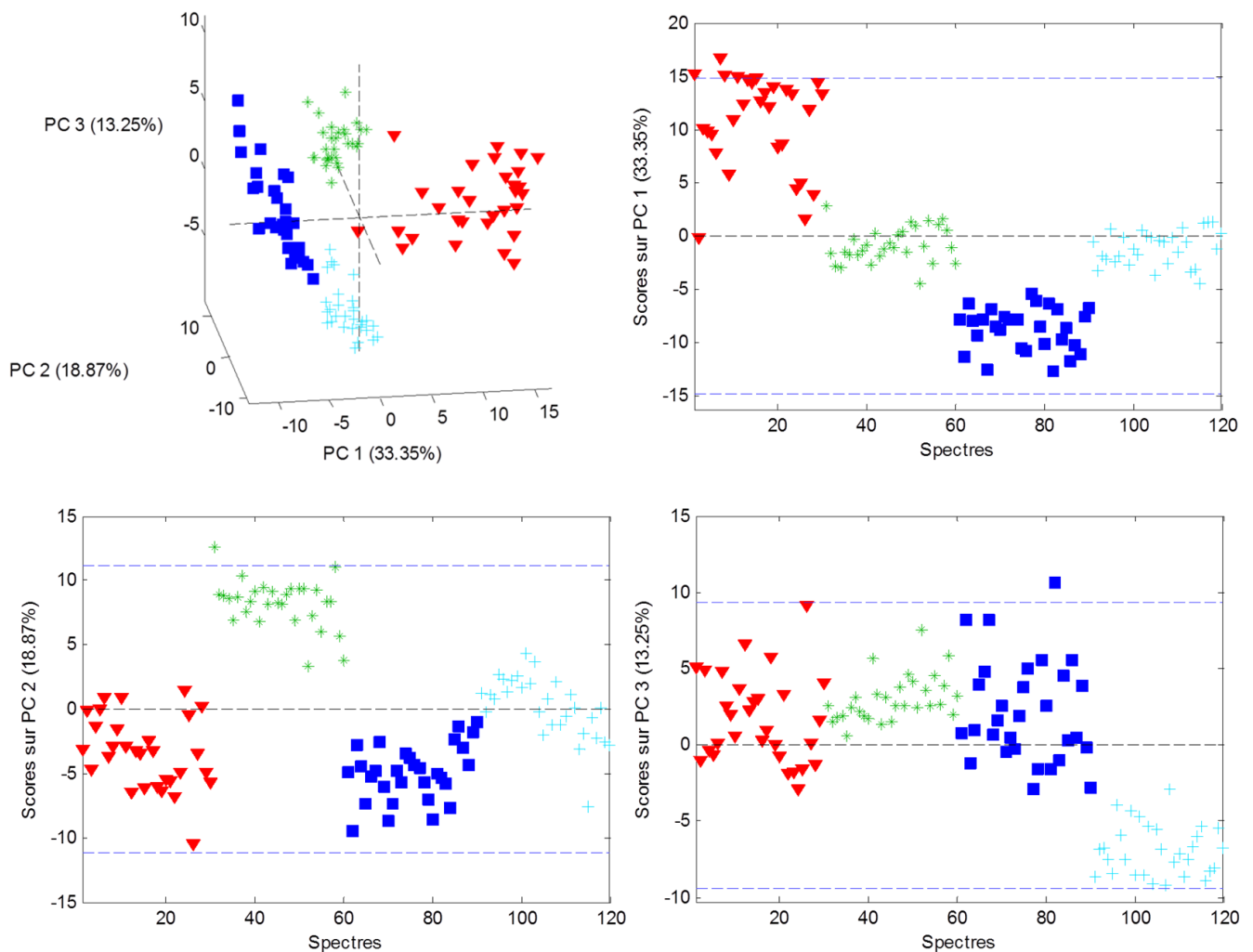
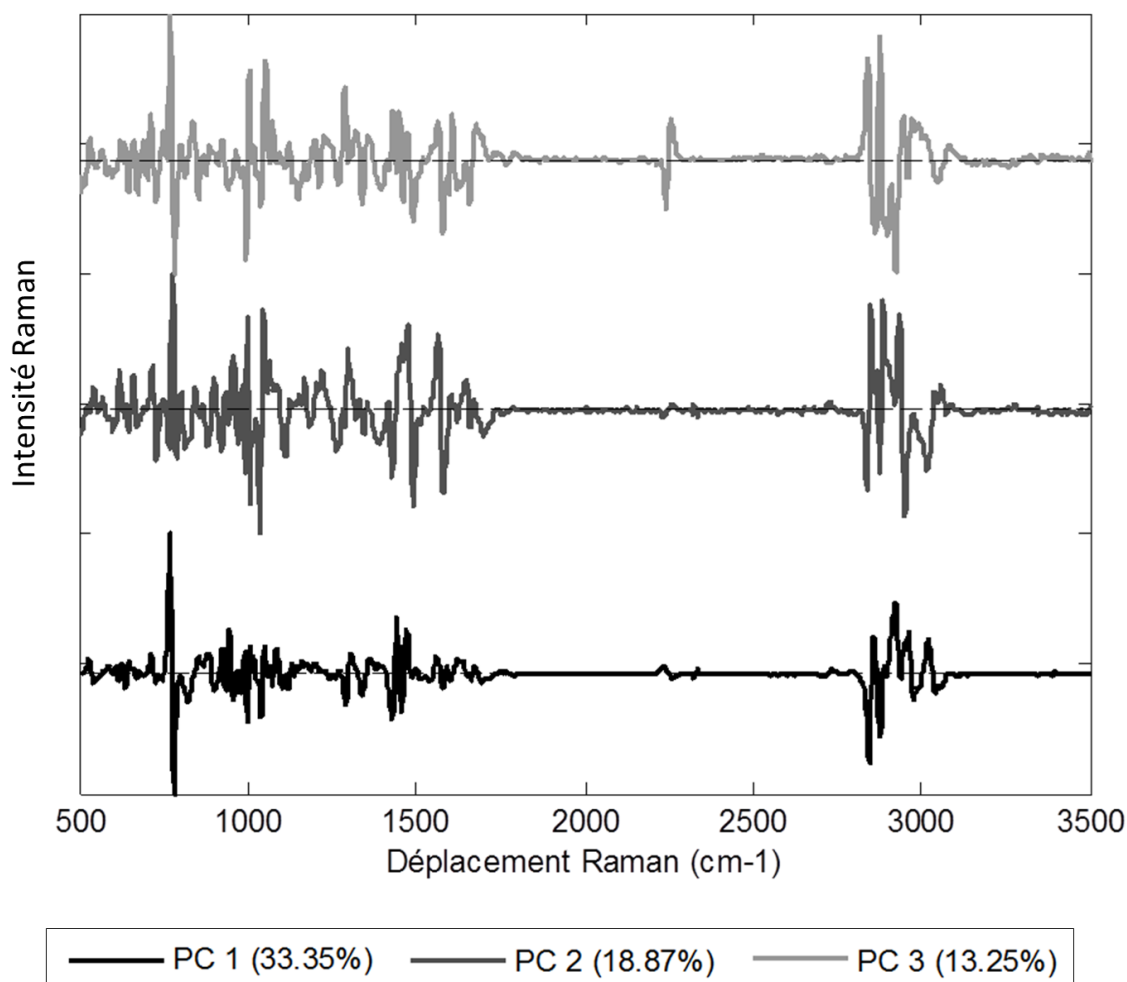


Figure 23 - Séparation des espèces par ACP. Représentation des scores associés.

Le graphique des *loadings* (Figure 24) propose une visualisation des variables permettant la séparation des spectres. Ainsi, on observe qu'une des principales bandes permettant la discrimination des espèces selon la première composante est la bande Raman présente à  $780\text{ cm}^{-1}$  correspondant à la présence de phénylalanine, cytosine et uracile au sein de la cellule. En règle générale, ce n'est pas une seule contribution en particulier qui est à l'origine de la séparation des espèces mais plutôt un ensemble de variables réparties dans l'intégralité du spectre.



Il est donc possible de montrer des différences entre les espèces en se basant sur la composition chimique des cellules, révélée par la microspectroscopie Raman. Cette première exploration permet de montrer que les choix instrumentaux permettent d'observer des différences significatives entre les espèces même si elles sont faibles. C'est d'ailleurs une condition nécessaire à la construction du modèle de classification.

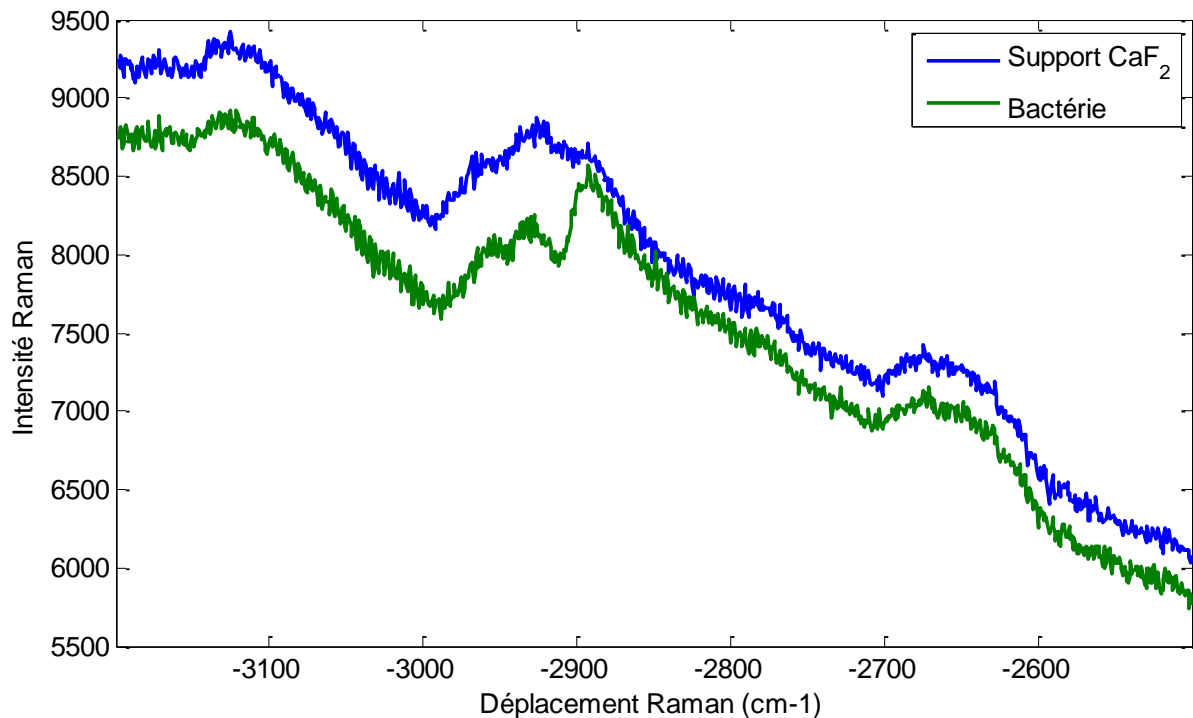


## **1.4. Potentiel de la spectroscopie M-CARS et EPI-CARS**

En parallèle des travaux menés dans le cadre de ma thèse, un autre projet s'est développé dans le laboratoire Xlim de l'Université de Limoges. Le but de ces travaux était le développement d'un nouveau dispositif Raman M-CARS [60] (de l'anglais *Multiplex Coherent Anti-Stokes Raman Spectroscopy*) permettant d'acquérir des spectres beaucoup plus rapidement qu'il est possible en spectroscopie Raman conventionnelle. Il serait alors possible d'intégrer un tel système au sein d'un cytomètre en flux liquide à haut débit. Ce couplage devrait permettre l'analyse de plusieurs milliers de particules par seconde. En intégrant un modèle de classification suffisamment robuste au système, l'instrument devrait satisfaire les exigences d'un détecteur biologique en temps réel.

Nous nous sommes rendus dans les laboratoires du Xlim afin de comparer les résultats obtenus via la spectroscopie Raman conventionnelle et la technique M-F-CARS (*Multiplex Forward Coherent Anti-Stokes Raman Spectroscopy*). Les travaux menés lors de cette étude se sont déroulés à l'aide d'un spectromètre Raman HR Evolution (Horiba Jobin-Yvon) équipé d'un détecteur CCD Synapse, d'un laser Infrarouge à 1064nm, d'un objectif x60, en moyennant 5 accumulations de 1 seconde. L'acquisition de spectres sur de petits objets comme les bactéries à l'aide du système M-F-CARS est pour le moment une tâche complexe qui nécessite des travaux de recherche plus approfondis. Lorsque les spectres sont acquis sur cellule unique, celles-ci sont brûlées à cause de la puissance moyenne trop importante impactant les cellules. Cependant nous avons tout de même réussi à faire l'acquisition de spectre sur des amas de particules. Le spectre enregistré est présenté sur la Figure 25. En bleu est représenté le spectre du support de CaF<sub>2</sub> sur lequel est l'échantillon et en rouge le spectre d'un amas de cellules. Le signal observé à -2900 cm<sup>-1</sup> correspondant à la présence de groupement C-H, traduit la présence de matière organique, qu'il est possible de retrouver dans les spectres Raman conventionnels des bactéries. Cependant, les spectres acquis ne permettent en aucun cas de différencier des espèces entre-elles à cause de la qualité du spectre trop pauvre en information. Il convient de préciser que la zone spectrale de la *fingerpint* pour le support et la bactérie ne présentent aucune différence. Il est difficile pour le moment d'observer quoique ce soit pour les bactéries dans cette région. Les premiers travaux relatant l'observation de signaux dans la *fingerpint* à l'aide de cette technique datent de 2010 et ont été acquis sur des

cellules de plus de 20  $\mu\text{m}$  de diamètre [61], objet beaucoup plus gros qui dissipe certainement plus facilement l'énergie du laser.

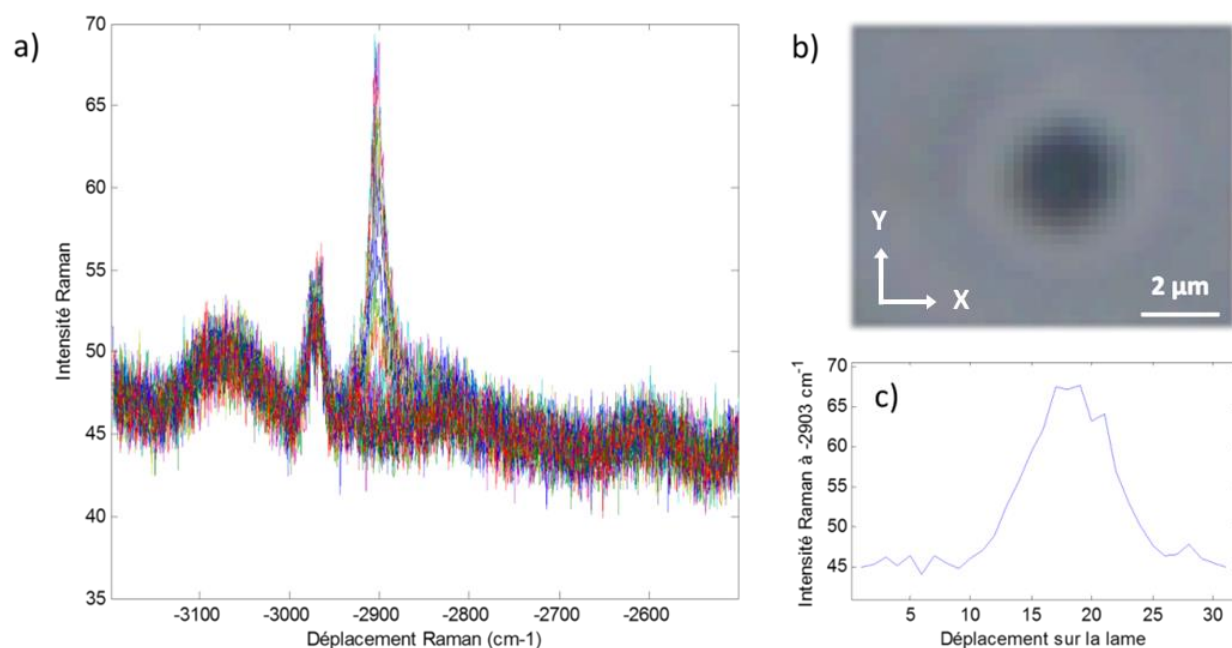


**Figure 25 - Spectres du support en  $\text{CaF}_2$  en bleu et d'une bactérie en vert**

Nous avons ensuite modifié quelques peu le système afin d'utiliser le spectromètre en mode EPI-CARS. L'avantage de l'EPI-CARS par rapport au M-F-CARS, réside dans la possibilité d'augmenter la puissance crête pour améliorer la qualité spectrale tout en conservant une puissance moyenne suffisamment faible pour ne pas brûler les particules. Cette technique est plus appropriée pour les petits objets comme les bactéries. En EPI-CARS, un signal est également observable dans la région des C-H à  $-2900\text{cm}^{-1}$  et aucun signal n'est observé dans la *fingerprint* ( $-400$  à  $-2000\text{cm}^{-1}$ ). Une acquisition transversale a été réalisée sur une cellule représentée sur la Figure 26.

Le rapport signal sur bruit des spectres est très mauvais (Figure 26 a)) mais il est tout de même possible d'observer une variation de l'intensité de la bande centrée à  $-2903\text{cm}^{-1}$  lorsque la table se déplace selon l'axe X. Cette variation d'intensité est représentée sur la Figure 26 c) et correspond à la forme de la cellule photographiée à l'aide du microscope (Figure 26 b)). Il y a donc bien des contributions liées à la cellule mais, pour l'heure, les spectres ne

sont pas d'une qualité suffisante pour permettre de différencier les espèces les unes des autres.



**Figure 26 - Analyse EPI-CARS transversale d'une bactérie**

Cette étude est une première approche dans la démarche de réunification des deux projets ASTRID de la DGA à savoir NEOSPRAM (Nouveaux procédés de spectroscopie Raman pour l'analyse d'agents biologiques) et SIBIRAM (Signature des agents biologiques par microspectroscopie Raman) faisant tous deux partie du projet RACSBio.

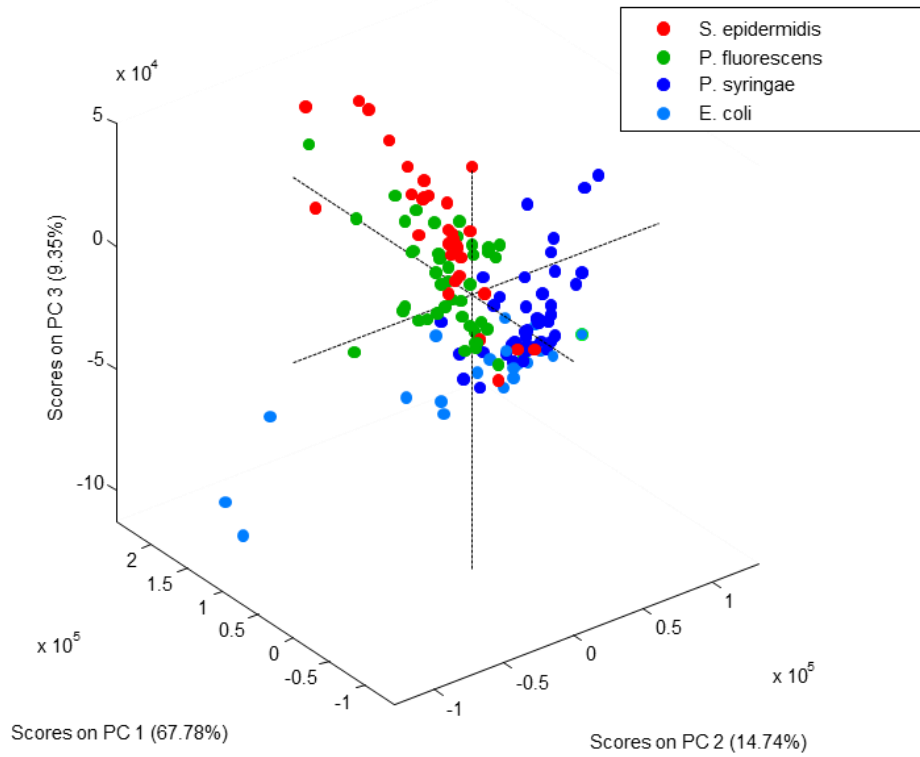
## 2. Classification par méthode SVM

Jusqu'à maintenant, tous les développements qui ont été menés ont eu pour but d'optimiser la mesure Raman avec pour objectif, l'obtention de spectres présentant une qualité suffisante pour la discrimination des espèces. Tout au long du projet de thèse, différents échantillons ont été générés permettant donc l'acquisition de plusieurs séries de mesures mais toujours dans des conditions identiques d'acquisition.

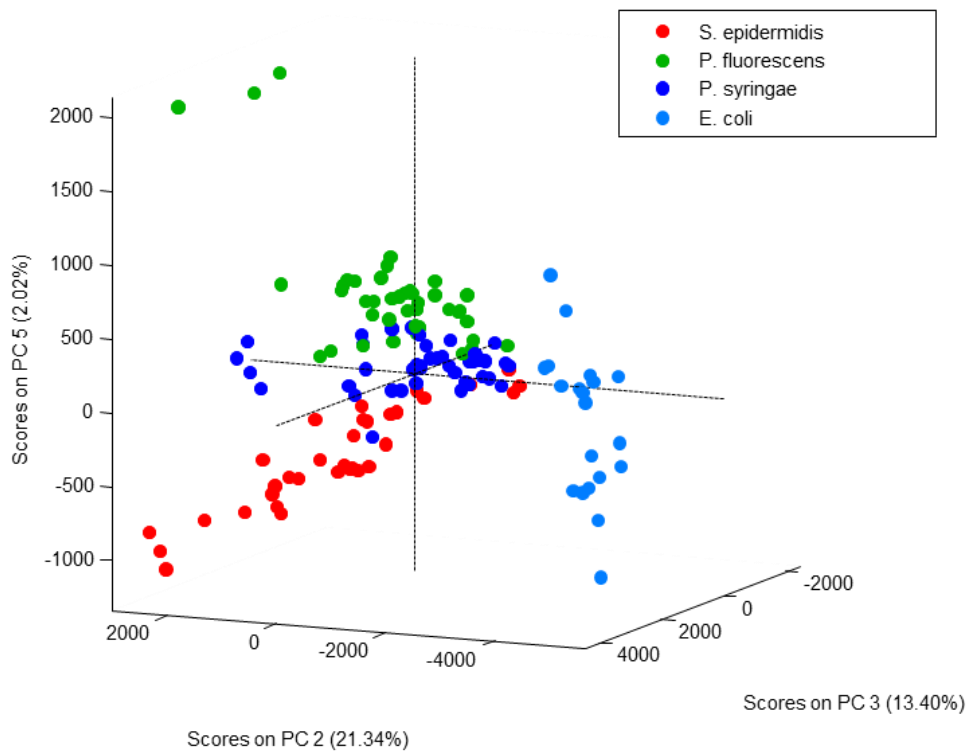
## **2.1. Optimisation des prétraitements des données spectrales**

Au cours des mesures et entre plusieurs séries de mesures, il existe une variabilité naturelle qui s'installe et qui peut complexifier la discrimination des espèces lors d'une analyse. Il existe aussi une variabilité naturelle entre les bactéries d'une même espèce. Même si les souches utilisées sont identiques d'un lot à un autre, avec des conditions de culture et d'échantillonnage identiques. Ainsi, il existe des facteurs incontrôlables comme la taille finale de chaque cellule, leur résistance, les conditions de transport, la température et l'humidité ambiante. Puisqu'il existe une multitude de facteurs qui peuvent influencer directement l'acquisition spectrale, il faut apporter des corrections aux spectres, des prétraitements mathématiques, afin de minimiser l'influence de ces facteurs pour ne conserver que l'information pertinente qui permettra de discriminer les espèces les unes des autres. Ces prétraitements permettent donc de réduire la variabilité entre les échantillons d'un même groupe et d'accroître la variabilité entre des groupes d'échantillons.

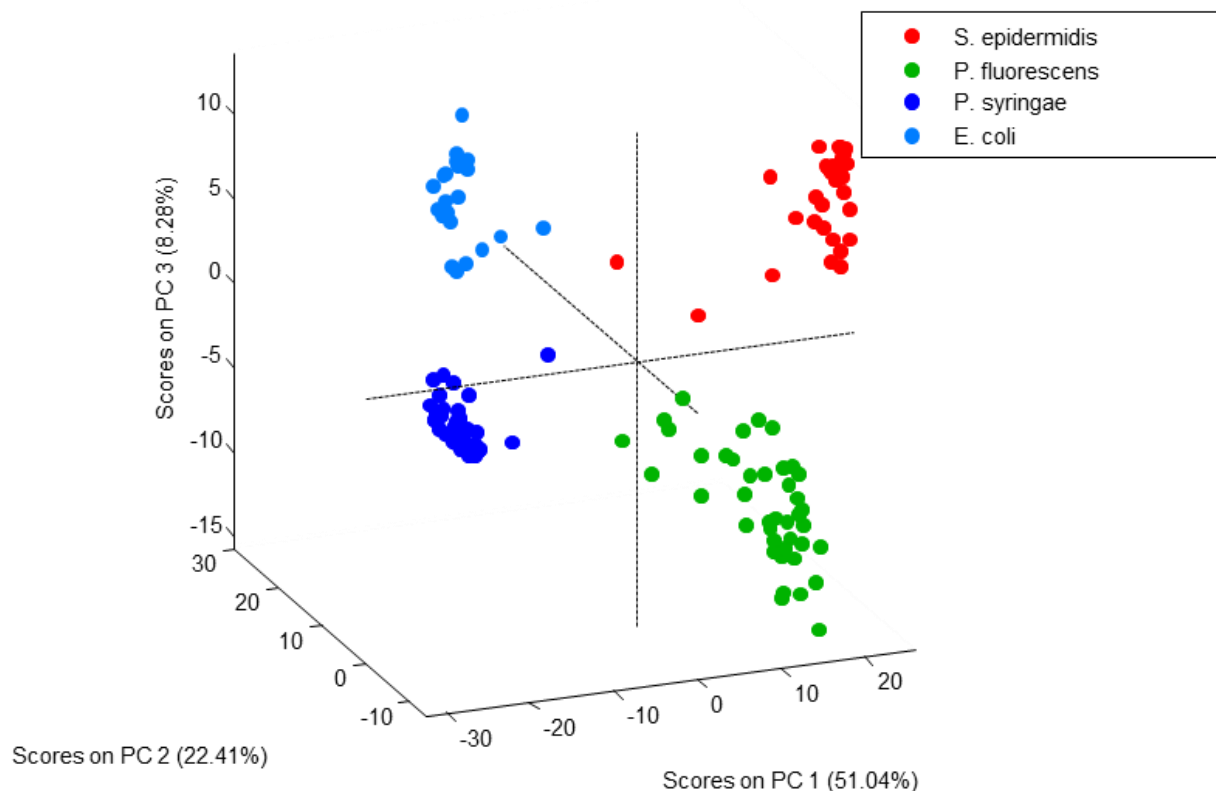
Sur la Figure 27 sont représentées trois analyses en composantes principales réalisées sur le même jeu de données comprenant nos quatre espèces de bactéries sur le domaine  $500\text{ cm}^{-1}$  à  $3500\text{ cm}^{-1}$ . Les prétraitements appliqués sur le jeu de données sont différents et ont une influence directe sur la qualité de séparation des données. Ainsi, sans prétraitement particulier (centrage par la moyenne), il est impossible de séparer les espèces. Lorsqu'on applique une dérivée première, permettant de corriger un possible fond de fluorescence, la séparation commence à se dessiner mais n'est toujours pas suffisamment bonne car certains groupes se superposent encore. On voit ici que la variance d'une faible évolution de la ligne de base peut ainsi masquer les faibles différences spectrales entre les bactéries. Lorsqu'un prétraitement optimal est appliqué, comme dans le dernier cas (dérivée 2<sup>nde</sup>, SNV, et centrage par la moyenne) la séparation entre les groupes est nette et il est désormais possible de construire un modèle. La méthode SNV est une méthode de normalisation. Elle est très influente ici car on peut imaginer que le volume confocal ne couvre pas toujours parfaitement la bactérie analysée.



### Aucun prétraitement



### Dérivée 1<sup>ère</sup>



## Dérivée 2<sup>nde</sup> + SNV

**Figure 27 - Influence du prétraitement appliqué sur la séparation des spectres lors d'une ACP.**

Après cette première étude du prétraitement et avant de commencer la construction d'un modèle de classification, il nous a semblé que nous devions faire une étude plus systématique afin de savoir s'il existait une combinaison de prétraitements permettant de maximiser l'écart inter-groupe et de minimiser l'écart intra-groupe. Pour cela, nous avons fixé à trois le nombre maximum de prétraitements à appliquer consécutivement sur l'ensemble des spectres. Au total ce sont 23 prétraitements qui ont été testés (Tableau 5).

**Tableau 5- Liste des prétraitements.**

Prétraitement	Options
Aucun	/
Correction de la ligne de base	Points spécifiés
Correction de la ligne de base	Moindres carrés pondérés
Lissage	ordre: 0, fenêtre: 7 pts
Lissage	ordre: 0, fenêtre: 15 pts
Lissage	ordre: 0, fenêtre: 21 pts
Lissage	ordre: 1, fenêtre: 7 pts
Lissage	ordre: 1, fenêtre: 15 pts
Lissage	ordre: 1, fenêtre: 21 pts
Lissage	ordre: 2, fenêtre: 7 pts
Lissage	ordre: 2, fenêtre: 15 pts
Lissage	ordre: 2, fenêtre: 21 pts
Dérivée 1 <sup>ère</sup>	ordre: 2, fenêtre: 7 pts
Dérivée 1 <sup>ère</sup>	ordre: 2, fenêtre: 15 pts
Dérivée 1 <sup>ère</sup>	ordre: 2, fenêtre: 21 pts
Dérivée 2 <sup>nde</sup>	ordre: 2, fenêtre: 7 pts
Dérivée 2 <sup>nde</sup>	ordre: 2, fenêtre: 15 pts
Correction de tendance linéaire	/
Correction de l'effet multiplicatif (MSC)	/
Normalisation de l'aire sous la courbe	Aire = 1
Normalisation SNV	/
Centrage par la moyenne	/
Centrage par la médiane	/

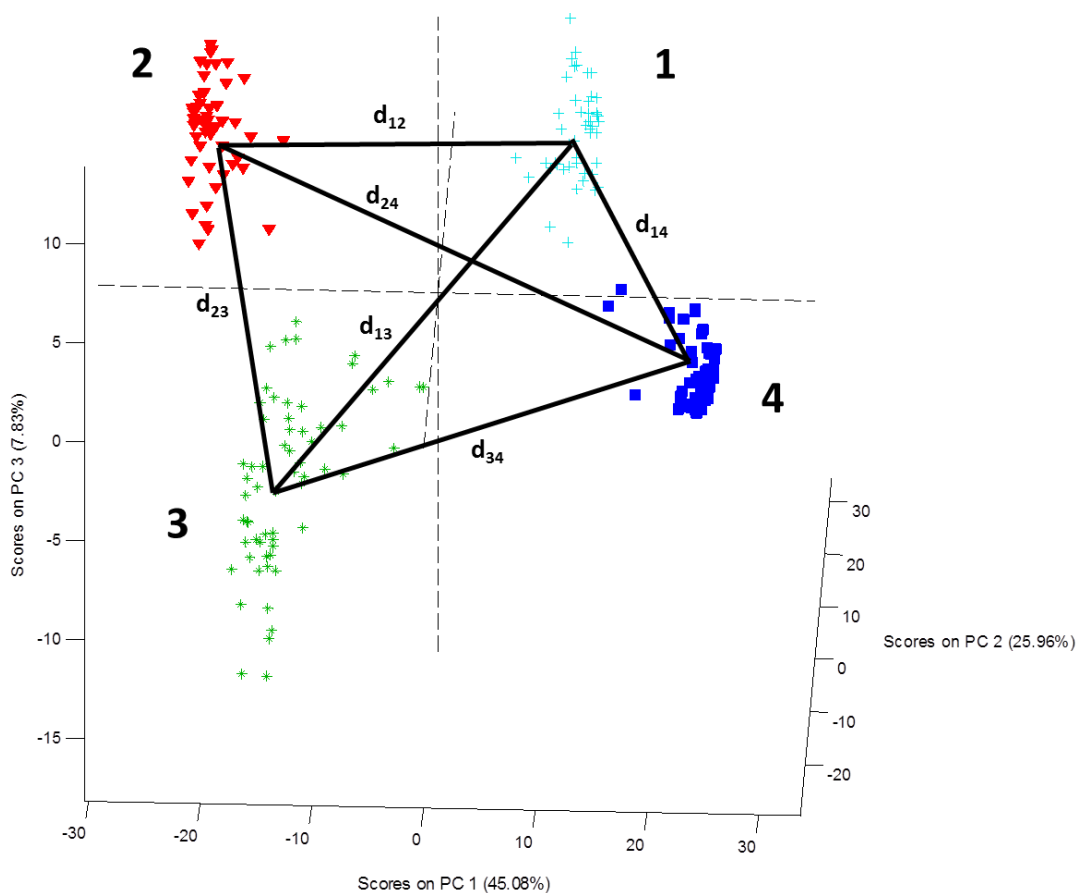
C'est donc un total de plus de 12000 combinaisons qui ont été réalisées et analysées. Pour chacune des combinaisons, un modèle ACP a été construit et des calculs de distances effectués.

Le centre de chacune des classes est déterminé à partir du graphique des scores selon les trois premières composantes. Le centre de la classe de chacune des quatre espèces est défini comme la moyenne des scores sur 3 PC. Puis, la distance euclidienne entre chacun des centres de classe est calculée (Eq.14).

$$distance(Espèce_A, Espèce_B) = \sqrt{(x_B - x_A)^2 + (y_B - x_A)^2 + (z_B - z_A)^2} \quad (\text{Eq.14})$$

où  $x_i$ ,  $y_i$  et  $z_i$  représentent les coordonnées de l'espèce  $i$  dans un espace à 3 composantes principales.

Ensuite, il faut procéder au calcul du critère de distance entre les centres de chaque groupe (Eq.15). Plus cette distance globale est grande, meilleure est la séparation des groupes. Il arrive parfois qu'un groupe soit tellement éloigné des trois autres, que le résultat



**Figure 28 – Représentation du calcul des distances entre le centre de chacune des diverses classes.**



est faussé puisque dans ce cas il n'assure pas la séparabilité maximale entre tous les groupes. Ces cas particuliers seront écartés manuellement en fin d'analyse.

$$distance\ totale = \sum (d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34})$$

(Eq.15)

où  $d_{ij}$  représente la distance entre les deux centres de deux classes  $i$  et  $j$ . La Figure 28 représente le calcul de ces distances entre les centres des différentes classes.

Selon les prétraitements utilisés, une normalisation des données n'est pas toujours réalisée, c'est pourquoi il est préférable de calculer le barycentre (ou centre de gravité, CDG) de l'ensemble des points, puis la distance de chacun des centres des différentes classes au barycentre (Eq.16).

$$distance\ CDG = \sum (d_{1/CDG}, d_{2/CDG}, d_{3/CDG}, d_{4/CDG})$$

(Eq.16)

où  $d_{i/CDG}$  représente la distance entre le centre d'une classe et le centre de gravité du modèle à 3 composantes principales. Il est maintenant possible de normaliser la distance totale par la distance au barycentre (Eq.17) afin de supprimer l'influence de l'ordre de grandeur des axes du graphique des scores.

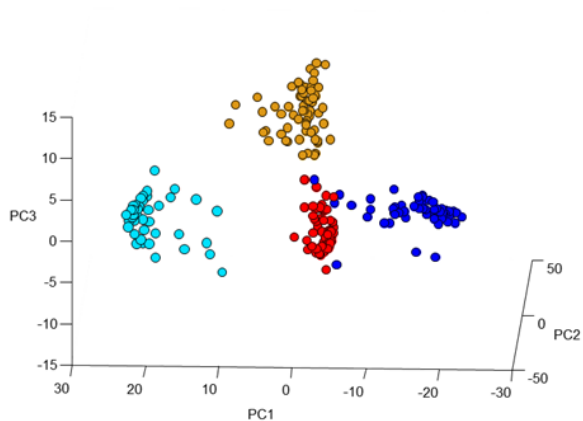
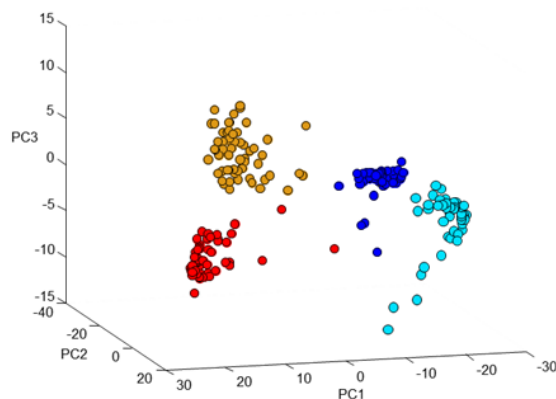
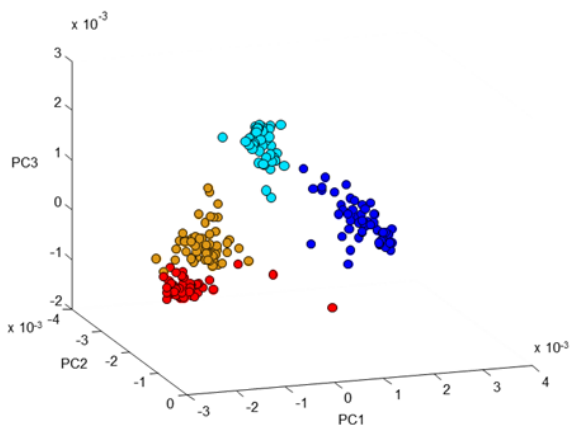
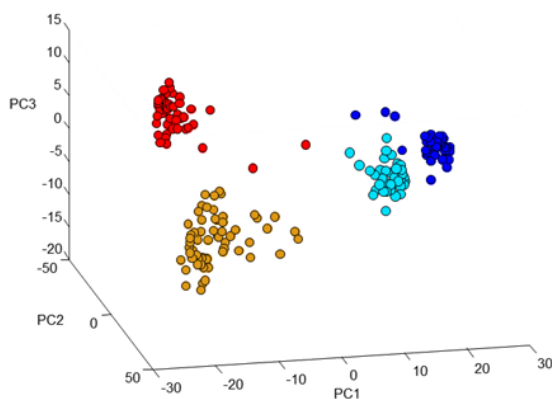
$$distance\ normalisée = \frac{distance\ totale}{distance\ CDG}$$

(Eq.17)

C'est donc ce rapport qui est considéré comme notre critère de discrimination pour trouver les meilleures combinaisons de prétraitements. À l'issue du calcul, les analyses présentant les meilleures séparations (top 10%) ont été conservées et une vérification manuelle a été opérée afin d'écarter les cas aberrants.

Sur la Figure 29 sont représentées différentes projections des spectres sur les trois premières composantes avec des prétraitements différents. L'étude révèle que plusieurs combinaisons sont possibles pour séparer les différents groupes les uns des autres (Figure 29).

Mais dans l'ensemble, les prétraitements récurrents permettant une bonne séparation des groupes sont les dérivées 1<sup>ère</sup> et 2<sup>nde</sup>, la normalisation SNV. Les dérivées permettent d'apporter une correction lorsqu'un fond de fluorescence est présent et une correction sur des variations de ligne de base. Le prétraitement SNV quant à lui, permettra une normalisation des intensités Raman d'un spectre à un autre.

Dérivée 2<sup>nde</sup> (ordre 2, 15 points)+ SNVDérivée 2<sup>nde</sup> (ordre 2, 7 points)+ SNVDérivée 1<sup>nde</sup> (ordre 2, 15 points)+ SNVDérivée 1<sup>nde</sup> (ordre 2, 7 points)+ SNV

**Figure 29 - Séparation possible obtenue à l'aide de différents prétraitements appliqués sur les spectres.**

Toutefois, lors de la construction d'un modèle final de classification, il sera nécessaire de réaliser une étude plus approfondie des résultats sur les différents prétraitements.

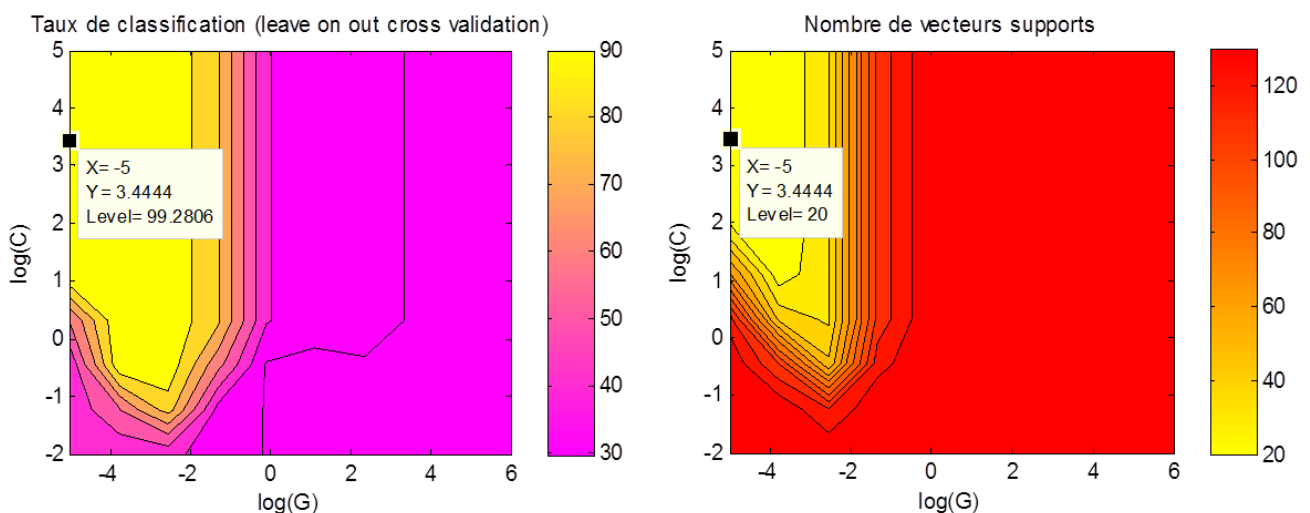
## 2.2. Construction et validation du modèle de classification

### SVM

Un des objectifs de ces travaux est de démontrer qu'il est possible de créer un modèle de classification robuste, pour des bactéries, à partir de leur signature chimique acquis par microspectroscopie Raman. Comme il a été vu précédemment, il existe plusieurs méthodes permettant la classification de ces agents microbiologiques mais la méthode des SVM (Machine à Vecteurs de Support) s'est révélée être la plus performante. Pour la construction d'un modèle de classification, il faut procéder en plusieurs étapes. En premier lieu, une estimation des paramètres C et G est réalisée, puis à l'aide de ces paramètres, le modèle sera ensuite construit sur la base d'une validation croisée, puis finalement nous procéderons à la validation du modèle à l'aide d'un jeu de données indépendant.

#### 2.2.1. Estimation des paramètres SVM

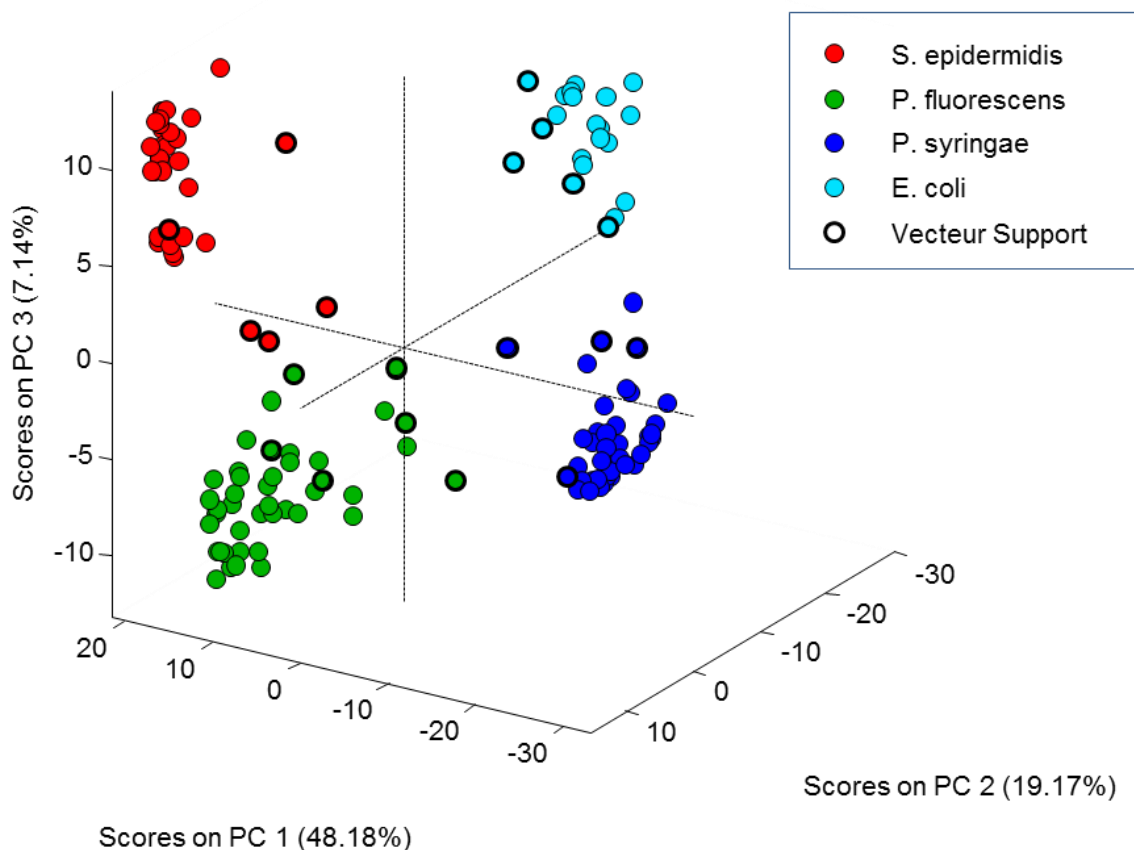
Les modèles de classification que nous construisons sont basés sur la méthode des Kernel utilisant la fonction noyau RBF gaussienne. Pour cette fonction, il est nécessaire de procéder à l'optimisation des méta-paramètres C et G pour éviter tout risque de surentraînement du modèle et une complexité des frontières trop importantes. Pour estimer



**Figure 30 - Optimisation des méta-paramètres C et G pour la construction d'un modèle de classification SVM. Le carré noir représente un optimum pour  $C=10^{3,44}$  et  $G=10^{-5}$  avec un taux de classification de 99,28% et 20 vecteurs supports.**

les méta-paramètres, nous nous basons sur deux critères qui sont le nombre de vecteurs supports nécessaires pour construire les frontières et le taux de bonne classification obtenu par validation croisée (*Leave One Out Cross Validation*).

Pour cette estimation il est possible d'utiliser les spectres prétraités ou bien les scores obtenus suite à une analyse en composantes principales. Pour une meilleure visualisation des résultats, nous préférons utiliser la projection des individus sur chacune des composantes principales que les spectres Raman directement. Le jeu de données utilisé pour cette étude est composé de 139 spectres appartenant à quatre espèces différentes afin de prendre en compte la variabilité biologique des espèces en considération. Pour déterminer la valeur optimale de chacun des paramètres, deux critères seront utilisés : minimiser le nombre de vecteurs supports nécessaires pour séparer les classes et maximiser



**Figure 31 - Représentation des vecteurs supports dans le graphique des scores. En rouge, bleu, vert et bleu clair sont représentés les spectres des quatre différentes espèces, en noir les spectres vecteurs supports.**

le pourcentage de bonne classification par validation croisée. La Figure 30 est une représentation de la grille d'optimisation obtenue pour ce jeu de données. Le nombre minimum de vecteurs supports déterminé par l'algorithme est de 20 pour un taux de bonne classification par validation croisée de 99,3% correspondant aux valeurs  $C=10^{3,44}$  et  $G=10^{-5}$ .

Diminuer la valeur du paramètre C permettrait d'augmenter la taille de la marge, mais au détriment du nombre de vecteurs supports nécessaires qui verrait sa valeur augmenter considérablement. De plus, l'augmentation du nombre de vecteurs supports ne permet pas d'augmenter le taux de bonne classification, donc il n'est pas nécessaire d'augmenter ces valeurs. Les vecteurs supports déterminés lors de l'étape précédente correspondent aux spectres les plus en périphérie de chacun des groupes (Figure 31) représentant les quatre différentes espèces analysées. Ils sont symbolisés par les cercles noirs. Ce sont les coordonnées de ces spectres, ainsi que les valeurs C et G qui vont définir les différentes marges qui séparent les groupes.

L'ensemble des calculs que nous avons accompli jusqu'à présent a permis de déterminer la valeur optimale des méta-paramètres C et G pour que nous puissions maintenant construire un modèle de classification.

### **2.2.2. Calcul du modèle de classification SVM**

Pour qu'un modèle de classification soit utilisable, il faut procéder en deux étapes. La première étape consiste en la construction du modèle de classification lui-même et la seconde étape est la validation du modèle à l'aide d'un jeu de données indépendant, c'est-à-dire qui n'a pas servi à construire le modèle. Nous avons à disposition un jeu de données contenant 139 spectres de quatre espèces différentes. Ce jeu est séparé de façon à obtenir un jeu de calibration correspondant au 2/3 de l'ensemble des spectres et en un jeu de validation correspondant à 1/3 de l'ensemble des spectres. Ainsi ce sont 93 spectres qui ont été sélectionnés aléatoirement pour la construction du modèle et 46 spectres pour la validation. Une ACP est effectuée sur le jeu de calibration puis les scores de chacun des spectres sur les dix premières composantes sont conservés expliquant plus de 99% de la variance totale. Les prétraitements appliqués sont une dérivée seconde (polynôme d'ordre 2, 7 points) suivi d'une

SNV et d'un centrage par la moyenne. Une valeur numérique est attribuée à chacun des individus pour définir leur classe d'appartenance (Tableau 6).

**Tableau 6 - Attribution d'une valeur numérique pour chacune des espèces représentées.**

Valeur	Classe
1	<i>S. epidermidis</i>
2	<i>P. fluorescens</i>
3	<i>P. syringae</i>
4	<i>E. coli</i>

Une optimisation des paramètres C et G est effectuée afin de s'assurer que les valeurs préalablement calculées sont toujours applicables à la construction de ce modèle. Le modèle de classification SVM est construit à l'aide des paramètres C et G déterminés précédemment ( $C=10^{3.44}$  et  $G=10^{-5}$ ) en utilisant la fonction noyau RBF gaussienne. Pour calculer le modèle nous utilisons les fonctions de la boîte à outils Matlab LIBSVM [62]. Lorsque l'étape de calcul des marges et des vecteurs supports est effectuée, le modèle est calculé. Il faut désormais valider le modèle à l'aide d'un jeu de données indépendant. Pour cela, nous reprenons le modèle ACP calculé à l'aide du jeu de calibration (93 spectres), puis une projection du jeu de validation est effectuée sur les 10 composantes principales du modèle. Les scores des individus du jeu de validation sont conservés et injectés dans le modèle de classification avec leurs classes également référencées comme pour le jeu de calibration (Tableau 6). Le modèle va alors déterminer la classe des nouveaux individus et celle-ci sera comparée à la classe référencée. Le taux de bonne classification sera calculé et fera gage de la puissance du modèle de classification SVM.

Pour le modèle de classification SVM calculé, le taux de bonne classification du jeu de validation est de 100%. Ce résultat n'est pas représentatif de la réalité mais donne une idée de la puissance du modèle. En effet, il est tout à fait possible de trouver un maximum local. C'est pourquoi toutes les opérations, allant du tirage aléatoire du jeu de calibration/validation jusqu'au calcul du taux de bonne classification du jeu de validation, ont été réalisées 1000 fois pour obtenir une vision plus réaliste de la puissance du modèle se rapprochant ainsi d'une

procédure de *bootstrap*. Toutes les étapes du processus d'estimation du pourcentage de bonne reconnaissance des classes suivent le schéma présenté Figure 32.

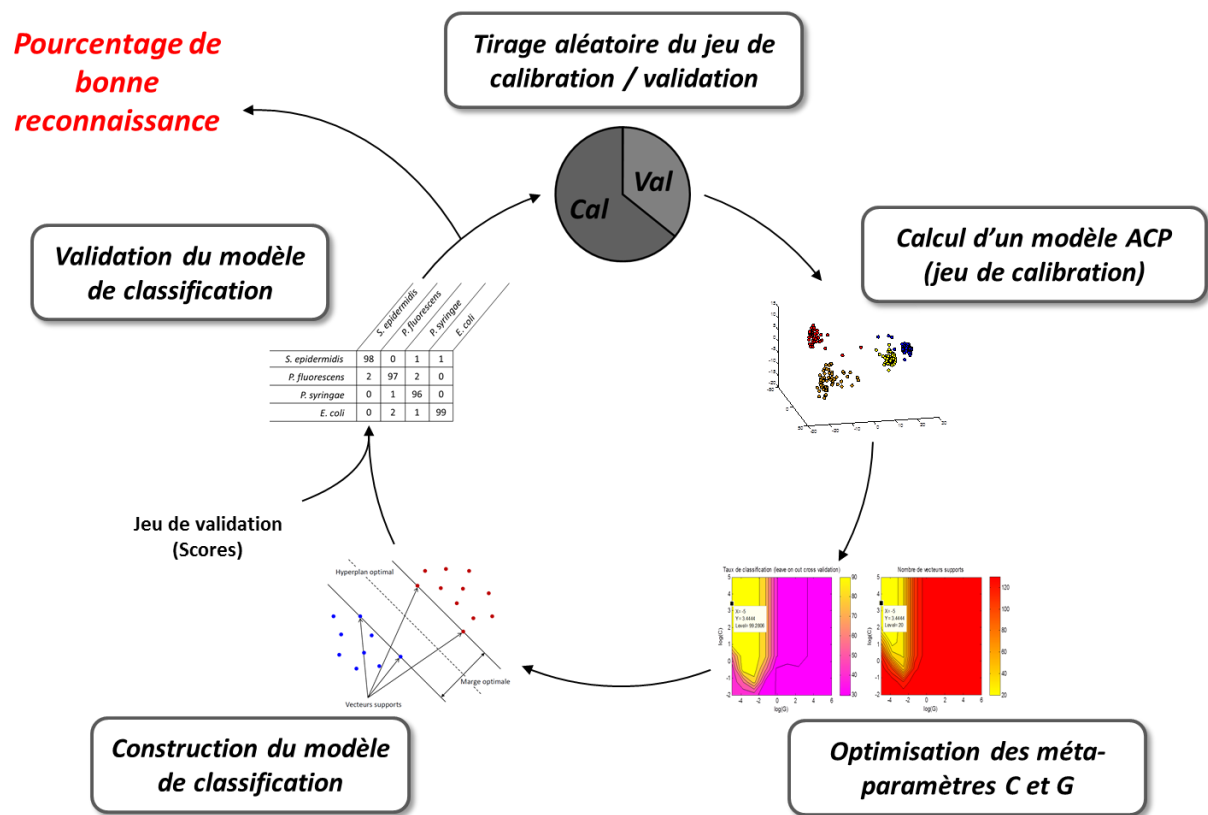
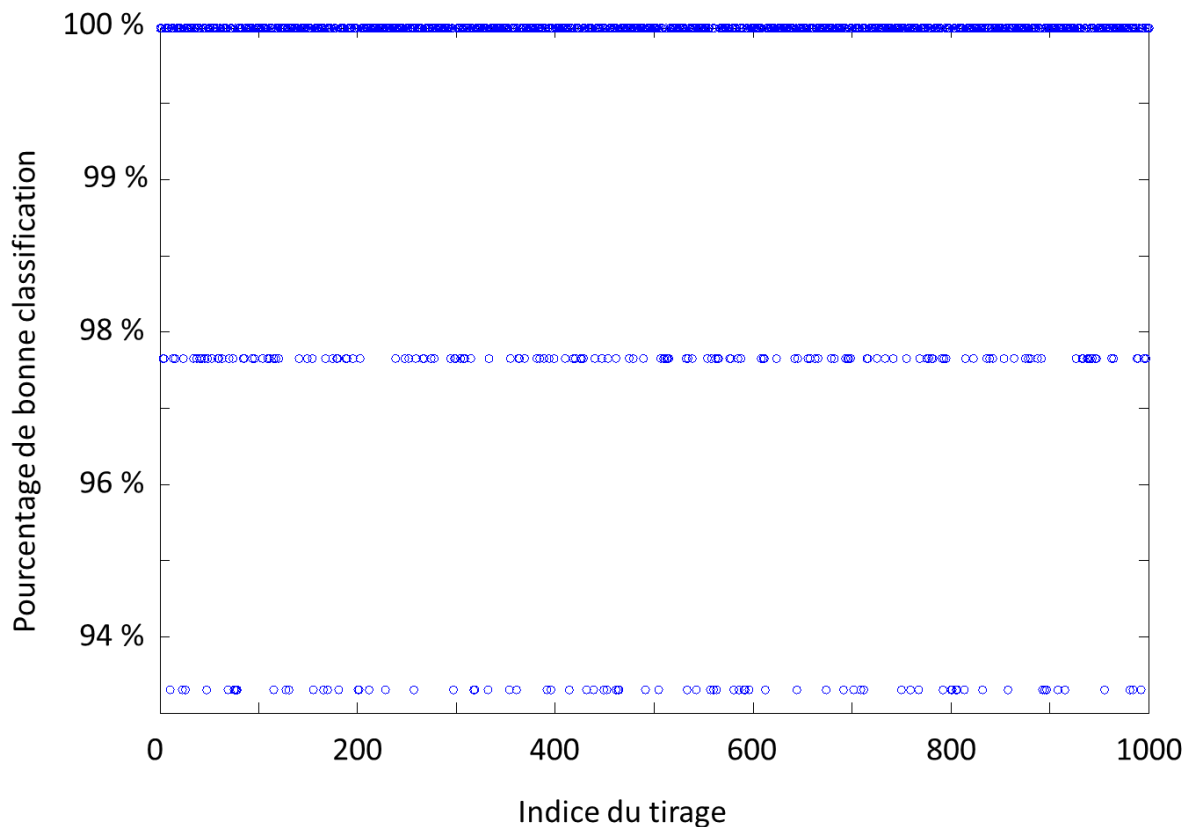


Figure 32 - Schéma des étapes pour la détermination du pourcentage de bonne reconnaissance de la classification SVM.

Selon le tirage qui aura lieu, le modèle de classification SVM n'utilisera pas les mêmes individus comme vecteurs supports pour calculer la limite de chacune des classes. Pour chaque tirage, les marges sont recalculées. Certains individus s'écartant des individus de sa classe augmentent le risque de mauvaise classification s'ils se retrouvent dans le jeu de validation. Ces individus auront une influence négative sur le taux de bonne classification. Les résultats obtenus pour l'étude sont présentés dans la Figure 33.



**Figure 33 - Pourcentage de bonne classification pour chacun des 1000 tirages réalisés sur les spectres des quatre espèces bactériennes.**

À l'issue de cette étude, pour les 1000 tirages nous avons obtenu :

- 70,3 % des tirages effectués présentent un pourcentage de reconnaissance égal à 100 % de bonne classification ;
- 24,4 % des tirages présentent un pourcentage de reconnaissance de 97,83 % de bonne classification, c'est-à-dire un seul individu mal classé par le modèle ;
- 5,3 % des tirages présentent un pourcentage de reconnaissance égale à 95,65% de bonne classification, c'est-à-dire deux individus mal classés par le modèle.

En moyenne, le taux de bonne classification des individus à l'aide d'un modèle SVM est de 99,24 % pour ces quatre espèces. Les résultats obtenus sont excellents, mais il faut noter qu'ils sont valables seulement pour quatre espèces qui ont été cultivées dans des conditions maîtrisées et qu'elles n'ont subi aucun stress dû à des variations de conditions environnementales. Par conséquent, nous allons poursuivre sur l'évaluation de l'adaptabilité



de notre protocole à des échantillons d'aérosols naturels. Nous démontrons néanmoins ici que nous avons optimisé l'ensemble de la méthodologie à savoir l'acquisition spectrale et la chimométrie prenant en compte la variabilité biologique naturelle des bactéries.

### **3. Adaptation du protocole aux aérosols naturels**

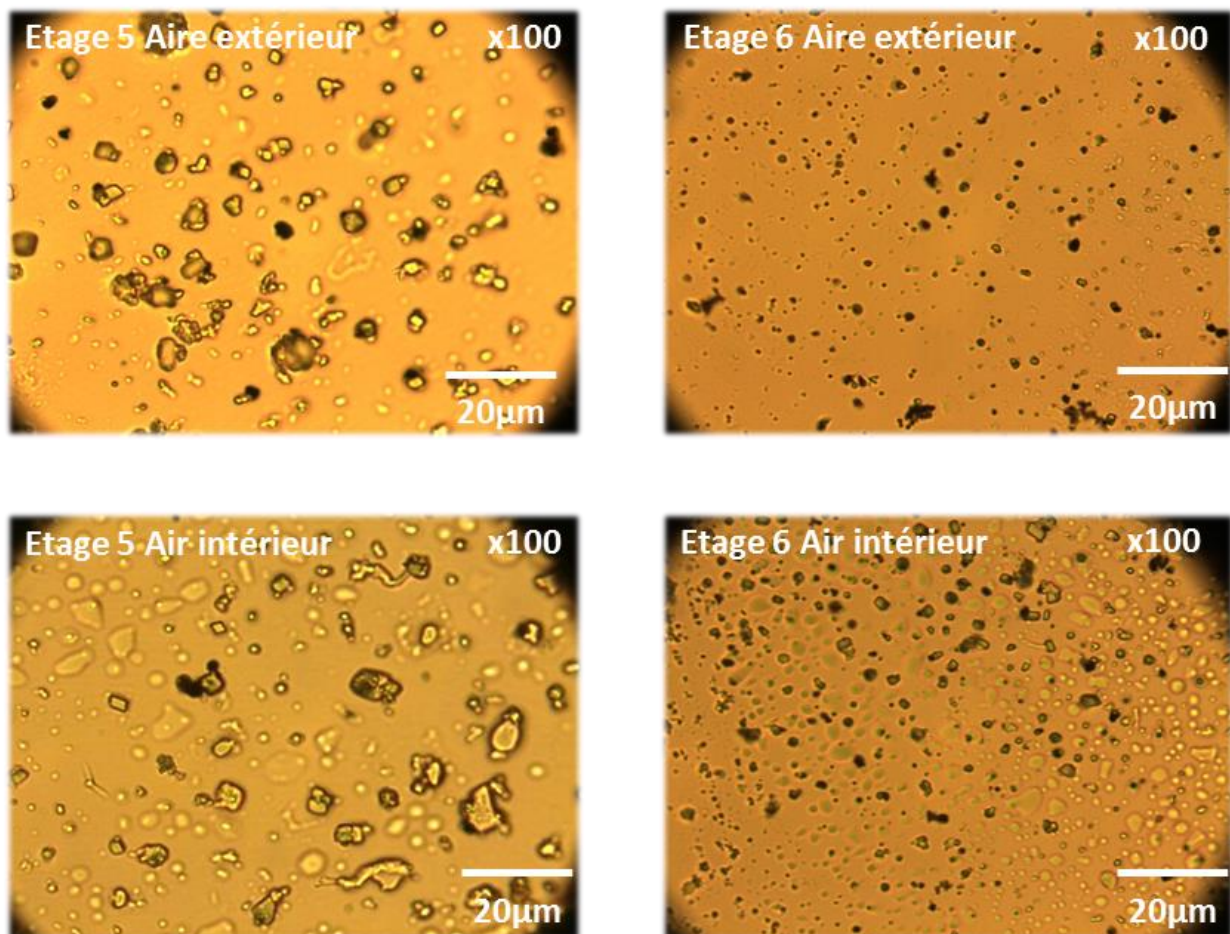
L'air est un milieu riche en particules d'origines diverses et variées. Il est possible de trouver des particules aussi bien de nature minérale, qu'organique ou biologique. Ces particules ont aussi des tailles sur des échelles nanométrique et micrométrique. L'air est donc un moyen de transport rapide et efficace pour les virus, champignons et bactéries. Dès lors, il est primordial de surveiller la qualité bactériologique de l'air afin d'alermer les autorités sanitaires en cas de détection d'espèces à risque. Les études menées jusqu'alors ont démontrées qu'il est possible d'analyser des échantillons cultivés dans un temps relativement court par spectroscopie Raman et que le modèle SVM est un outil puissant de classification. Nous avons donc tout naturellement voulu appliquer tout ce que nous avons fait jusqu'à maintenant à des particules naturelles. Cependant, la complexité des aérosols rend la tâche ardue, nous obligeant alors à apporter des améliorations au protocole que nous avons mis en place.

#### **3.1. *Complexité des aérosols naturels***

Les aérosols naturels sont de nature complexe et présentent une importante diversité en particules. Cette diversité se traduit par des différences de tailles, de morphologie et de composition. On retrouvera aussi bien des particules minérales qu'organiques, de toutes les formes, avec des tailles allant de plusieurs dizaines de micromètres pour les plus grandes à plusieurs dizaines de nanomètres pour les plus petites. C'est pourquoi il est nécessaire de réaliser un premier criblage lors des collectes. Pour cela nous utilisons lors des campagnes de collecte de bruits de fond d'air, un impacteur en cascade de type Andersen. Celui-ci est équipé de différents filtres dont la taille des mailles rétrécit au fur et à mesure que les particules descendent d'étage en étage. Le prélèvement des particules est réalisé dans deux environnements différents à savoir un environnement intérieur et un environnement extérieur car les populations de bactéries et autres particules sont certainement différentes.

### 3.1.1. Choix de la maille de l'impacteur (étage 5 et 6)

Les bactéries sont des micro-organismes de petite taille (environ  $1\ \mu\text{m}$ ), c'est pourquoi nous avons expérimenté deux tailles de mailles. Les mailles du filtre de l'étage 5 laissent passer des particules dont la taille est inférieure à  $10\ \mu\text{m}$  alors que les mailles du filtre de l'étage 6 laissent passer des particules de tailles inférieures à  $5\ \mu\text{m}$ . Des clichés de ces lames ont été enregistrés à l'aide d'un microscope optique avec un objectif x100 ( Figure 34).



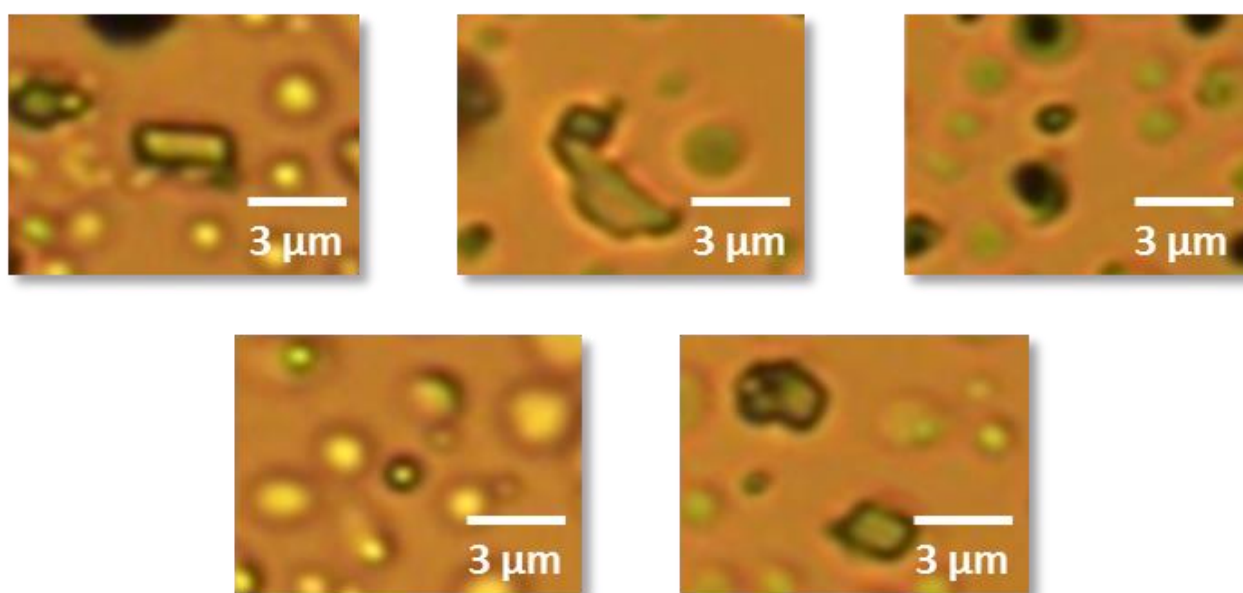
**Figure 34 - Images des prélèvements effectués à l'aide de l'impacteur en cascade aux étages 5 et 6, en milieu extérieur et intérieur (objectif x100).**

La taille des particules récoltées au 5<sup>ème</sup> étage du collecteur est plus importante que celles récoltées au 6<sup>ème</sup> étage et sont trop importantes pour correspondre à des bactéries. Que le prélèvement soit effectué en milieu intérieur ou extérieur, aucune différence de taille

n'est observée. Pour la suite du protocole, nous privilégierons l'utilisation du filtre 5  $\mu\text{m}$  correspondant à la collecte faite au 6<sup>ème</sup> étage de l'impacteur en cascade.

### 3.1.2. Morphologie des particules

Les particules récoltées dans ces bruits de fond d'air ne sont pas seulement de tailles différentes mais également de formes différentes (Figure 35). Certaines particules présentent des formes géométriques bien définies, d'autres particules présentent une morphologie circulaire ou encore des particules dont la forme est plus aléatoire.



**Figure 35 - Diversité morphologiques des particules sur les lames de bruit de fond d'air.**

Cette différence morphologique pourrait constituer un premier critère afin d'écarter certains individus de la liste des particules à analyser au microspectromètre Raman. Cependant, les microbiologistes estiment le ratio entre le nombre de bactéries et le nombre de particules totales à 1/10000. Une étude, datant de 1978, estime qu'il y aurait entre 0,013 et 1,88 bactéries par litre d'air [63] selon les saisons. Il est donc impossible de se baser sur le seul critère morphologique pour repérer les bactéries parmi l'ensemble des particules. Plus encore, la limite de diffraction ne permet pas non plus de faire la différence entre une particule solide de moins d'un micron et une bactérie.

### **3.1.3. Densité particulaire**

La composition en particules n'est pas le seul frein révélé par cette étude. La densité particulaire de l'échantillon va également poser des problèmes, non pas pour l'analyse Raman car la taille du spot laser est de l'ordre du micromètre et nous pouvons toujours trouver des bactéries isolées, mais pour l'étape de microdissection laser. En effet, dans ce processus, le film polymère va couler sur la surface de l'échantillon pour emprisonner la particule à extraire. Or ce film thermosensible va recouvrir une surface circulaire d'une vingtaine de micromètres de diamètre autour de la particule ciblée. Si les particules ne sont pas suffisamment espacées les unes des autres, le risque de prélever plusieurs particules est accru, rendant l'analyse génomique inutile. Si plusieurs bactéries sont prélevées, l'étape d'amplification de l'ARN 16S se fera sur l'ensemble du matériel biologique, pouvant ainsi mener à une mauvaise identification. Dans le cas des échantillons présentés Figure 34, la densité particulaire est tellement importante qu'une microdissection de cellule unique par LCM est impossible. Diminuer le temps durant lequel l'impaction est opérée pourrait être une solution, mais cela se ferait au détriment du nombre de particules impactées, diminuant ainsi le nombre de chance de collecter une bactérie.

Réaliser une cartographie ou acquérir un spectre Raman particule par particule n'est pas une solution envisageable à cause du temps que dureraient les analyses. Essayer de repérer une bactérie parmi toutes les particules minérales de l'échantillon serait trop long et il est impossible d'avoir la certitude d'en trouver suffisamment pour créer un jeu de données exploitable. Il faut donc trouver une alternative et des solutions pour repérer plus facilement les bactéries dans cette masse de particules.

## **3.2. Amélioration du protocole**

Pour remédier à tous les problèmes listés précédemment concernant la complexité des aérosols, plusieurs actions et outils ont été déployés pour permettre un repérage efficace et rapide des agents bactériologiques parmi les particules minérales.

### 3.2.1. Marquage par agent fluorescent

Le problème majeur de ces échantillons est le repérage des bactéries parmi le nombre important de particules présentes sur la lame. Il faut donc mettre en place une étape supplémentaire dans notre protocole pour que le repérage se fasse plus facilement et plus rapidement. Nous avons dans un premier temps cherché à observer l'autofluorescence des bactéries au microscope optique mais les essais ont rapidement aboutis à un échec. Une des techniques employée pour remédier à ce problème est l'utilisation de sondes fluorescentes greffées à l'agent bactériologique. Les microbiologistes utilisent ces sondes fluorescentes pour déterminer des paramètres comme la viabilité des cellules. Les fluorophores se greffent directement à l'ADN des bactéries en ciblant les acides nucléiques. Certains fluorophores marquent aussi bien les cellules vivantes que mortes (DAPI, SYTO9) ; quant à d'autres (SYTOX Blue, PI), ils marquent exclusivement les cellules mortes à cause de la faible perméabilité de la membrane des cellules à Gram + lorsqu'elles sont vivantes. En excitant le fluorophore à la longueur d'onde appropriée, il devient possible d'observer directement les particules marquées au microscope optique.

Pour étudier l'intérêt d'un marquage des bactéries par agents fluorescents, nous avons sélectionné et testé deux fluorophores bien connus : le DAPI et le SYTO9. Les marqueurs SYTO 9 et DAPI sont tous deux des fluorophores qui pénètrent les bactéries Gram+ et Gram-, vivantes et mortes, et se lient à leurs acides nucléiques. Après liaison, leur intensité de fluorescence est sensiblement augmentée, ce qui permet en principe de différencier les particules biologiques des particules non biologiques. Les couples de longueurs d'onde excitation/émission de deux fluorophores sont résumés dans le Tableau 7. Pour l'étude, les fluorophores ont été fixés sur des cellules de *S. epidermidis* et déposés sur lames de CaF<sub>2</sub>.

**Tableau 7 - Couples excitation/émission des deux fluorophores utilisés : DAPI et SYTO9.**

Fluorophore	Excitation	Emission
<b>DAPI</b>	340-380 nm (max 359 nm)	435-485 nm (max 461 nm)
<b>SYTO9</b>	465-495 nm (max 480 nm)	515-555 nm (max 525 nm)

L'observation des lames est effectuée sur un microscope d'épifluorescence équipé d'un objectif x50. Ce microscope est différent et indépendant de celui qui est couplé au système de mesure Raman. Le microscope d'épifluorescence est équipé d'une lampe blanche et de différents jeux de filtres en fonction du fluorophore que nous souhaitons observer. Pour l'observation des bactéries marquées au DAPI, nous utilisons un miroir dichroïque 400 nm (U-MWU2, Olympus) équipé d'un filtre d'excitation passe bande 330 - 385 nm et d'un filtre d'émission passe bande 459 - 481 nm. Pour l'observation des bactéries marquées au SYTO9, nous utilisons un miroir dichroïque 500 nm (U-MWB2, Olympus) équipé d'un filtre d'excitation passe bande 460 - 490nm et d'un filtre d'émission passe bande 524 - 546 nm. La Figure 36 illustre les observations faites sur toutes les cellules marquées (*S. epidermidis*, *P. fluorescens* et *E. coli*) par les deux agents fluorescents.

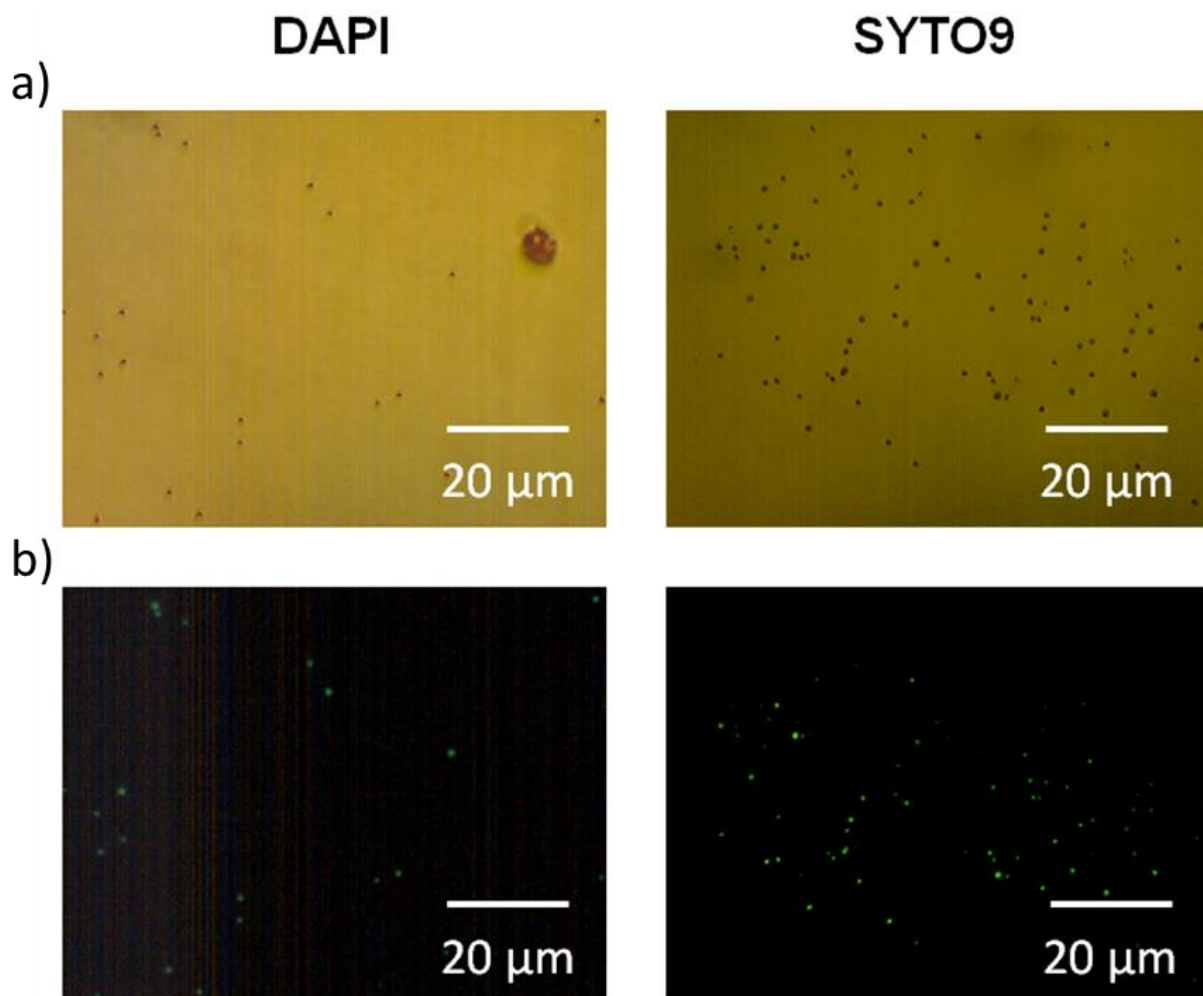
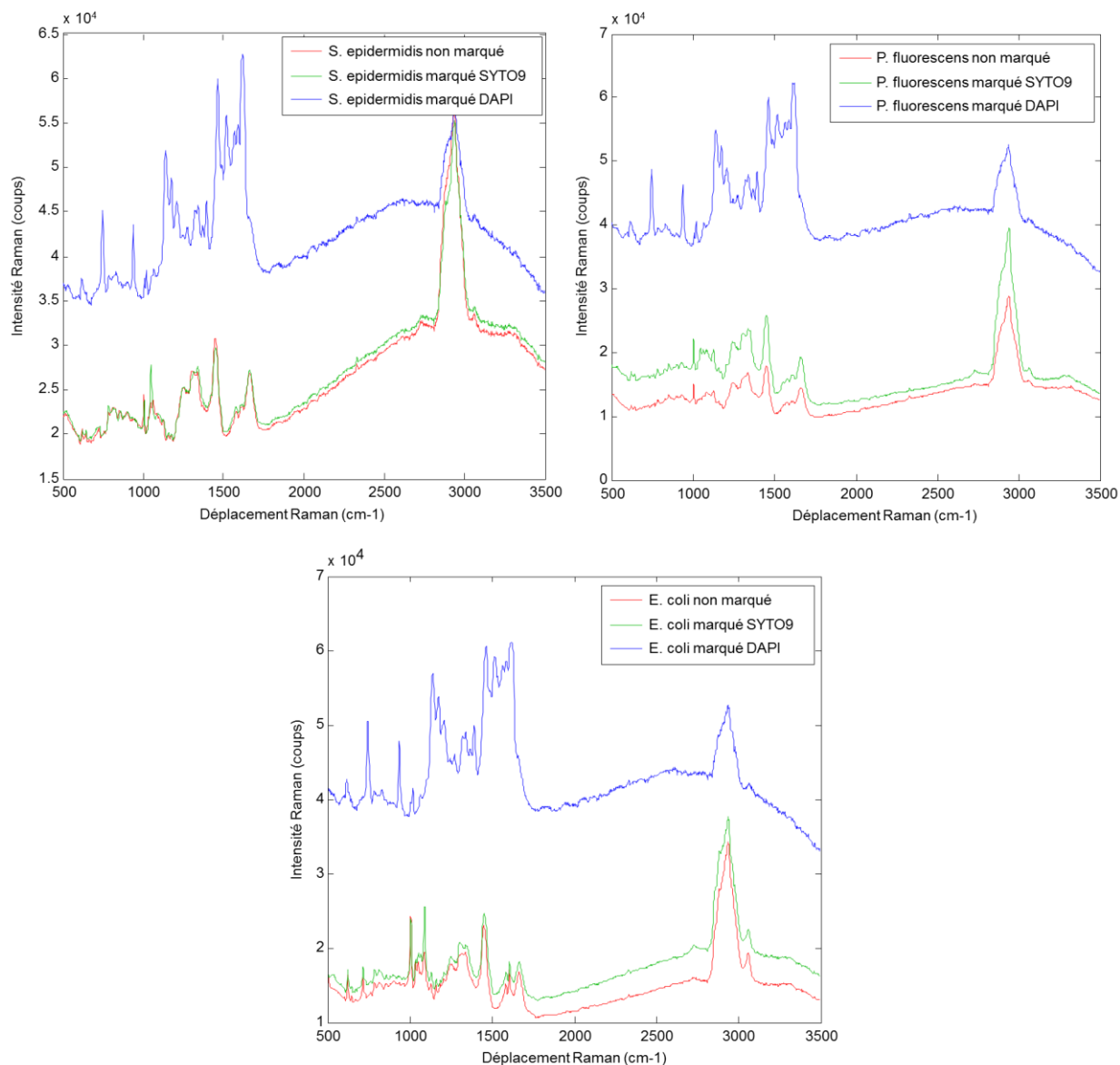


Figure 36 - Images des cellules de *S.epidermidis* marquée au DAPI et au SYTO9 sous lumière blanche a) et fluorescence des particules b).



**Figure 37 - Impact du marquage par le DAPI et le SYTO9 sur les spectres Raman.**

Entre les deux marquages, on remarque une différence d'intensité de fluorescence. Les bactéries marquées avec SYTO9 sont plus facilement détectables que les cellules marquées au DAPI. Ces observations ne sont néanmoins pas suffisantes pour faire un choix entre les deux fluorophores. Les échantillons sont ensuite transférés sur le microspectromètre Raman pour évaluer l'impact que le marquage entraîne sur la qualité des spectres Raman acquis à 633 nm.

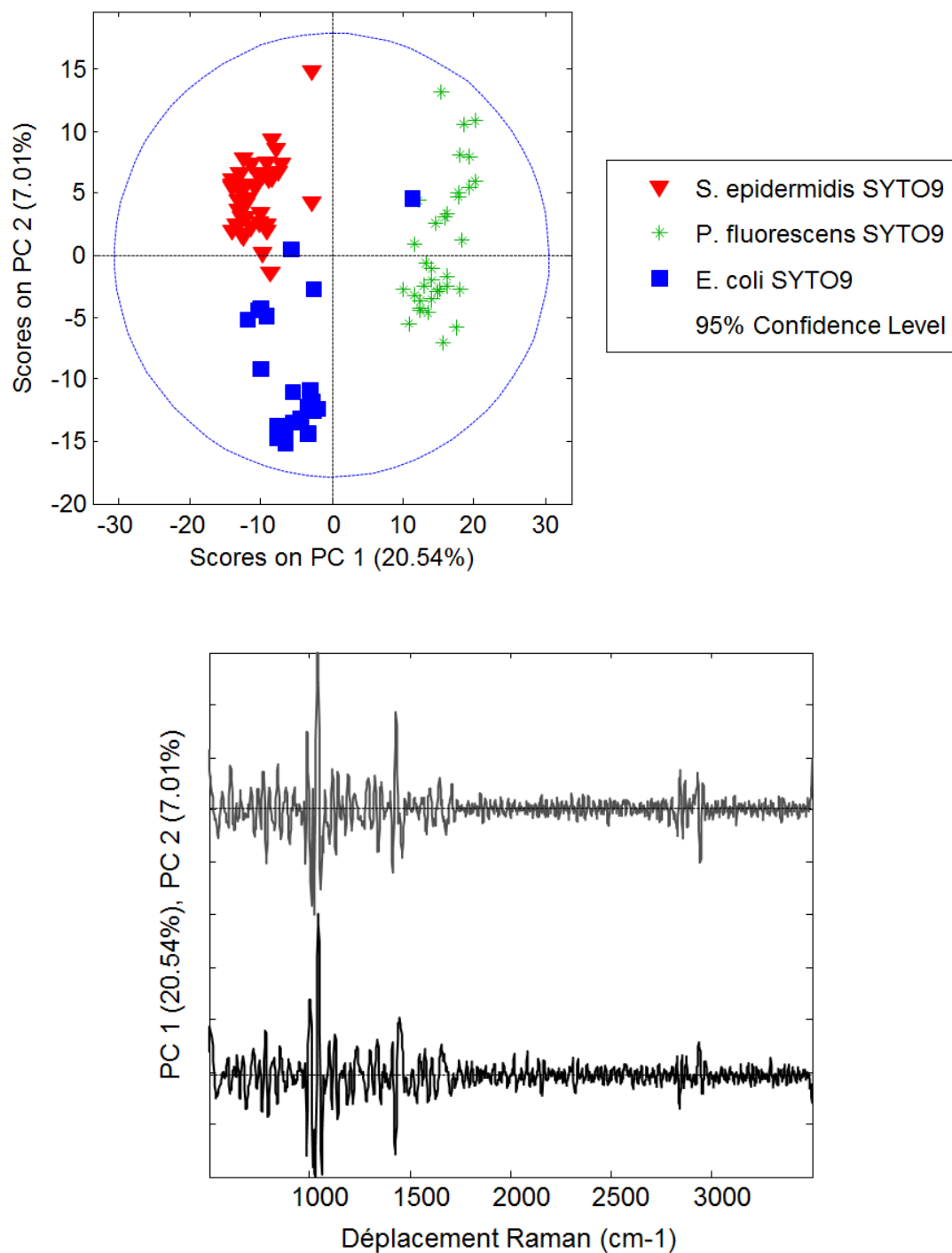
Sur la Figure 37, sont représentés les spectres moyens acquis à 633 nm sur les trois souches non marquées, marquées au DAPI et marquées au SYTO9. Nous observons une

différence flagrante entre les spectres des espèces marquées au DAPI et les deux autres. Le spectre bleu pour chacune des espèces est dominé par d'intenses bandes dues à la présence de DAPI. Ce fluorophore lorsqu'il n'est pas couplé ou déposé sur une surface ne présente pas de bande caractéristique, en revanche dans le cas contraire, un nombre important de bandes sont observables dans la *fingerprint* [64]. Ces bandes masquent le signal Raman provenant des bactéries, ce qui rend ce marqueur inadapté à l'analyse des bactéries par spectroscopie Raman. À l'inverse, il n'y a pas de contributions majeures de la part du SYTO9 qui masquent le signal Raman des cellules marquées. Comme nous pouvons le voir sur la Figure 38, il est toujours possible de discriminer les espèces entre elles même lorsqu'elles sont marquées par le SYTO9.

Suite à ces résultats, nous avons décidé d'utiliser le SYTO9 comme marqueur fluorescent pour repérer les bactéries sur la lame. Cependant, même si les bactéries sont observables sur le microscope d'épifluorescence, il faut être capable de réaliser les mêmes observations sur le système de microspectroscopie Raman qui n'est pas optimisé pour ces observations.

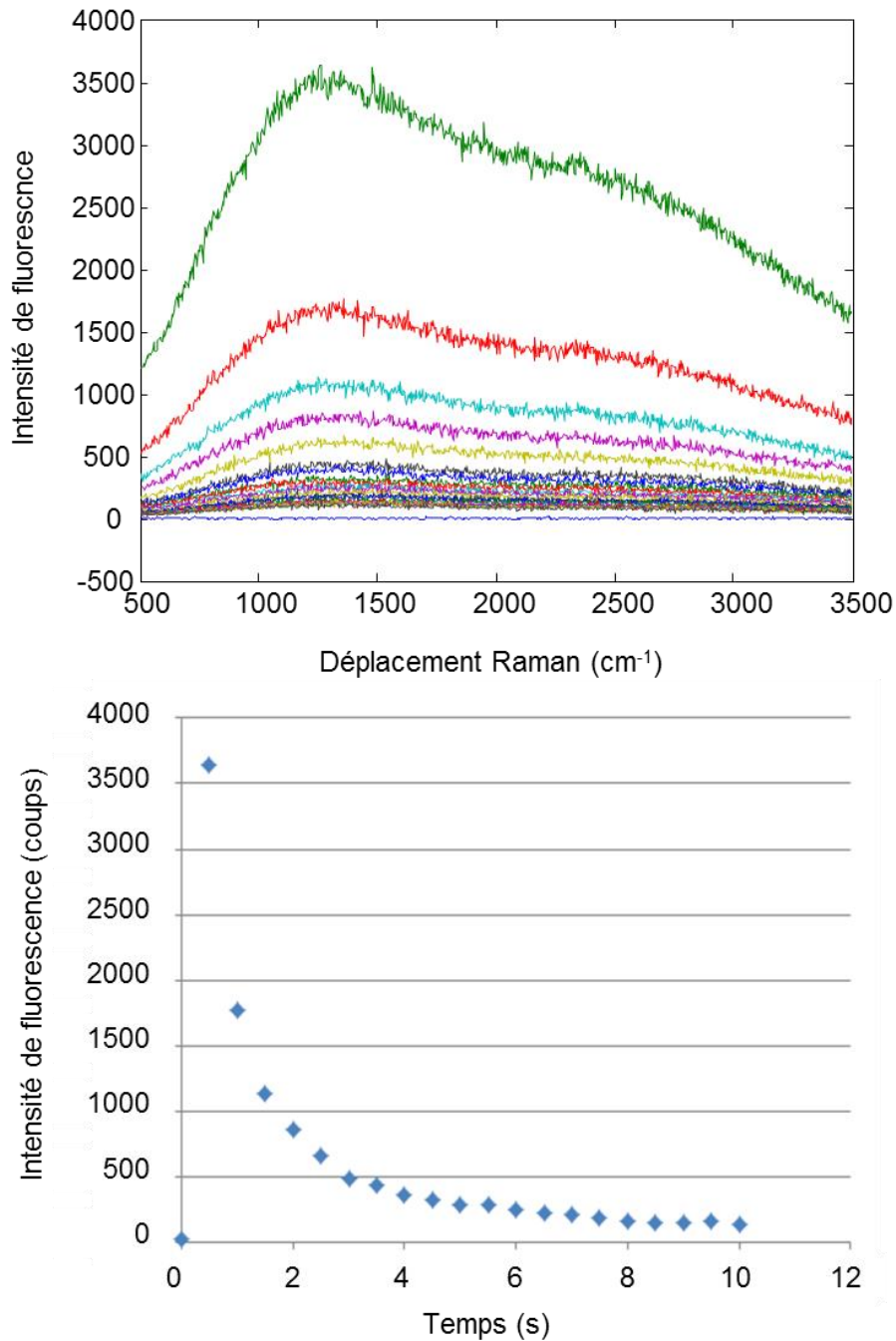
Nous avons tenté d'adapter différentes sources lumineuses et jeux de filtres pour exciter les cellules, mais les sources ne sont pas assez puissantes pour observer la fluorescence via la caméra du microspectromètre. Cependant, grâce à l'utilisation du laser bleu à 473 nm, il est possible en excitant pendant un temps court et à faible puissance de mesurer un spectre de fluorescence de la cellule marquée. L'opération doit être réalisée rapidement à cause du phénomène de *photobleaching* du marqueur [65]. La décroissance de la fluorescence du SYTO9 est rapide (Figure 39). Après 1s d'exposition, l'intensité de fluorescence enregistrée n'est plus qu'à 50% de sa valeur maximale. Dans notre étude, l'acquisition d'un spectre avec un temps d'accumulation 0,1s à une puissance de 0,3 mW (1% de la puissance du laser) est suffisante pour observer une fluorescence des agents marqués au SYTO9.





**Figure 38 - Graphique des *scores* et des *loadings* pour une ACP faite sur les spectres de trois espèces bactériennes marquées au SYTO9.**

Un maximum d'émission est mesuré à 506 nm pour le SYTO9. Il n'est cependant pas possible de déterminer un seuil de fluorescence à cause du phénomène de *photobleaching*. L'intensité mesurée possède une trop grande variabilité. De plus, la taille de la particule aura



**Figure 39 - Évolution de la fluorescence d'une bactérie marquée au SYTO9 sous un laser à 473 nm.**

une influence directe sur l'intensité de fluorescence. Plus la particule est grande, plus la quantité de SYTO9 susceptible de se greffer est importante. Il faut également noter que le SYTO9 ne marque pas exclusivement les cellules bactériennes mais qu'il est également capable de se greffer à tout autre matériel contenant de l'ADN (débris végétaux,

protozoaires...). C'est pourquoi ce critère ne se suffit pas à lui seul pour discriminer les bactéries des autres particules.

### 3.2.2. Discrimination morphologique

Nous avons vu précédemment qu'un critère morphologique n'était pas suffisant pour repérer les bactéries parmi l'ensemble des particules présentes sur un échantillon de bruit de fond d'air. Cependant, il est tout de même possible de mesurer un diamètre moyen des cellules auxquelles nous faisons face. Pour cela nous utilisons le logiciel ParticleFinder (Labspec, HORIBA Jobin Yvon) (Figure 40). Ce module du logiciel d'acquisition Raman utilise l'image vidéo obtenue par le microscope optique du LabRam pour déterminer des paramètres morphologiques des cellules comme le diamètre moyen, la circularité ou encore

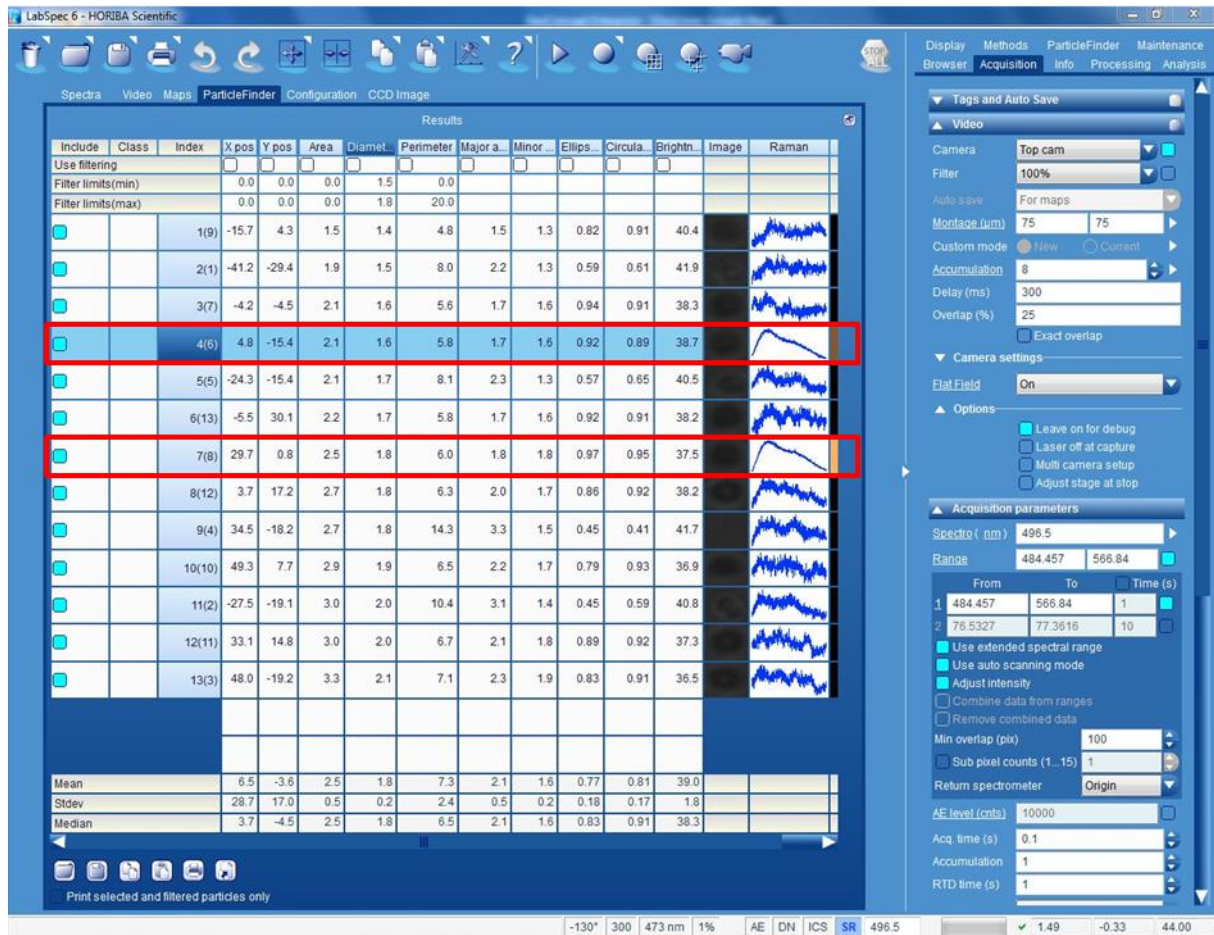
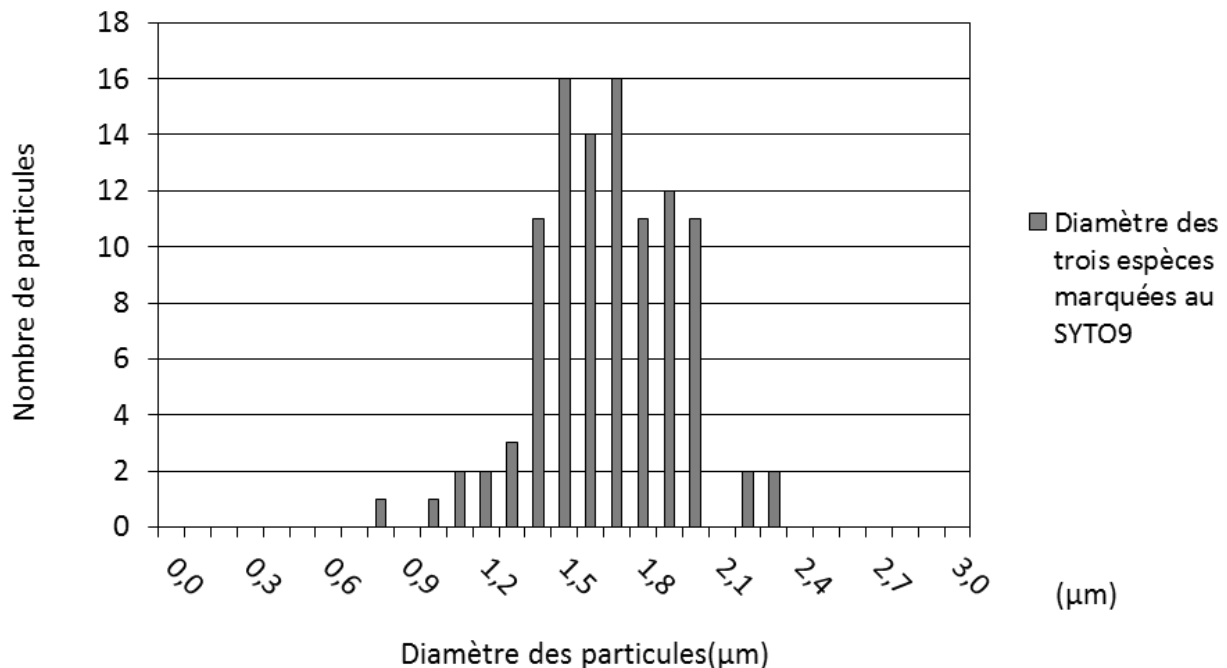


Figure 40 – Interface de l'outil *ParticleFinder* permettant d'obtenir des informations sur les paramètres morphologiques des cellules analysées.

le périmètre. Nous pouvons remarquer que certaines cellules présentent une fluorescence, comme les deux cellules dans les cadres rouges et d'autres, même si elles répondent au critère de taille, ne présentent aucune fluorescence mais seulement du bruit.

Après avoir étudié la variabilité de chacun des paramètres, il est apparu que le seul paramètre pertinent est le diamètre moyen des cellules. Il permet d'écarter suffisamment de particules pour n'obtenir qu'un nombre limité d'individus à analyser par spectroscopie Raman. Avec le *ParticleFinder*, nous avons pu déterminer un intervalle de tailles à partir d'une centaine de cellules analysées. Pour les trois espèces analysées (*S. epidermidis*, *P. fluorescens* et *E. coli*), près de 90% des cellules présentent un diamètre moyen compris entre 1,4  $\mu\text{m}$  et 2,0  $\mu\text{m}$  et plus de 50% des cellules ont un diamètre compris entre 1,5  $\mu\text{m}$  et 1,7  $\mu\text{m}$ .

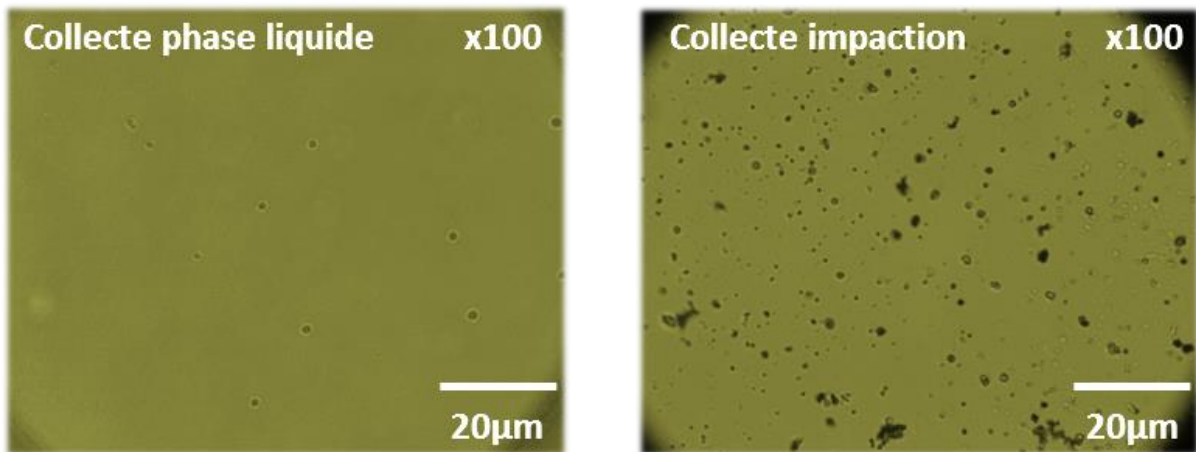


**Figure 41 - Distribution des bactéries selon le diamètre moyen déterminé à l'aide du *ParticleFinder*.**

Ainsi, lorsqu'il s'agira de repérer des cellules microbiennes parmi toutes les particules minérales dans un échantillon de bruits de fond d'air, nous veillerons à vérifier que la taille de la particule analysée respecte bien un critère de taille, à savoir un diamètre moyen compris entre 1,4  $\mu\text{m}$  et 2,0  $\mu\text{m}$ .

### 3.2.3. Concentration en bactéries et diminution de la densité particulaire

Nous avons vu que la densité particulaire est un problème majeur auquel nous sommes confrontés. Même s'il est possible de repérer une bactérie à l'aide de la fluorescence, du spectre Raman et du critère de taille, il faut pouvoir l'extraire seule via le système de microdissection laser. Pour remédier à ce problème, les particules collectées ne seront plus impactées directement sur la lame de CaF<sub>2</sub> depuis les différents étages de l'impacteur Andersen. Désormais la biocollecte des échantillons de bruit de fond d'air se fera en phase liquide à l'aide du *BioSampler* (SKC). Les échantillons sont récoltés pendant environ 60 min, à raison de 3 m<sup>3</sup> d'air pompés par heure, à travers un filtre ne laissant passer que les particules inférieures à 5µm.

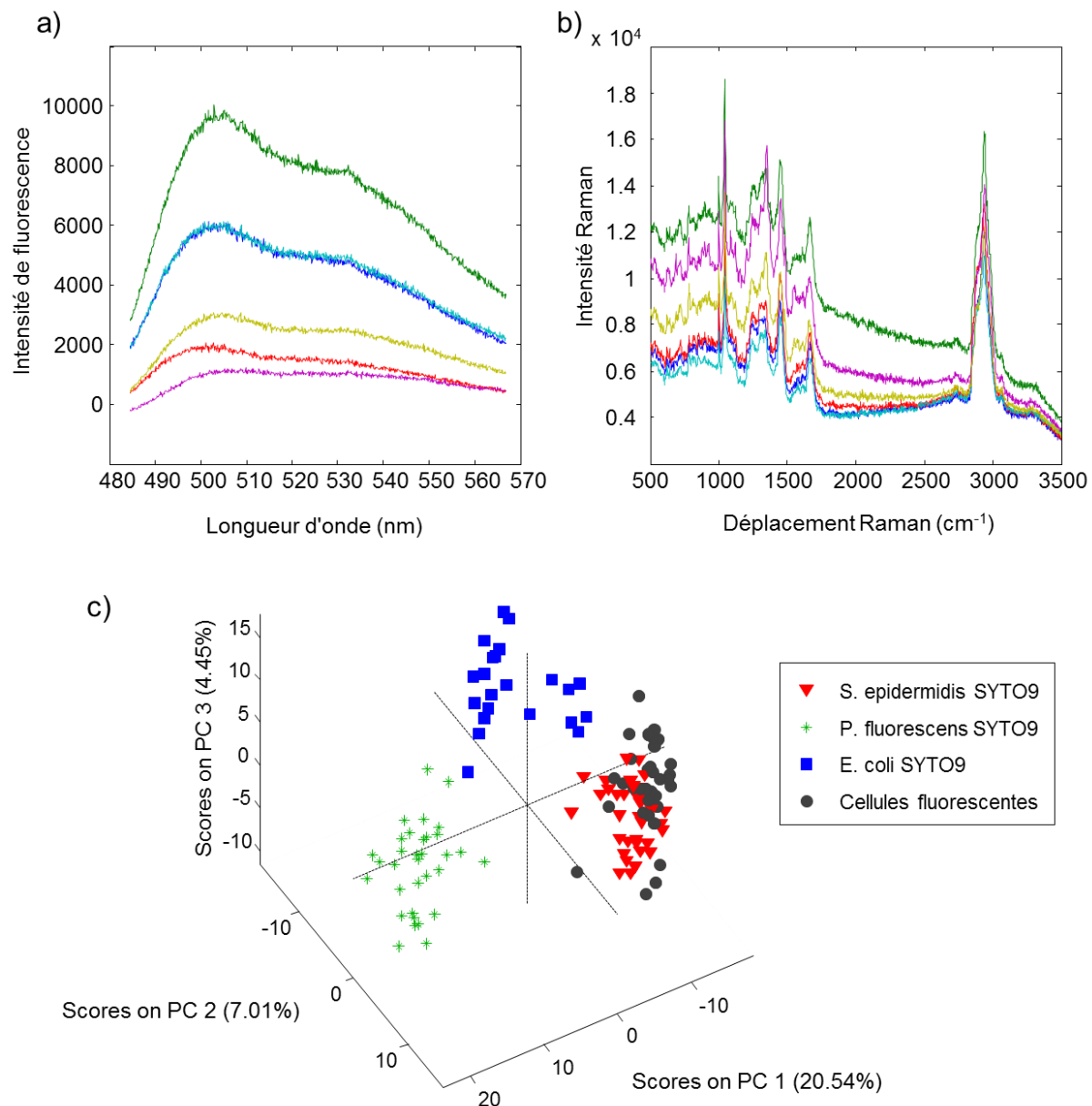


**Figure 42 - À gauche : une image d'un échantillon obtenu après une collecte en phase liquide; à droite : une image d'un échantillon obtenu après une collection par impaction.**

Les échantillons sont ensuite dopés en bactéries en incorporant directement dans la solution contenant les particules récoltées, des cellules de *S. epidermidis*. Puis un marquage au SYTO9 est réalisé afin d'assurer le repérage d'agents bactériologiques parmi les particules minérales. Un dépôt d'une gouttelette de 10 µL est effectué sur la lame de CaF<sub>2</sub>. Une concentration moyenne de  $3,5 \cdot 10^{-3}$  cellules/µm<sup>2</sup> est déterminée par comptage. Les observations faites sur les lames révèlent en effet une densité particulaire très fortement diminuée en comparaison aux prélèvements faits directement avec l'impacteur en cascade

Andersen. Les particules présentes sur la lames sont suffisamment espacées les unes des autres pour permettre une collecte par LCM.

L'analyse de la fluorescence des particules récoltées en phase liquide révèle l'existence de particules non-fluorescentes et de particules fluorescentes avec un maximum



**Figure 43 – a) spectres de fluorescence, b) spectres Raman de particules présentant une fluorescence et répondant au critère de taille et c) projection des spectres Raman des cellules fluorescentes sur un modèle ACP à trois composantes principales.**

d'intensité à 506 nm. Les particules examinées sont homogènes en taille et présentent un diamètre moyen compris entre 1,4  $\mu\text{m}$  pour les plus petites particules et 3,6  $\mu\text{m}$  pour les plus grosses. Les spectres Raman enregistrés présentent une signature similaire à celle d'une bactérie.

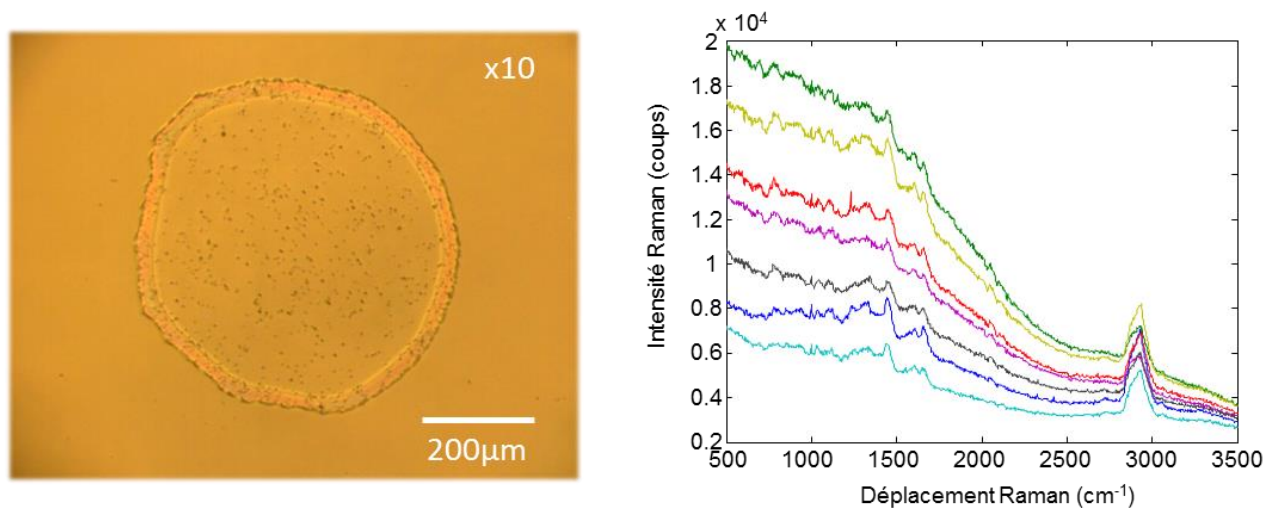
Les spectres Raman acquis sur les cellules fluorescentes sont ensuite projetés sur un modèle ACP à trois composantes principales (Figure 43 c)). Ce modèle ACP a été construit à l'aide des spectres des trois espèces *S. epidermidis*, *P. fluorescens* et *E. coli* marqués au SYTO9 (vus dans la partie 3.2.1 Marquage par agent fluorescent). La projection des spectres des cellules fluorescentes, qui sont en théorie des cellules de *S. epidermidis*, est tout à fait pertinente car ces individus sont bien localisés dans l'espace ACP. Cette partie démontre que l'utilisation simultanée de la fluorescence et de la discrimination en taille permet de faire un criblage efficace des particules.

#### **3.2.4. Dépôt des particules sur $\text{CaF}_2$**

Le dépôt des particules sur les supports en  $\text{CaF}_2$  a présenté des difficultés lors des analyses. En effet, la façon de déposer un échantillon aura une influence sur la qualité de l'échantillon et sur l'étape de prélèvement des particules par microdissection laser. Deux protocoles de dépôts ont été étudiés au cours de ces travaux, à savoir un dépôt par goutte ou un dépôt par sédimentation.

Lorsque le dépôt par goutte est effectué directement à partir des souches de laboratoire, une pellicule se dépose sur l'ensemble des particules et les cellules sont agglomérées sur différentes zones comme sur l'image de la Figure 44. L'analyse Raman effectuée sur ces cellules a mené à l'obtention de spectres de faible intensité, bruités et présentant un fond de fluorescence important.

De plus, les particules sont très proches les unes des autres et l'adhérence au support est augmentée lorsque le dépôt par goutte est effectué. Il devient alors impossible de prélever les cellules par microdissection laser.



**Figure 44 - Effet du dépôt par goutte sur la répartition des cellules sur la lame. Influence sur les spectres Raman de bactéries *S. epidermidis*.**

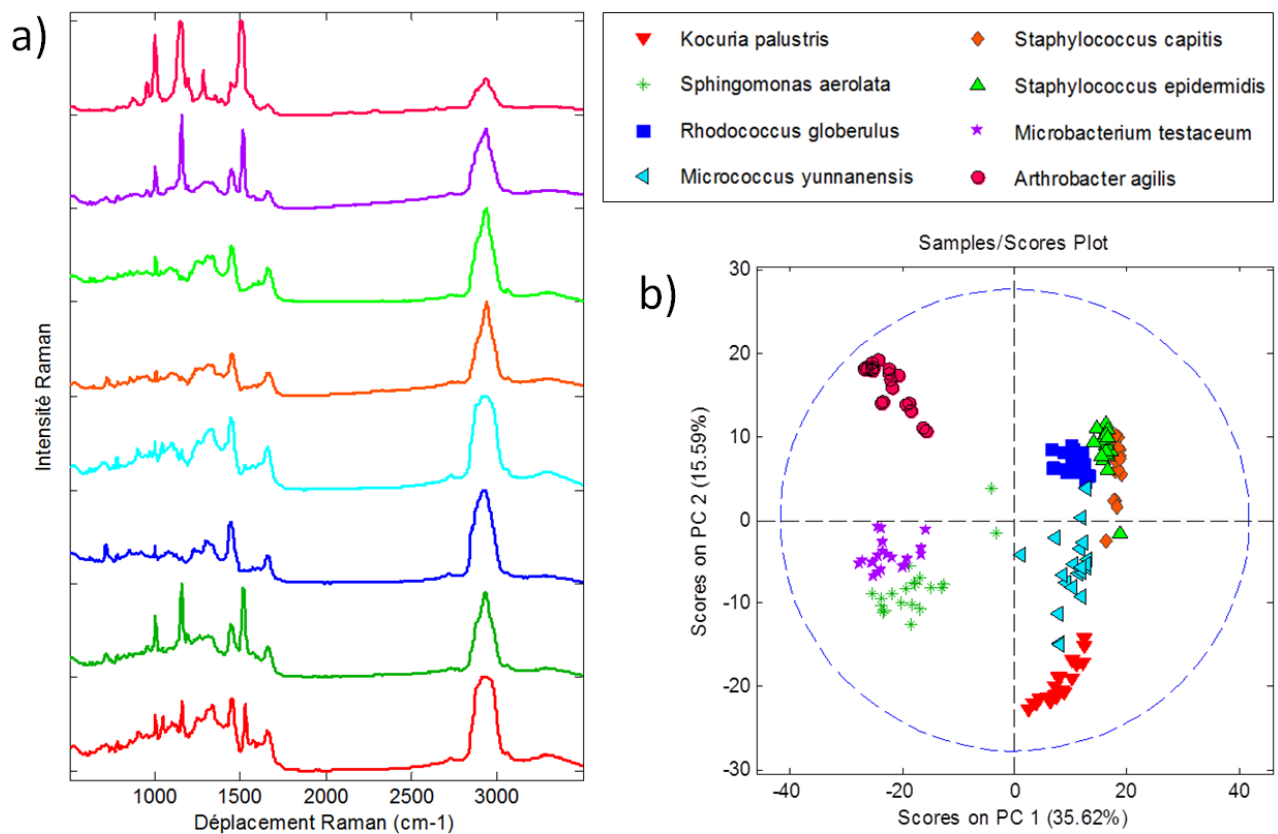
Pour remédier à ces problèmes, nous avons opté pour un dépôt par sédimentation. Les bactéries sont cultivées et aérosolisées pour obtenir un dépôt plus uniforme sur la totalité de la lame avec des cellules bien espacées les unes des autres. Ainsi l'analyse Raman n'est plus perturbée et la microdissection laser des cellules devient possible.

### **3.3. Étude de bactéries provenant de souches environnementales**

Une campagne de prélèvement a été réalisée par le CSTB (Centre Scientifique et Technique du Bâtiment) en vue de récolter plusieurs souches bactériennes représentatives du milieu extérieur et intérieur. Les prélèvements ont permis d'isoler 8 espèces différentes qui seront mises en culture en vue de réaliser un dépôt par sédimentation sur des lames de  $\text{CaF}_2$ . En fin d'incubation, des prélèvements pour identification par analyse génomique de bactéries, depuis le milieu de culture, ont été réalisés pour identifier les espèces isolées et cultivées. L'analyse a révélé la présence de 8 espèces différentes à savoir : *Staphylococcus epidermidis*, *Microbacterium testaceum*, *Arthrobacter agilis*, *Kocuria palustris*, *Sphingomonas aerolata*, *Rhodococcus globerulus*, *Micrococcus yunnanensis*, *Staphylococcus capitis*. Chacune des espèces est aérosolisée et déposée sur lame de  $\text{CaF}_2$  par sédimentation puis analysée par microspectroscopie Raman.



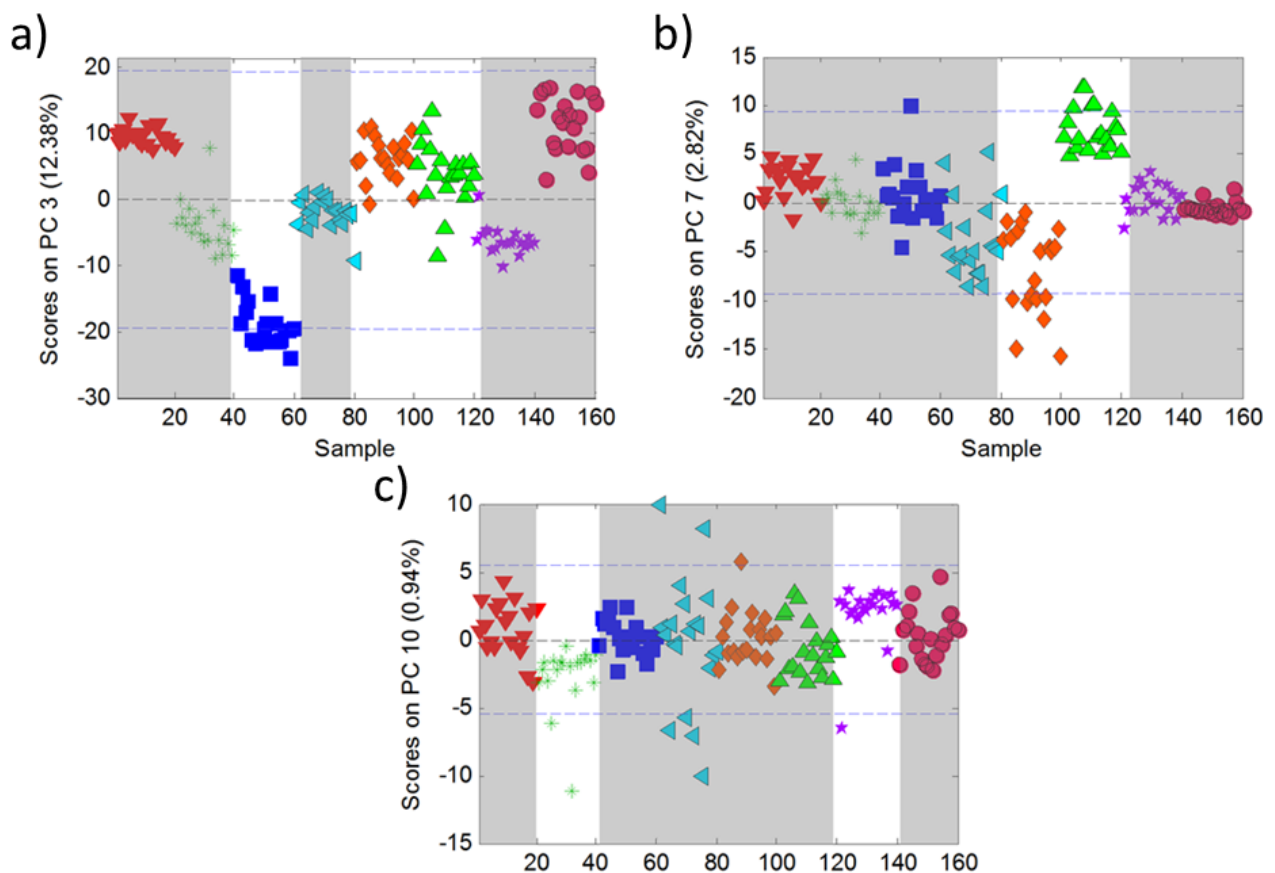
Pour chacune des huit souches environnementales, 20 spectres ont été acquis dans des conditions identiques et les spectres moyens des différentes espèces sont présentés dans la Figure 45 a). Au premier abord, les spectres présentent quelques différences, ce qui laisse présager une bonne séparation de certaines des souches. Les prétraitements appliqués sur les spectres sont une dérivée seconde (ordre 2, 15 points), une normalisation SNV et un centrage par la moyenne. Une ACP est ensuite réalisée sur ces spectres prétraités. Sur le graphique des scores (Figure 45 b)) on observe une discrimination entre les différentes espèces uniquement à l'aide des deux premières composantes. Notons également que certaines espèces sont très proches les unes des autres dans cette représentation, mais l'analyse des représentations sur d'autres composantes que PC1 et PC2 permet une séparation.



**Figure 45 - Spectres Raman des huit espèces isolées d'un prélèvement d'air extérieur.**

En poursuivant l'exploration des données sur les composantes suivantes, on s'aperçoit que des espèces proches sur les deux premières composantes à savoir les spectres de la bactérie *Rhodococcus globerulus* (carré bleu) et les deux espèces de *Staphylococcus* (losange orange et triangle vert clair) peuvent être discriminées par PC3 expliquant 12,38 % de la

variabilité totale (Figure 46 a)). Les deux espèces *Staphylococcus epidermidis* et *Staphylococcus capitis*, ne seront séparables qu'à partir de la PC7 représentant 2,82 % de la variance totale (Figure 46 b)). Un autre couple difficilement séparable est l'espèce *Sphingomonas aerolata* et *Microbacterium testaceum*. Ces deux espèces seront correctement séparées par PC10 représentant 0,94% de la variance totale (Figure 46 c)). C'est pourquoi pour les calculs de classification SVM, nous utiliserons les scores des individus sur les 10 premières composantes de l'analyse ACP.



**Figure 46 - Séparation des espèces proches sur les deux premières composantes. Scores de chaque individu sur a) PC2, b) PC7 et c) PC10.**

L'utilisation de 10 composantes principales est nécessaire pour aboutir à une bonne séparation des différentes classes. L'ensemble des graphiques des scores de chaque individu sur chacune des composantes est détaillé dans la Figure 47.

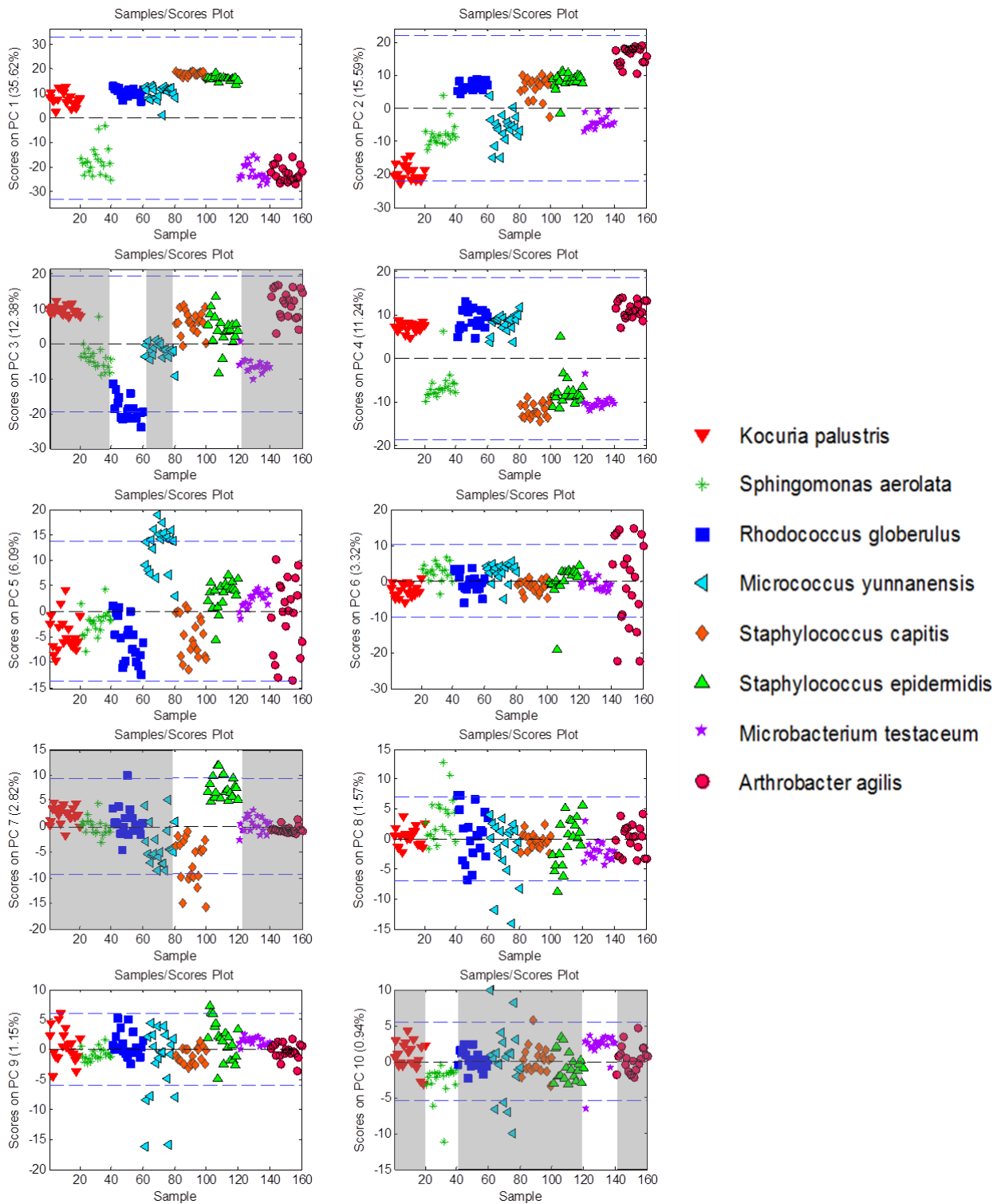
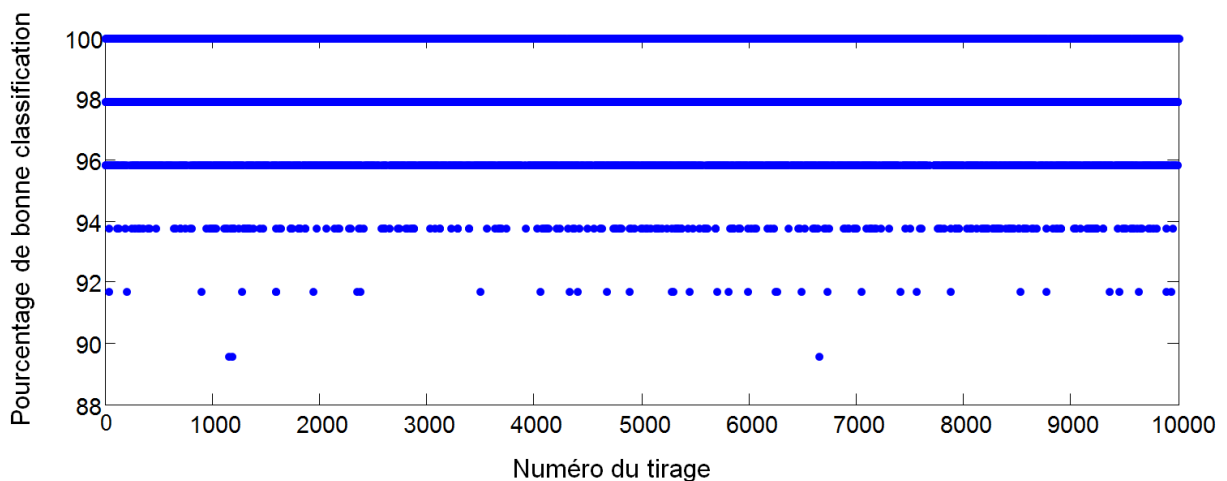


Figure 47 - Scores de chacun des spectres sur chacune des 10 premières composantes calculées.

Pour réaliser un modèle de classification nous conservons donc les scores de chacun des individus sur chacune des 10 premières composantes principales. Ceux-ci sont enregistrés et implémentés dans l'algorithme de construction du modèle de classification SVM.

Pour chacune des 8 classes, 20 spectres Raman ont été acquis. Parmi ces 20 spectres, 70%, soit 14 individus, sont sélectionnés, aléatoirement pour constituer le jeu de données servant à la construction du modèle et 30%, soit 6 individus serviront pour la validation du modèle. Ainsi pour les 8 espèces, nous auront un total de 112 spectres pour le jeu de calibration et 48 spectres pour le jeu de validation. Les valeurs des méta-paramètres sont fixées par défaut à  $C=10^5$  et  $G=10^{-5}$ . Pour obtenir une estimation globale des performances d'un modèle de classification SVM sur ce jeu de données, nous employons la technique du *bootstrap* qui consiste à faire un ré-échantillonnage du jeu de calibration en réincorporant à chaque itération les échantillons déjà utilisés. Au total ce sont 10000 modèles SVM qui ont été calculés et les résultats obtenus sont présentés dans la Figure 48.



**Figure 48 - Estimation globale des performances du modèle de classification SVM**

- 36,16% des tirages effectués ne présentent aucun individu mal classé sur 48 soit 100 % de bonne classification ;
- 45,05% des tirages effectués présentent 1 individu mal classé sur 48 soit 97,9% de bonne classification ;
- 15,68% des tirages effectués présentent 2 individus mal classés sur 48 soit 95,8% de bonne classification ;

- 2,72% des tirages effectués présentent 3 individus mal classés sur 48 soit 93,8% de bonne classification ;
- 0,36% des tirages effectués présentent 4 individus mal classés sur 48 soit 91,7% de bonne classification ;
- 0,03% des tirages effectués présentent 5 individus mal classés sur 48 soit 89,6% de bonne classification.

Ainsi l'estimation globale des performances d'un modèle de classification SVM sur le jeu de données est de 98,2% de bonne classification. Ce taux de classification est très satisfaisant.

## 4. Conclusion

Ce chapitre présente les points clés concernant l'optimisation de la mesure des spectres Raman, du calcul des modèles de classification SVM sur des échantillons issus de souches de laboratoire ou de souches environnementales.

L'optimisation de la mesure des spectres Raman est un point essentiel dans l'ensemble du protocole que nous développons. Toutes les études que nous avons menées ont permis d'optimiser l'ensemble des paramètres expérimentaux allant du choix du support à la détermination du temps d'acquisition en passant par le choix du laser et du réseau de diffraction afin de travailler dans les meilleures conditions d'analyse possible. Les spectres acquis présentent un bon rapport signal sur bruit permettant ainsi une bonne discrimination des espèces analysées.

Les résultats obtenus suite aux études faites en spectroscopie M-CARS et EPI-CARS sur les bactéries sont encourageants car un signal est observé mais il serait nécessaire d'entreprendre des recherches plus approfondies pour réussir à obtenir un signal d'une qualité suffisante pour être exploitable.

Nous avons pu observer la complexité des aérosols naturels et la difficulté à repérer les bactéries parmi l'ensemble des particules collectées. C'est pourquoi nous avons mis en place un protocole de collecte en phase liquide suivi d'une étape de marquage des agents microbiologiques par le fluorophore SYTO9. Les particules sont ensuite déposées sur lames de

CaF<sub>2</sub> par ré-aérosolisation et sédimentation des particules pour limiter l'adhésion au support d'analyse. La fluorescence des particules est observée en procédant à l'acquisition d'un spectre sur le système Raman à l'aide du laser bleu à 473 nm. En couplant ces observations au module *ParticleFinder*, permettant de déterminer un diamètre moyen des particules analysées, il devient possible de différencier les particules biologiques des particules minérales.

La méthode de classification SVM a démontré qu'elle est un outil puissant pour la reconnaissance des bactéries à condition d'utiliser les bons prétraitements mathématiques à appliquer pour corriger les spectres. Les résultats obtenus présentent des taux de bonne classification supérieurs à 98%.

# Chapitre 4 – Modules NanoGPS, microdissection laser et analyse génomique

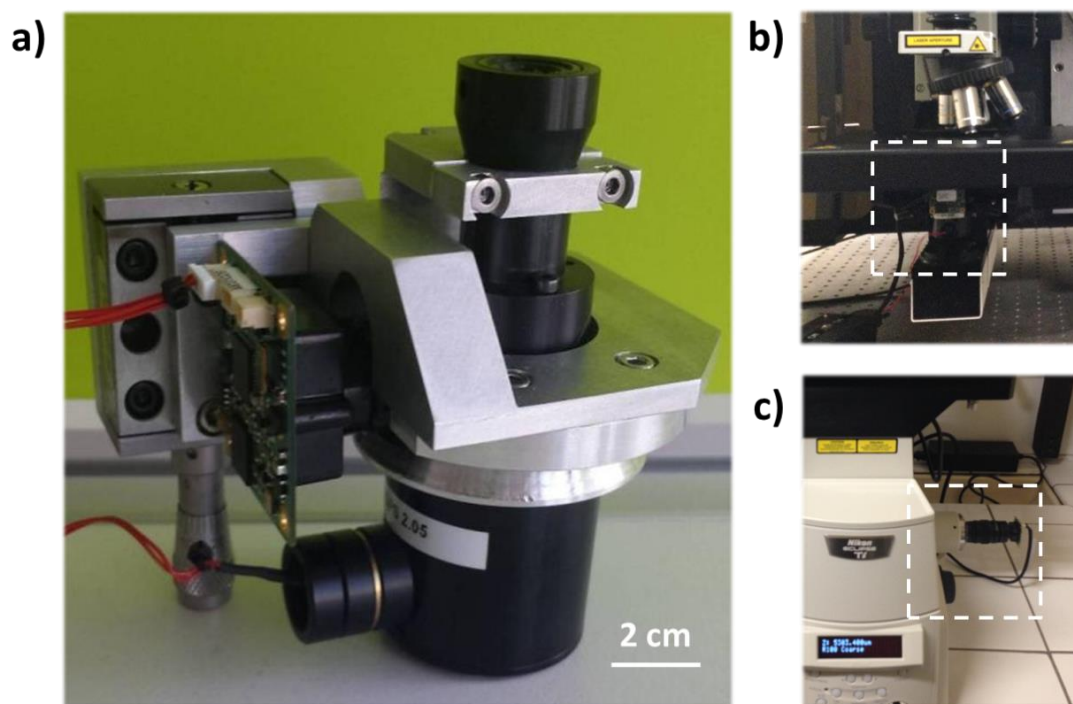
Dans ce chapitre il est question de détailler comment est réalisé le transfert des bactéries entre le microspectromètre Raman, où elles sont analysées pour acquérir leur signature chimique, et le système de microdissection laser pour leur prélèvement. Pour être sûr que nous procédons à l'extraction de la même particule qui a été analysée par microspectroscopie, nous avons effectué un couplage virtuel des deux appareils à l'aide de deux nanoGPS placés sur chacun d'eux. La précision des modules nanoGPS sera estimée pour permettre un retour sur particule unique. Nous aborderons ensuite en détail la mise en place de la microdissection laser sur cellule unique et les difficultés rencontrées lors de l'optimisation de cette étape. Enfin, dans une dernière partie, nous présenterons les difficultés et les premiers résultats obtenus par identification génomique sur cellule unique microdissectée.

## 1. Système de repositionnement absolu : le nanoGPS

Avec le développement des nanotechnologies, la capacité à effectuer des mesures extrêmement précises sur divers objets à l'échelle micrométrique et nanométrique est devenue primordiale, tant dans la recherche universitaire que dans l'industrie. Néanmoins, nous sommes confrontés à la difficulté de retrouver les objets analysés sur différents instruments (colocalisation) ou bien sur un même instrument (relocalisation) jours après jours. Pour répondre à ce besoin, la société HORIBA Jobin Yvon a développé une solution originale, le nanoGPS (Figure 49). Nous avons ainsi pu exploiter des prototypes dans le cadre de cette

thèse. Le nanoGPS a pour fonction principale de permettre un retour sur zone, en vue d'une nouvelle analyse spectroscopique Raman ou bien d'un prélèvement par microdissection laser pour une identification génomique. Pour rappel, l'objectif à terme, est d'associer la signature spectrale avec la signature génomique de la cellule pour construire un modèle de classification robuste, prenant en compte toutes modifications engendrées par l'environnement sur une cellule bactérienne.

Dans cette partie, nous expliquons le fonctionnement du module nanoGPS, les différences entre les deux nanoGPS équipés sur le microspectromètre Raman et sur l'outil de microdissection laser. Nous évaluerons la précision en repositionnement de l'instrument selon différents cas de figure, en vue d'un prélèvement optimal de cellules pour analyse génomique.



**Figure 49 – a) Image d'un nanoGPS, b) fixé sous la platine du microspectromètre Raman et c) implémenté sur l'instrument de microdissection laser.**

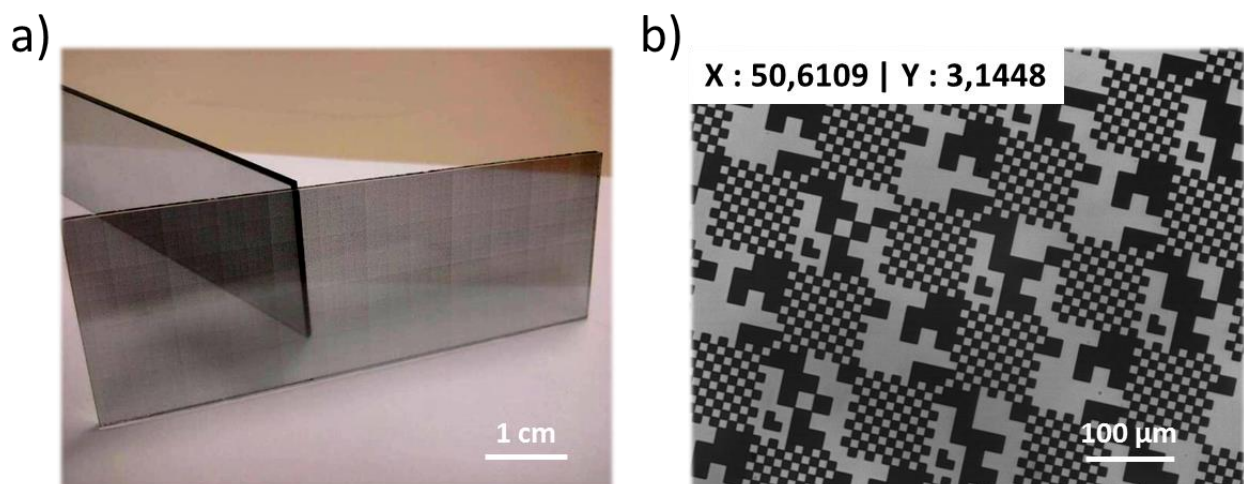
### **1.1. Principe général**

La technologie développée lors de la construction du nanoGPS permet un repositionnement absolu, précis et fiable sur une ou plusieurs zones analysées. Toutes les observations faites sur les échantillons sont enregistrées sur un logiciel de cartographie et de



navigation, entièrement adapté pour le système nanoGPS. Chaque observation enregistrée sur la cartographie peut être annexée d'une image de la zone observée au microscope et d'un spectre Raman. Le microscope du microspectromètre Raman est équipé de plusieurs objectifs à différents grossissements (x10, x50 et x100), tous pris en charge par le module nanoGPS et le logiciel de navigation, rendant possible la réalisation d'une cartographie à différentes échelles en superposant les images aux différents grossissements pour une meilleure navigation sur l'échantillon.

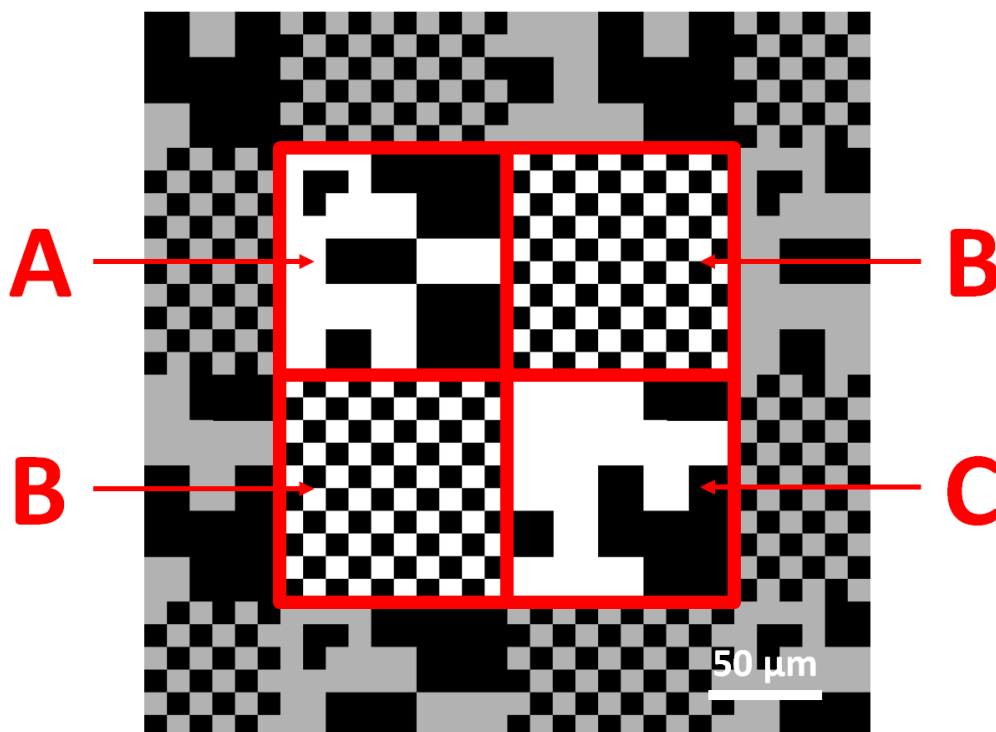
Pour réaliser une cartographie, il est nécessaire d'utiliser des lames spécialement conçues pour le nanoGPS (Figure 50 a)) sur lesquelles des motifs sont imprimés par lithographie permettant un repositionnement absolu sur n'importe quel échantillon. Ce sont sur ces lames que nous fixons nos lames de CaF<sub>2</sub> sur lesquelles sont déposées les bactéries. Ces motifs (Figure 50 b)), assimilables à des « QR codes », recouvrent l'ensemble de la lame et représentent l'emplacement observé sur la lame. Pour lire les motifs, une caméra est reliée à un petit microscope optique de faible ouverture numérique (0,13). Ce sont des motifs pseudopériodiques qui, à l'aide d'un algorithme d'analyse d'images, sont traduits en coordonnées pour créer et compléter la cartographie de l'échantillon.



**Figure 50 - Images a) des lames pour la localisation nanoGPS et b) des motifs imprimés par lithographie.**

Pour comprendre comment le logiciel d'analyse détermine la position selon l'image qui lui est transmis par le nanoGPS, il est nécessaire de comprendre la signification des différents motifs sur la lame. Lorsqu'une image est transférée au logiciel, celui-ci va se

focaliser sur l'analyse de quatre motifs. Sur la Figure 51, sont détaillés les différents motifs présents sur la lame nanoGPS. Le premier motif, noté A dans la Figure 51, est le plus important. En effet, celui-ci traduit les coordonnées X et Y sur lesquelles le nanoGPS est centré. On remarque qu'à l'intérieur de ce motif il y a toujours une marque en forme d'équerre, dans la Figure 51, qui correspond à une marque d'orientation de la lame. Cette marque est nécessaire pour prendre en compte le décalage angulaire lorsque la lame n'est pas correctement placée sur la platine. Un motif sur deux sur l'ensemble de la lame, est une grille de localisation (sous forme de damier) notée B dans la Figure 51. Le dernier motif, reproduit à l'identique sur l'ensemble de la lame (noté C), permet de l'identifier.

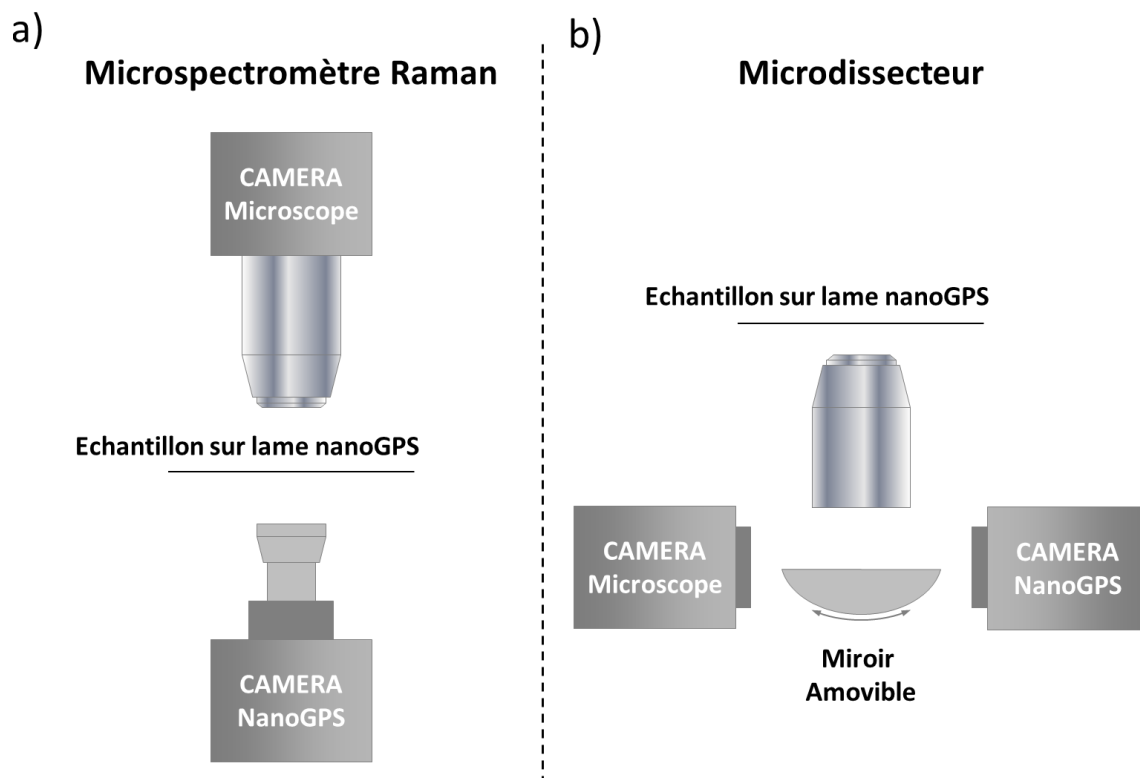


**Figure 51 - Représentation des différents motifs imprimés sur les lames nanoGPS. A) Coordonnées X et Y et marque d'orientation, B) grille de localisation, C) identifiant de la lame.**

Ainsi, grâce à ce module nanoGPS, toutes les observations effectuées au microscope peuvent être enregistrées. L'ensemble des cellules analysées sont donc localisées et répertoriées sur une cartographie globale de la surface de l'échantillon à l'aide des observations faites au microscope.

## 1.2. Couplage virtuel de différents instruments

Au cours de nos analyses, nous avons implémenté deux modules nanoGPS sur les deux principaux instruments mis en jeu lors de l'élaboration de ce nouveau protocole que nous développons. Un premier module est implémenté sur le système de mesure Raman, grâce auquel nous pouvons effectuer l'enregistrement de la localisation des particules sur l'échantillon. Le second module est installé sur le microdissecteur. Avec celui-ci nous pouvons revenir sur les bactéries pour réaliser une extraction par microdissection laser dans le but d'effectuer plus tard une analyse génomique. Toutefois, il existe une différence au niveau optique dans le mode de fonctionnement entre les deux modules nanoGPS. Sur la Figure 52 sont schématisés les deux dispositifs nanoGPS en question.



**Figure 52 - Positionnement des nanoGPS sur le microspectromètre Raman et sur le microdissecteur.**

Le premier module nanoGPS (Figure 52 a)), adapté à l'instrument Raman est totalement indépendant du microscope qui nous permet d'enregistrer simultanément les observations des motifs et de réaliser les analyses Raman des particules. Celui-ci est

positionné sous la platine du microscope du microspectromètre Raman, avec une vision directe sur la face inférieure de la lame sur laquelle sont imprimés les motifs permettant d'attribuer les coordonnées. Avec cette disposition, il est donc possible de travailler simultanément avec le nanoGPS et le microspectromètre Raman. Ce premier nanoGPS permet d'enregistrer toutes les observations nécessaires à la construction de la cartographie. L'autre avantage de ce nanoGPS est qu'il est libre vis-à-vis du microspectromètre, et peut donc être désinstallé lorsque nous ne l'utilisons pas. Le second nanoGPS est fixé au microdissecteur (Figure 52 b)). Pour des raisons instrumentales et principalement d'encombrement, l'adaptation de ce nanoGPS n'a pas pu se faire de la même façon que le premier. En effet, comme la microdissection est réalisée sur la face supérieure de l'échantillon, la tourelle d'objectifs est obligatoirement située sous la platine mobile où est déposée la lame. Ainsi, il n'est donc pas possible de placer le nanoGPS sous l'échantillon. Pour remédier à cette difficulté, le microdissecteur est équipé d'un miroir amovible permettant de basculer le trajet optique soit vers la caméra du microscope du microdissecteur soit vers le nanoGPS. Une étape de refocalisation des objectifs sur la face inférieure ou supérieure est nécessaire en fonction de la caméra que nous utilisons et entre chacun des prélèvements effectués.

La cartographie ainsi obtenue peut être constamment enrichie à l'aide des divers systèmes nanoGPS. Il est tout à fait possible de réaliser une multitude d'analyses, à l'aide de plusieurs instruments, à condition que chaque instrument soit équipé d'un nanoGPS. Néanmoins, la tâche ne peut pas encore être automatisée à cause de la faible taille des cellules bactériennes. La principale difficulté rencontrée lors des manipulations réside dans le repositionnement des objectifs. En effet, si l'objectif est légèrement décalé d'une observation à une autre, la superposition des images ne se fera pas correctement et la position sera suffisamment erronée pour empêcher une automatisation de la manipulation. Néanmoins, le système présente une grande précision que nous allons désormais évaluer.

### **1.3. Précision du système nanoGPS**

Le système nanoGPS propose une grande fiabilité de repositionnement lorsqu'il s'agit de superposer des images. Pour illustrer la précision du système, nous avons étudié les décalages qu'il peut y avoir lorsque deux observations sont enregistrées et superposées sur un même échantillon. Pour illustrer notre propos, nous nous basons sur les images acquises à

partir des lames sur lesquelles sont imprimés les motifs pseudopériodiques. Toutes les observations présentées sont réalisées à l'aide des objectifs x50 et x100 mais le nanoGPS et le logiciel de cartographie sont également configurés pour utiliser également un objectif x10. Dans cette partie différents cas de figure sont étudiés afin d'évaluer la précision du repositionnement du nanoGPS.

### 1.3.1. Cas n°1 : Superposition de deux images sans réétalonnage

Le premier cas est le cas le plus simple en matière de superposition d'images. Le nanoGPS est étalonné, une première observation est enregistrée et ajoutée à la cartographie, puis en déplaçant la platine de quelques dizaines de micromètres sans retirer la lame de la platine, une seconde observation est enregistrée puis également ajoutée à la

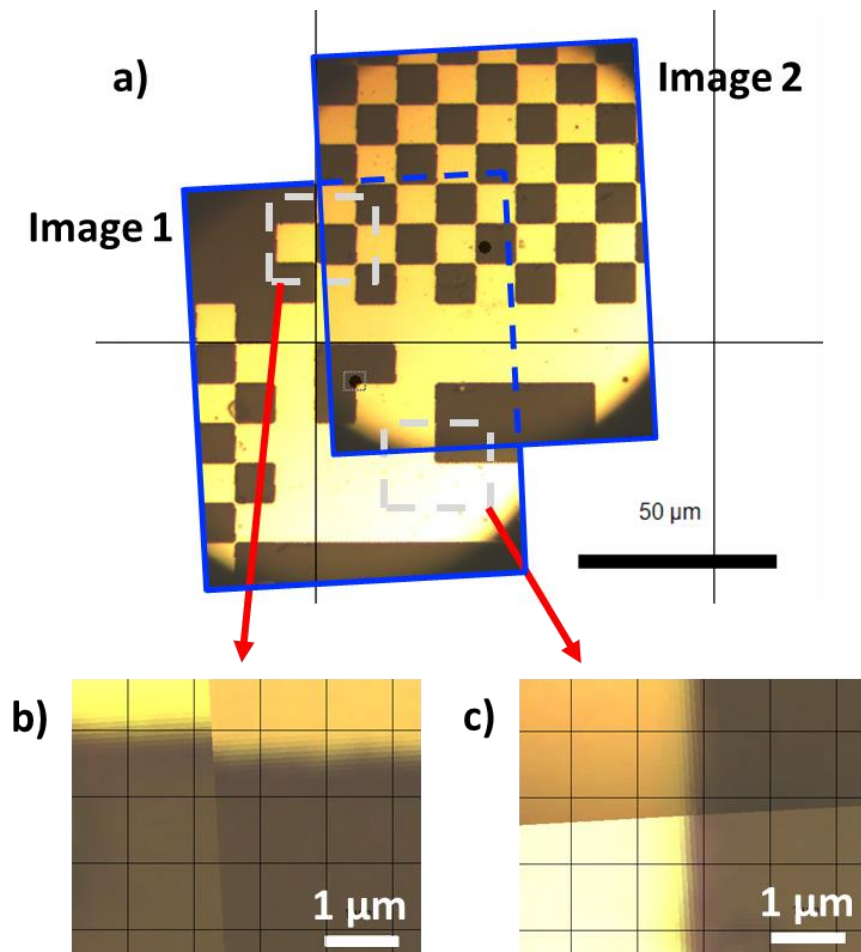


Figure 53 – Précision du nanoGPS. Premier cas : superposition de deux observations avec un même objectif.

cartographie. La superposition des deux images est illustrée dans la Figure 53 a). Les images b) et c) représentent respectivement le décalage vertical et horizontal causé par l'erreur de lecture du nanoGPS. L'écart vertical constaté est inférieur au micromètre et l'écart horizontal est négligeable. Cette étude démontre que l'erreur entre deux mesures n'est pas impactante pour la construction de la cartographie et la localisation d'une bactérie dont la taille avoisine le micromètre.

### 1.3.2. Cas n°2 : Utilisation de deux objectifs pour construire la cartographie

Ce deuxième cas illustre la construction d'une cartographie à l'aide de plusieurs objectifs. Le système nanoGPS a été étalonné, puis une première image est enregistrée à l'aide de

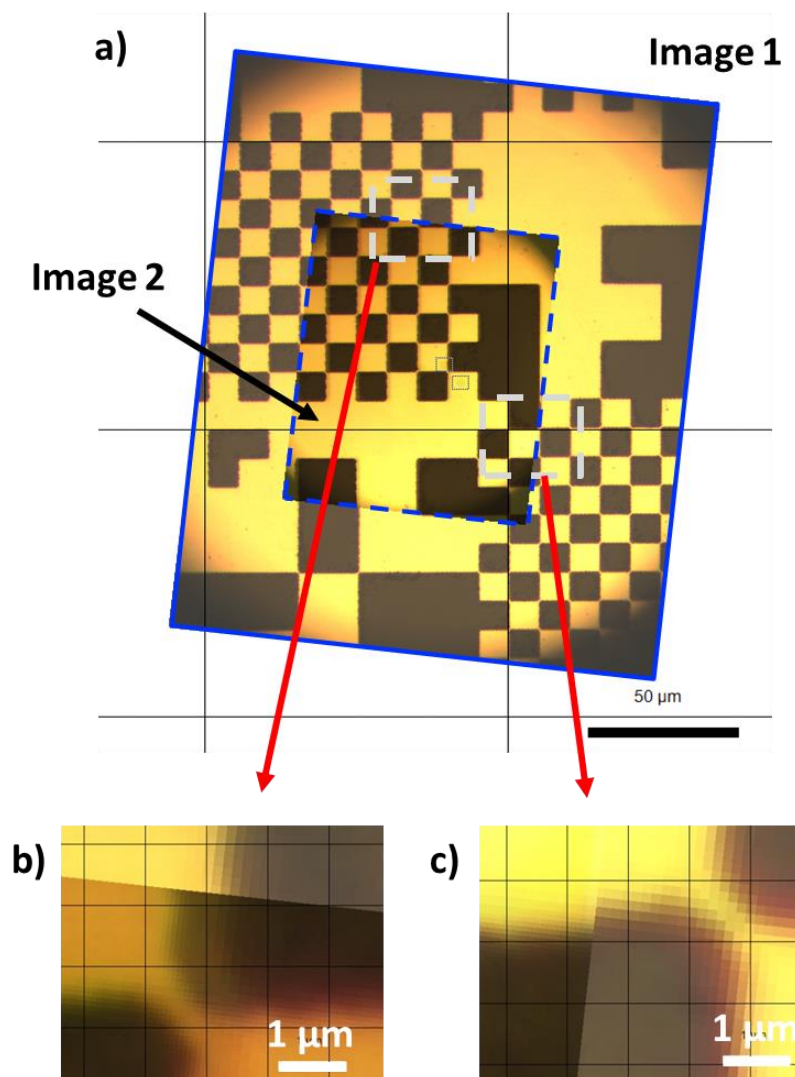


Figure 54 - Précision du nanoGPS. Deuxième cas : superposition de deux images acquises à l'aide de deux objectifs différents (x50 et x100) sans déplacement de la platine.

l'objectif x50. Sans déplacer la platine et sans retirer la lame, une seconde observation est réalisée mais cette fois-ci à l'aide de l'objectif x100. Les deux images sont superposées et la précision du décalage peut être estimée (Figure 54).

De prime abord, les deux images se superposent correctement. En observant de plus près les frontières de la superposition, l'écart horizontal et l'écart vertical constatés sont inférieurs au micromètre ce qui reste acceptable pour notre utilisation.

### 1.3.3. Cas n°3 : Superposition de deux images après rotation de la lame sans réétalonnage

Ce troisième cas illustre la superposition de deux images sans réétalonnage entre les deux observations. Avant l'acquisition de la seconde observation, la lame est retirée de la

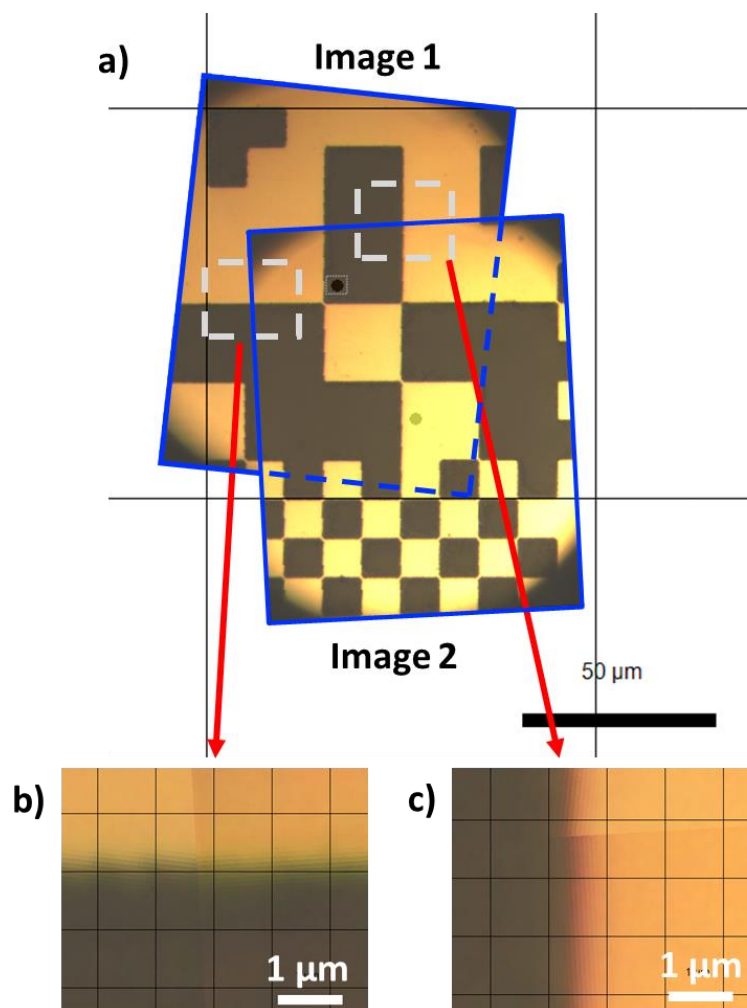


Figure 55 – Précision du nanoGPS. Troisième cas : Superposition de deux images après rotation de la lame.

platine et repositionnée avec un angle différent. Le but étant d'évaluer la capacité à prendre en compte la rotation angulaire de l'échantillon.

L'écart horizontal et l'écart vertical observés sont très largement inférieurs au micromètre. Le système appréhende parfaitement la manipulation d'une lame entre deux observations. Il n'y a donc aucune influence de la rotation de la lame lorsqu'elle est replacée sur la platine du microscope pour la construction de la cartographie.

#### 1.3.4. Cas n°4 : Superposition de deux images avec réétalonnage entre les deux observations

Le troisième cas illustre la construction d'une cartographie avec des observations enregistrées avec plusieurs jours d'écart. Pour simuler ce cas, nous avons réalisé avant

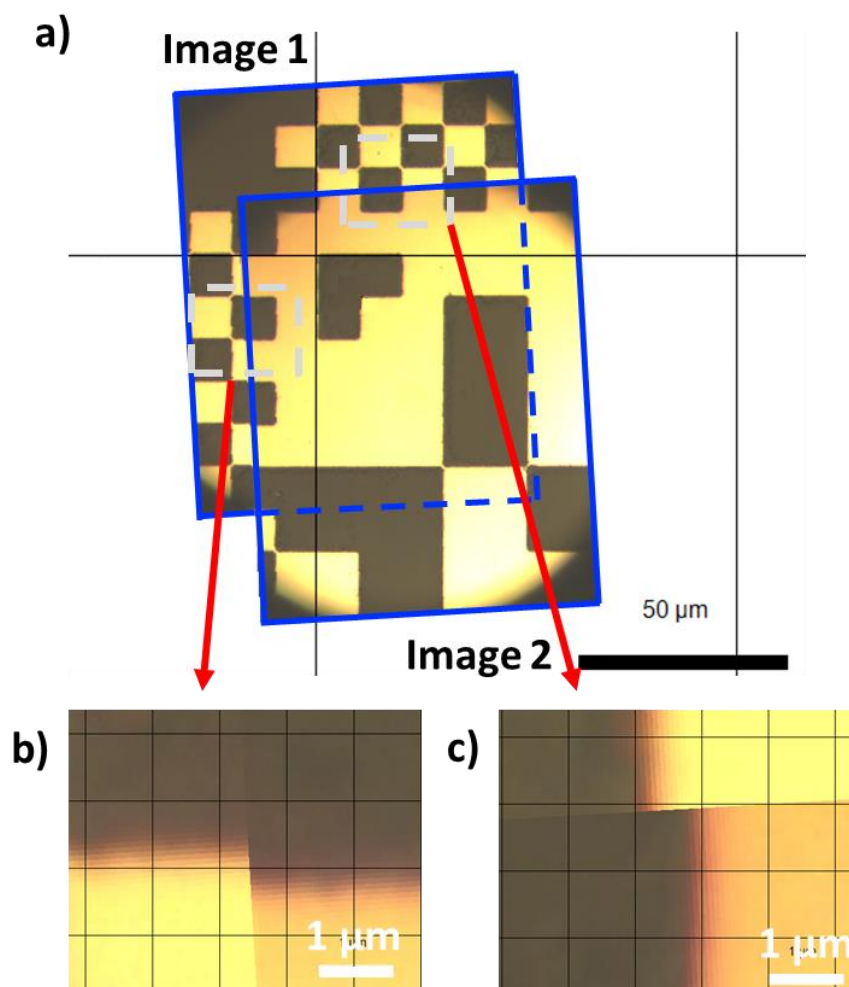


Figure 56 - Précision du nanoGPS. Quatrième cas : Superposition de deux images avec un réétalonnage entre chaque observation.

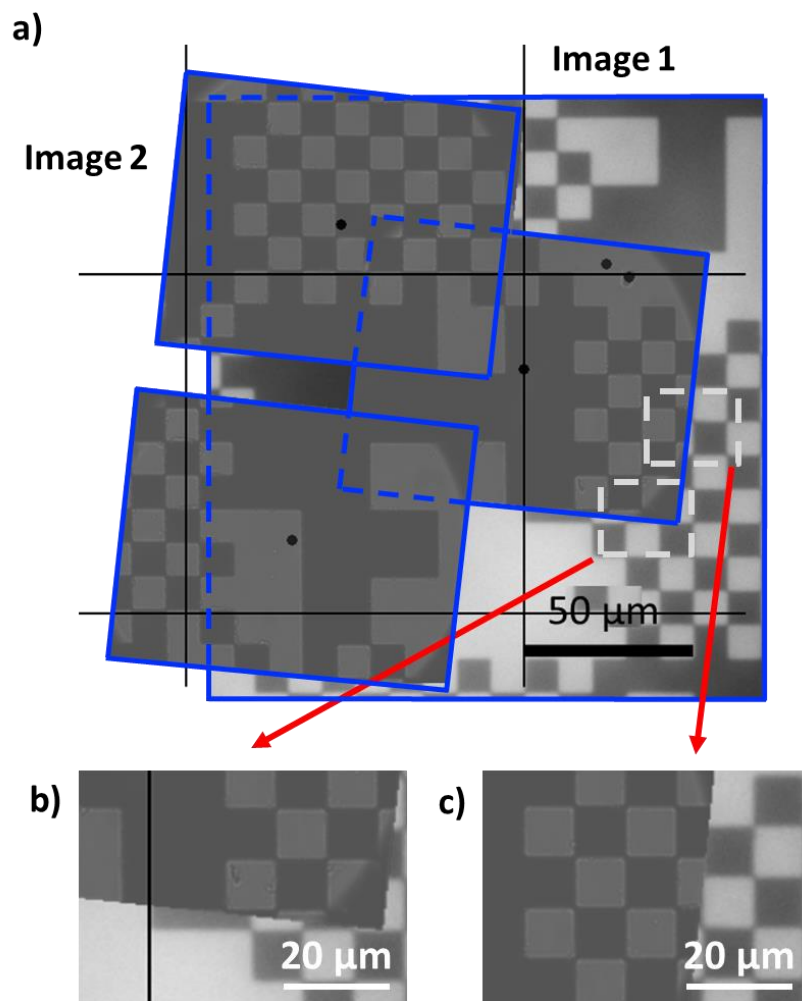


chacune des observations un réétalonnage du nanoGPS. La superposition des images est représentée sur la Figure 56.

Un réétalonnage entre les différentes mesures permet de corriger les erreurs dues à un décalage pouvant apparaître lors de la permutation des différents objectifs. Pour éviter d'accumuler les erreurs, un réétalonnage régulier, *a minima* quotidien, du nanoGPS est effectué lors des analyses.

**1.3.5. Cas n°5 : Superposition entre les images acquises sur le microscope du microspectromètre Raman et sur le microdissecteur**

Ce cinquième et dernier cas illustre la précision de colocalisation obtenue lorsque l'échantillon est déplacé d'un instrument à un autre. La précision du repositionnement est



**Figure 57 - Précision du nanoGPS. Cinquième cas : Superposition de deux images acquises sur les deux instruments.**

moins bonne que dans les cas précédents. L'étalonnage des nanoGPS sur le microdissecteur présente de plus grands écarts lors de la superposition des images dans le logiciel de cartographie. Ces décalages sont dus à des erreurs de repositionnement de la platine durant l'étape d'étalonnage du nanoGPS.

En effet, lors de la phase d'étalonnage du nanoGPS sur le microdissecteur, les images des motifs sont enregistrées à l'aide d'un objectif x40 par la caméra du microscope. Le flux vidéo doit ensuite être basculé vers la caméra du nanoGPS qui utilise l'objectif x10 pour lire les motifs. La platine effectue alors un déplacement pour permettre la rotation des objectifs et se replace à sa position initiale. Ce mouvement de l'instrument lors de l'étalonnage du nanoGPS entraîne un décalage non maîtrisé qui explique l'écart constaté dans la superposition des images sur la cartographie (Figure 49). Dans le pire des cas, le décalage peut être de 5  $\mu\text{m}$ . C'est la raison principale pour laquelle, le travail de prélèvements des bactéries par microdissection ne peut pas être réalisé automatiquement mais doit se faire manuellement. Il est donc indispensable de vérifier à l'aide des images enregistrées, que les cellules ciblées correspondent bien à celles analysées par microspectroscopie Raman. La dispersion des bactéries sur la lame permettra néanmoins d'assurer le prélèvement de la bactérie attendue.

## 2. Microdissection

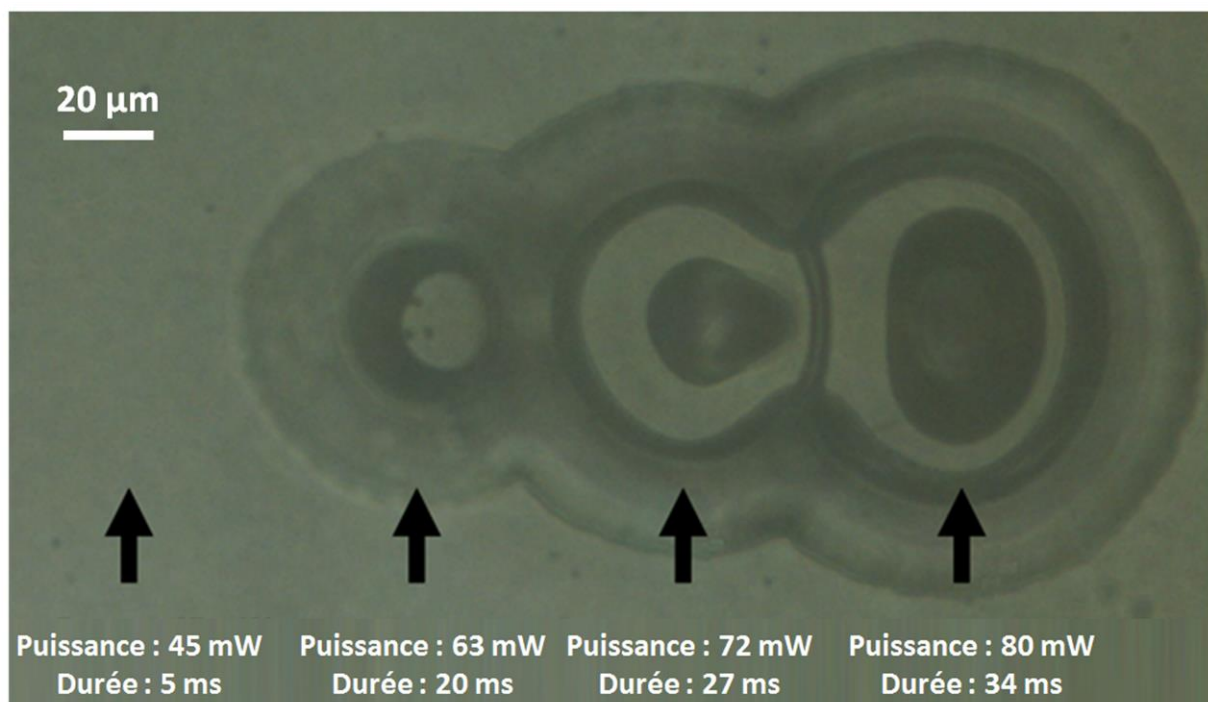
La microdissection laser est une étape clé dans le protocole. Elle permet d'extraire les bactéries en vue d'une analyse génomique. Néanmoins cette étape est très délicate car nous devons nous assurer que la majorité des cellules analysées sont effectivement prélevées.

### 2.1. *Optimisation de la taille du spot laser*

Pour réaliser cette tâche de manière optimale, il faut faire attention à ne prélever qu'une seule bactérie. Pour cela il a fallu procéder à l'optimisation des paramètres de l'impulsion laser permettant de faire fondre le polymère de capture. Ainsi initialement, le polymère n'est pas directement en contact avec l'échantillon afin d'éviter tout risque de contamination. Des rails maintiennent une distance entre la surface de l'échantillon et le film thermosensible. Il faut donc envoyer une impulsion laser suffisante afin de faire fondre le

polymère, en veillant à ce qu'elle ne soit pas trop importante au risque de recouvrir une trop grande surface de l'échantillon. Le principal risque en recouvrant une zone trop grande est le prélèvement de plusieurs cellules en même temps, ce qui n'est pas concevable lorsque nous voulons procéder à une analyse génomique sur cellule unique.

Sur la Figure 58, nous observons l'influence des deux principaux paramètres, à savoir la puissance du laser et la durée de l'impulsion. Si la puissance du laser et le temps de l'impulsion sont trop faibles alors le polymère risque de ne pas fondre. À l'inverse, si la puissance du laser et le temps d'impulsion sont trop importants, la surface de recouvrement de l'échantillon sera trop importante et le risque de prélever plusieurs cellules sera accru. Pour limiter au maximum ce risque, il faut réussir à obtenir une zone de recouvrement la plus étroite possible. Dans les cas les plus importants, c'est-à-dire en envoyant une impulsion laser avec un couple puissance/durée de 80mW/34ms et 72mW/27ms, la zone de recouvrement de l'échantillon par le film polymère thermosensible est trop importante. Les diamètres moyens obtenus pour ces zones de collecte sont respectivement de 65  $\mu\text{m}$  et de



**Figure 58 – Influence des paramètres de l'impulsion laser sur la taille des spots pour le prélèvement par microdissection laser.**

50  $\mu\text{m}$ . Les essais effectués à des puissances et des durées plus faibles que 63 mW et 20 ms ne sont pas suffisants pour réussir à obtenir un spot visible venant recouvrir la surface de la lame.

Les résultats des différents essais ont convergés vers une puissance de 63 mW pour une durée de 20 ms. Toutefois, la puissance et le temps de l'impulsion laser doivent être réajustés selon le lot de « caps » utilisé. En utilisant ces paramètres, les spots présentent un diamètre moyen d'une vingtaine de micromètres. Cette observation engendre une contrainte supplémentaire lors de l'analyse des particules. Il ne faut donc pas qu'une deuxième particule soit dans un rayon de moins de 10  $\mu\text{m}$  autour de la particule ciblée au risque de la prélever également.

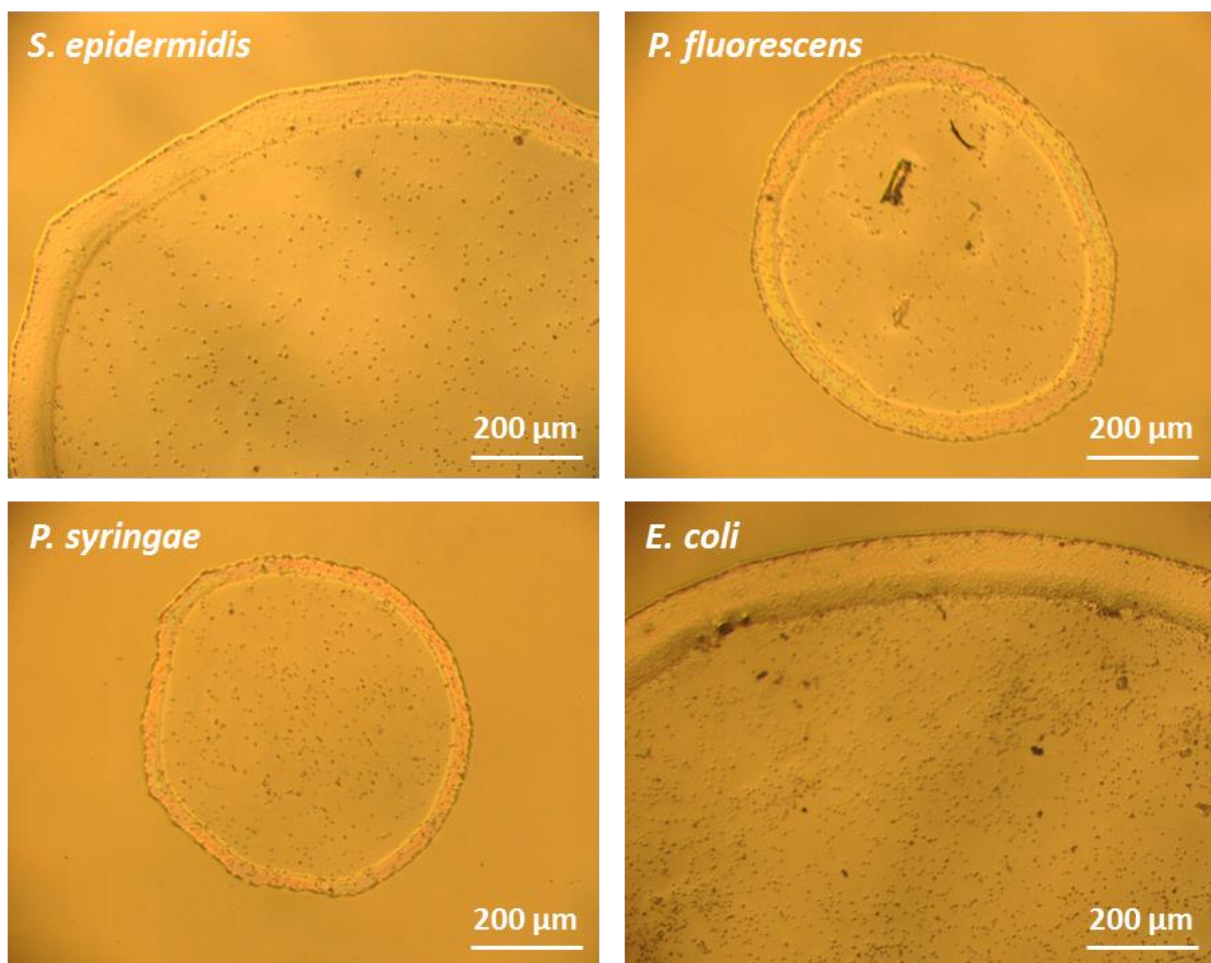
La microdissection laser est une méthode généralement effectuée sur des tissus ou des cellules de plus grande taille que les bactéries. Son utilisation pour un prélèvement sur cellule unique n'est pas une pratique courante. En effet, de par la taille de ces microorganismes, il est difficile de les capturer dans le film polymère. La surface de contact entre la cellule et le polymère est extrêmement faible réduisant alors le taux de réussite des prélèvements. Après avoir déterminé les paramètres permettant d'obtenir une zone restreinte d'extraction, nous avons souhaité évaluer le taux de prélèvements des cellules à partir de la lame de  $\text{CaF}_2$ .

## ***2.2. Influence des modes de dépôt des bactéries sur l'extraction des cellules par microdissection laser***

Le matériel microbiologique est si petit que la surface de contact entre le polymère et la cellule est très faible, ce qui rend la tâche de prélèvement complexe. Le taux de prélèvement, dépendant de facteurs souvent incontrôlables comme l'adhérence des particules au support de dépôt, est donc une variable que nous devons prendre en compte et maîtriser pour augmenter les chances de futures identifications génomiques.

### 2.2.1. Influence du dépôt par goutte

Nous avons vu que différents modes de dépôt des particules sur les lames de CaF<sub>2</sub> ont été testés. Nous savons déjà que le dépôt par goutte a une influence directe sur les résultats spectroscopiques mais ce n'est pas tout. En effet, lorsqu'un dépôt par goutte est effectué, nous avons remarqué la formation d'une fine pellicule de matière sur les cellules à analyser (Figure 59).



**Figure 59 - Influence du dépôt par goutte sur le prélèvement par microdissection laser.**

À cause de ce recouvrement, il devient non seulement impossible de réaliser l'acquisition des spectres Raman mais également impossible de prélever les bactéries par microdissection laser. Aucune des cellules que nous avons ciblées pour tenter un prélèvement par microdissection laser n'a pu être extrait correctement. Ce mode de dépôt n'est donc pas envisageable pour le protocole que nous développons.

### 2.2.2. Influence du dépôt par sédimentation

Pour cette étude, des bactéries ont été localisées sur une lame de *P. fluorescens* déposées par sédimentation après culture. Sur un total de 49 cellules ciblées, 47 ont pu être extraites du support. Nous pouvons observer sur la Figure 60, trois images représentant les différentes phases lors du prélèvement de cellules de *P. fluorescens* par microdissection laser. Tout d'abord, à gauche, une image de la lame avant le prélèvement, les bactéries ciblées sont celles qui sont au centre de chacun des cercles rouges. Ensuite, au centre, une image de la surface du polymère après impaction laser, avec au centre de chaque impact la bactérie arrachée. Certaines particules peuvent sembler plus visibles que d'autres mais cela est uniquement dû à la différence de distance focale selon la cellule observée. Enfin, à droite, une image de la lame après le prélèvement par LCM. Cette vérification faite sur la lame confirme que les cellules ciblées sont, dans ce cas, parfaitement extraites de la surface de l'échantillon.

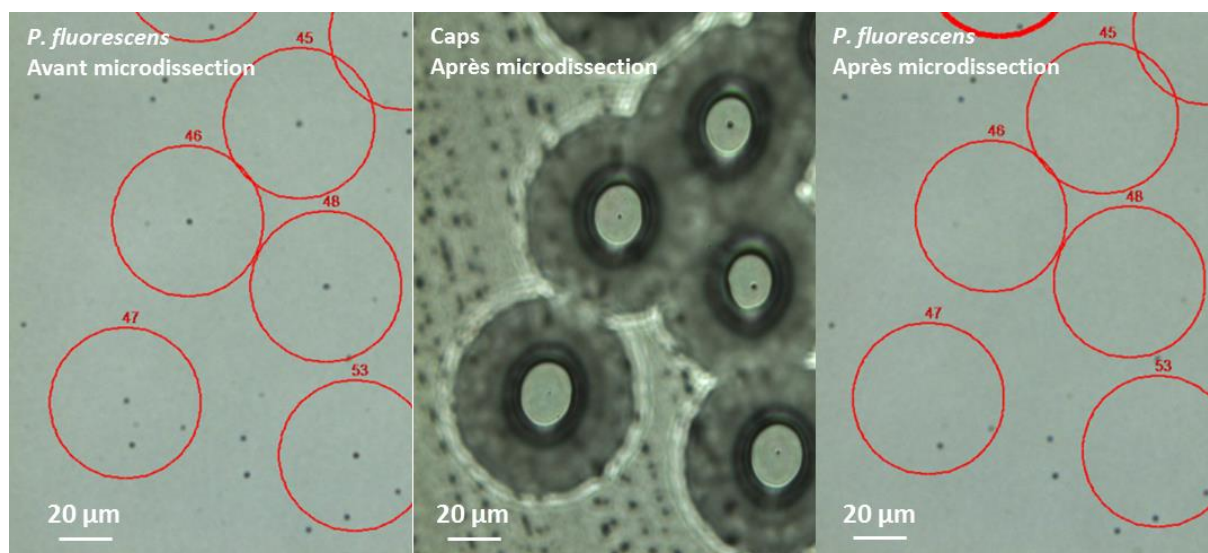
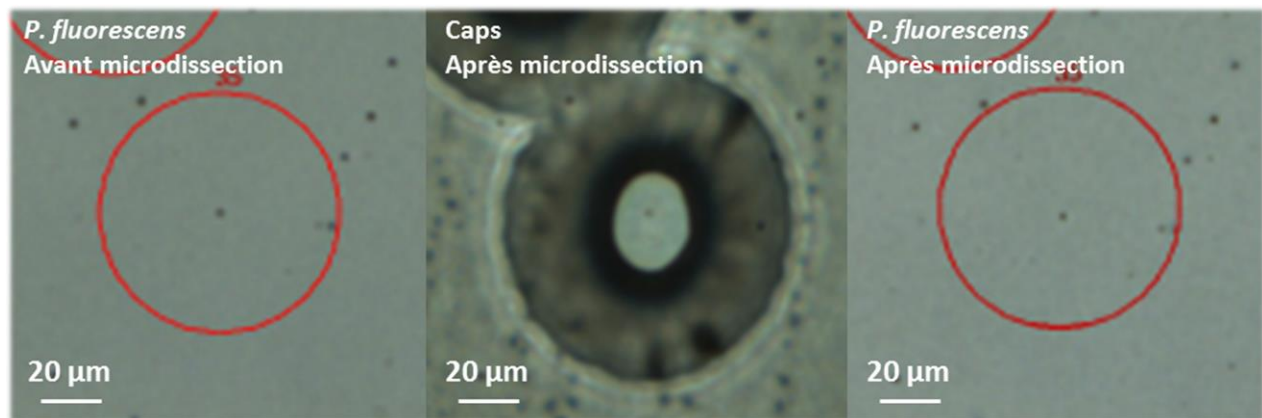


Figure 60 – Prélèvement de cellules de *P. fluorescens* par microdissection laser. À gauche, une image de la lame avant le prélèvement, au centre une image de la surface du polymère avec au centre de chaque impact la bactérie prélevée et à droite une image de la lame après le prélèvement.

### 2.3. Difficultés liées à l'adhérence des cellules

Même si le dépôt par sédimentation permet d'améliorer le taux de prélèvement des cellules, il y a tout de même des cellules qui semblent avoir une adhérence suffisamment forte

au support pour ne pas être extractibles. Lors de prélèvements sur certains échantillons, même si une empreinte est visible dans le polymère (Figure 61), la cellule ciblée pour la microdissection est en réalité toujours sur le support. Pour vérifier si la cellule est toujours présente sur la lame, l'échantillon est transféré sur le microspectromètre Raman et un spectre est acquis sur la cellule. Le nouveau spectre acquis est similaire à celui acquis avant le prélèvement. Pour vérifier que la marque au centre de l'impact dans le polymère n'est bien qu'une empreinte, nous effectuons une nouvelle impulsion laser pour faire fondre de nouveau le polymère. Lorsqu'une nouvelle impulsion laser est effectuée sur la zone d'impact, la marque disparaît. L'observation présentée sur l'image au centre de la figure n'est donc qu'une empreinte et non la cellule elle-même.



**Figure 61 – Tentative de prélèvement d'une cellule de *P. fluorescens*. À gauche, une image de la lame avant le prélèvement, au centre une image de la surface du polymère avec au centre l'empreinte de la bactérie ciblée pour le prélèvement et à droite, l'image de la lame après le prélèvement.**

Cette difficulté à prélever les cellules est indépendante de la préparation des échantillons. Nous avons été confrontés à ces problèmes avec des échantillons préparés selon un même mode opératoire, le même jour. En effet, avec les échantillons générés depuis les huit souches isolées d'un d'aérosol, l'espèce *K. palustris* n'a pas pu être microdissectée à cause de l'adhérence des cellules. Les sept autres n'ont pas présenté de difficultés particulières lors de l'extraction.

Plusieurs actions ont été entreprises pour remédier à ces difficultés. Nous avons dans un premier temps pensé qu'il pouvait s'agir d'un problème de conservation des « caps » ou

de validité du lot. Plusieurs lots ont donc été testés et les résultats se sont avérés être toujours les mêmes. Nous avons également modifié les paramètres de puissance et de durée de l'impulsion laser sans jamais réussir à extraire les cellules de *K. palustris*. La dernière tentative pour remédier aux difficultés liées à l'adhérence des cellules au support, a été l'utilisation d'une enceinte à atmosphère contrôlée. En faisant varier la température et le degré hygrométrique, nous pensions pouvoir diminuer l'adhérence. Nous avons fait varier l'hygrométrie de 40% à 80% par palier de 5% et la température de 20°C à 40°C par palier de 5°C dans l'enceinte à atmosphère contrôlée. À chaque fois qu'un paramètre est modifié, les cellules sont placées dans l'enceinte, puis une tentative de prélèvement par microdissection est effectuée. Toutes les tentatives menées sur la microdissection n'ont pas permis d'améliorer le taux de prélèvement. À ce jour, la seule certitude que nous avons est que le dépôt par sédimentation est la meilleure manière de déposer les bactéries sur les lames pour un prélèvement par microdissection. La diminution de l'adhérence des cellules est donc un point à améliorer pour augmenter les taux de prélèvements.

### 3. Analyse génomique sur cellule unique

À l'inverse de l'identification moléculaire sur bactéries de culture, l'identification sur cellule unique est une tâche complexe compte tenu de la faible quantité de matériel génétique présent sur la « caps ».

#### 3.1. Analyse génomique sur cellules non marquées

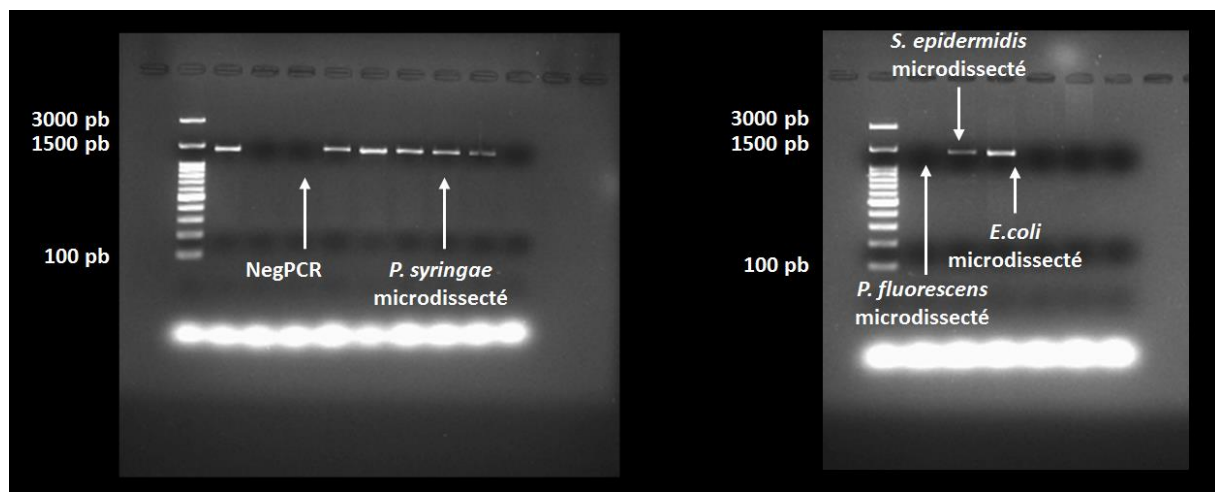
Pour rappel, les cellules microbiologiques arrivent au laboratoire du CSTB piégées dans le polymère recouvrant la surface inférieure de la « caps ». Chaque prélèvement est lysé et son génome est amplifié. Une fraction de l'ADN ainsi obtenu est amplifiée par PCR à l'aide d'amorces définissant un fragment de 1500 paires de bases dans le gène codant l'ARN 16S. Les produits de PCR sont ensuite séquencés et les séquences obtenues alignées sur la base GenBank du NCBI.

L'étude a été menée sur les quatre espèces principales. Les cellules prélevées sont issues de cultures déposées par sédimentation, elles n'ont pas subi l'étape de marquage au



SYTO9 et n'ont pas été analysées par microspectroscopie Raman. Le but premier de cette étude est de déterminer si le matériel génétique prélevé est en suffisamment grande quantité pour permettre une identification bactérienne par analyse génomique après microdissection.

Sur la Figure 62, sont présentés les résultats sur gel d'électrophorèse des produits obtenus suite à l'amplification génomique par PCR à l'aide des amorces 63F et 1387R.



**Figure 62 – Visualisation de l'amplification génomique par PCR sur gel d'électrophorèse des cellules microdissectées.**

Comme auparavant, un témoin négatif est réalisé sur une « caps » vierge pour attester de la présence de matériel génétique dans le cas d'une détection positive. Pour les prélèvements effectués sur les espèces *S. epidermidis*, *P. syringae* et *E. coli*, nous obtenons des produits de PCR à la taille attendue, visibles à 1500 pb. En revanche, pour la cellule de *P. fluorescens* microdissectée, le produit de PCR n'est pas visible sur le gel d'électrophorèse. Les séquences obtenues à partir des produits PCR sont alignées sur la base GenBank, mais les identifications obtenues ne correspondent pas aux espèces de départ. Pour les trois espèces qui sont observées sur le gel d'électrophorèse, des scores d'homologie de 100% sont obtenus pour le gène codant l'ARN 16S mais sont identifiées comme des bactéries non cultivables. Il est possible que des erreurs se soient infiltrées dans les séquences, ou bien qu'une dégradation trop importante de l'ADN au cours de l'une des étapes du procédé ait eu lieu, ce qui pourrait expliquer la difficulté à identifier les espèces analysées. Ces résultats suggèrent

tout de même qu'il est possible d'obtenir une séquence monoclonale à partir de bactéries microdissectées dans trois des quatre cas présentés.

### **3.2. Influence du marquage et du temps d'exposition laser sur l'analyse génomique**

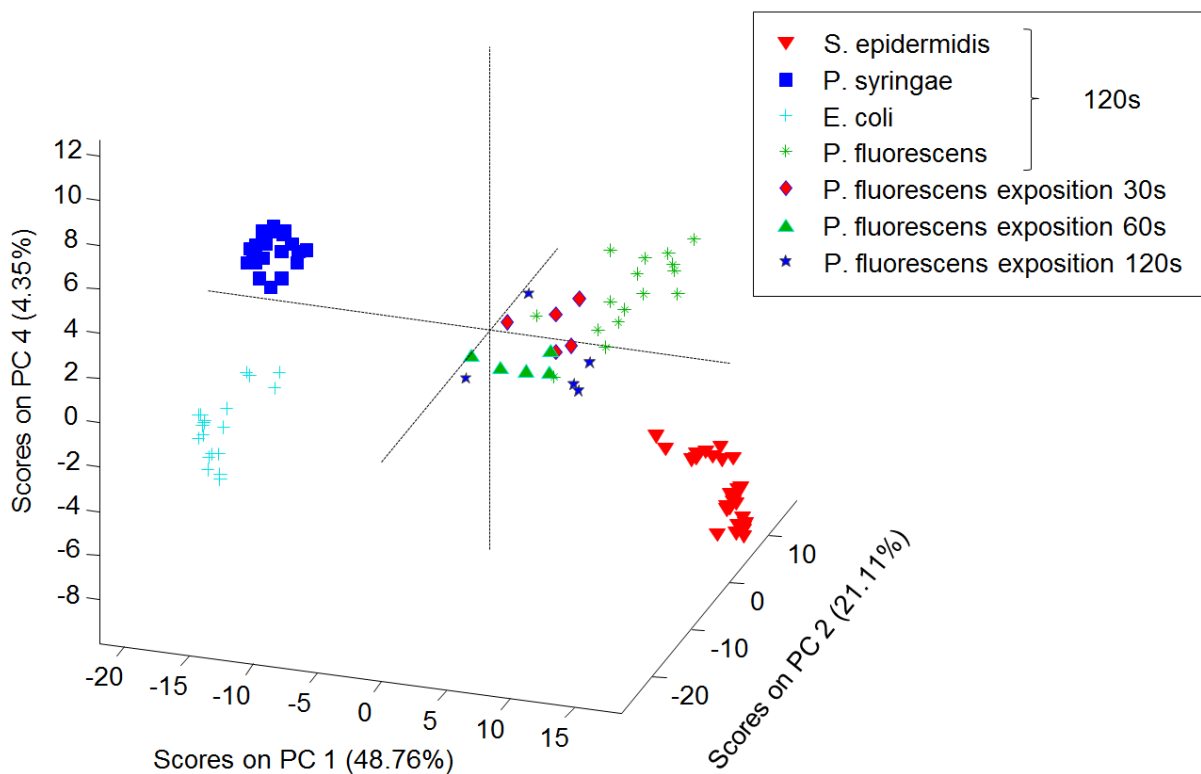
Durant cette nouvelle étude nous avons analysé simultanément, l'influence de deux paramètres pouvant se révéler critique vis-à-vis de l'identification des bactéries prélevées par microdissection. Des cellules de *P. fluorescens*, marquées par au SYTO9 sont utilisées pour évaluer l'impact du marquage et du temps d'exposition pour l'identification. Les bactéries sont exposées durant des durées variables (0, 30, 60 et 120 secondes) sous le laser émettant à 633 nm et sont prélevées par microdissection laser pour être identifiées par analyse génomique.

Les séquences obtenues sont clonales mais ne correspondent pas à la séquence de l'espèce *P. fluorescens* attendue uniquement dans un cas sur quatre. Seule la cellule correspondant à une acquisition du spectre Raman d'une durée de 120 secondes présente un résultat positif. Dans les autres cas, les bactéries sont des streptocoques distincts, qui pourraient être des bactéries commensales humaines.

À l'issus de cette étude, plusieurs conclusions peuvent être présentées. La première permet de dire qu'il y a une possible contamination lors de la manipulation des échantillons, causant une mauvaise identification des cellules prélevées. Ensuite, nous avons pu lever le verrou lié au marquage des agents bactériologiques. Le marquage des agents microbiologiques par le SYTO9 n'empêche pas l'identification génétique, même si le fluorophore vient s'insérer dans l'ADN. Enfin, une cellule qui a été exposée pendant 120 secondes au rayonnement laser lors de l'analyse Raman a pu être correctement identifiée. Nous supposons donc que les bactéries microdissectées, si le temps d'exposition au laser n'excède pas 120 secondes, peuvent être identifiées par analyse génomique.

La projection des spectres acquis sur les cellules aux différents temps d'exposition, sur un graphique ACP à quatre espèces, tend à dire que les bactéries analysées et prélevées sont de l'espèce *P. fluorescens*. Nous ne pouvons donc pas dire d'où vient la contamination si elle

a lieu. Le modèle ACP présenté ici ne considère que quatre espèces, et nous observons que les projections sont décentrées du nuage de points de l'espèce *P. fluorescens*. Même si la projection laisse à penser que nous sommes bien en présence de l'espèce souhaitée, nous ne pouvons pas l'affirmer formellement.



**Figure 63 – Projection des spectres de *P. fluorescens* acquis aux différents temps d'exposition sur un graphique ACP à quatre espèces.**

Nous voyons bien dans cette dernière partie, toute la problématique liée au prélèvement de cellules uniques et à l'identification basée sur une faible quantité de matière.

## 4. Conclusion

Dans ce chapitre, nous n'avons vu qu'une partie du potentiel qu'offre le module nanoGPS. Cet outil permet de coupler virtuellement deux instruments, comme nous l'avons fait avec le microspectromètre Raman et le microdissecteur à capture laser. Il permet un retour sur cellule unique avec une grande précision tant bien pour des problématiques de relocalisation et de colocalisation. Nous avons également pu appréhender les difficultés liées

à la microdissection sur cellule unique. Cette étape est une des plus délicates du protocole que nous avons développé. En effet, il suffit que l'adhérence d'une cellule soit légèrement trop importante pour que l'étape de microdissection soit compromise. Enfin, l'identification moléculaire de cellules uniques isolées par microdissection laser est possible. Néanmoins, cette étape est la plus difficile à optimiser et à reproduire avec un bon pourcentage d'identification. Le travail d'amplification génomique sur cellule unique est une tâche aisée sur des cellules cultivées ce qui laisse présager qu'elle le sera encore moins sur des bactéries provenant d'un aérosol naturel qui n'a pas subi de transformation.

# Conclusions et perspectives

La biosurveillance de l'air présente aujourd'hui un enjeu important tant dans le domaine civil que militaire. Il est donc nécessaire de déployer des moyens efficaces pour prévenir des risques d'une épidémie à cause de la libre circulation d'agents bactériens dans l'air. Ainsi, la DGA a lancé en 2012, le programme ASTRID qui permet le financement de projets scientifiques d'innovation défense comme le projet SIBIRAM, dans lequel s'inscrivent ces travaux de thèse. L'objectif principal des recherches menées au cours de cette thèse était d'apporter une preuve solide de la faisabilité de concept de détection et d'identification rapides des microorganismes présents dans un échantillon d'air par microscopie Raman, avec une résolution au niveau de l'espèce. Pour répondre à ce challenge, nous avons développé un protocole original permettant d'acquérir la signature spectrale Raman des microorganismes, suivi d'une extraction individuelle par microdissection laser des cellules analysées dans le but de réaliser une identification par analyse génomique sur cellule unique.

Le premier des objectifs de ces travaux a été de mettre en place le protocole d'acquisition des spectres Raman sur bactérie unique. Pour cela, nous avons testé différents supports avant de sélectionner des lames en  $\text{CaF}_2$ . Plusieurs lasers à différentes longueurs d'onde ont pu être testés. Les résultats ont conduit à conserver le laser émettant à 633 nm pour la suite des opérations. Nous avons également choisi le réseau de diffraction à utiliser et déterminé un temps d'exposition optimal pour ne pas dégrader le matériel biologique. Pour obtenir des spectres d'une qualité suffisante, c'est-à-dire un rapport signal/bruit suffisamment élevé pour permettre une discrimination des espèces, le temps d'analyse a été fixé à 60s par fenêtre spectrale en utilisant un réseau de diffraction à 300 traits/mm.

Concernant la collecte des aérosols, nous avons testé plusieurs modes de collecte. Cependant, face à la complexité des aérosols naturels, le protocole retenu a été une collecte en milieu liquide à l'aide du *SKC Biosampler*. Ce mode opératoire a permis à la fois une sélection des particules en taille et une diminution de la densité particulaire contrairement à une collecte à l'aide d'un impacteur en cascade. Nous avons également travaillé sur l'amélioration de la localisation des bactéries dans les échantillons naturels en procédant à un

marquage des bactéries par agent fluorescent (SYTO9). Cependant, comme certaines particules présentent une autofluorescence naturelle, nous avons également exploité un logiciel d'analyse d'image (ParticleFinder) nous permettant d'appliquer un critère de discrimination en taille. Nos expériences nous ont aussi permis d'observer que le mode de dépôt des particules sur les lames de CaF<sub>2</sub> avait une influence directe non seulement sur l'analyse Raman mais également sur la microdissection laser des cellules. C'est pourquoi, nous avons opté pour un dépôt par aérosolisation/sédimentation des particules sur les lames. Ce mode de dépôt évite la formation d'une pellicule sur la surface de l'échantillon et diminue l'adhérence des cellules au support permettant ainsi une meilleure extraction.

Le second objectif de ces travaux de thèse était la mise en place d'une méthode d'extraction de bactéries uniques par microdissection à capture laser en vue d'une identification par analyse génomique. Nous avons présenté les difficultés rencontrées lors de la phase d'optimisation de la procédure, comme l'influence de la puissance et de la durée de l'impulsion laser ou encore les problèmes d'adhérence observés sur les différents échantillons analysés. Les cellules extraites ont ensuite été expédiées aux laboratoires du CSTB pour subir un protocole d'amplification génomique et un séquençage des ARN 16S pour obtenir l'identité réelle des cellules.

Toutefois, nous devons être certain que la cellule microdissectée est bien celle qui a été analysée par microspectroscopie Raman. Pour cela nous avons réalisé un couplage virtuel des instruments à l'aide d'un module instrumental original nommé « nanoGPS ». Ce système permet un repositionnement absolu sur n'importe quel échantillon, à condition d'utiliser des lames spécifiques sur lesquelles sont imprimées par lithographie des motifs semblables à des « QRcodes ». Ainsi dans ce manuscrit, nous avons consacré toute une partie à l'explication du fonctionnement module nanoGPS, son implémentation sur les différents systèmes et à l'évaluation de ses performances. Les différentes études ont permis de montrer que les nanoGPS proposent un très bon repositionnement pour notre application, permettant de retrouver avec précision et certitude les cellules analysées par microspectroscopie Raman.

Un autre point important présenté lors de ces travaux est le développement d'une analyse chimiométrique sur les données acquises à partir des différents échantillons. Nous avons montré qu'une étude des prétraitements est impérative pour maximiser la

discrimination entre les différentes espèces analysées. Cette étude a révélée qu'il est nécessaire d'appliquer une dérivée de Savitzky-Golay suivie d'une normalisation SNV pour apporter des corrections sur les fonds de fluorescence, sur les variations de ligne de base et pour normaliser les intensités Raman d'un spectre à l'autre. Suite à cette étude des prétraitements mathématiques, nous avons travaillé sur le calcul d'un modèle de classification chimiométrique. Dans la littérature, la classification SVM est celle qui offre souvent les meilleures performances de classification pour les jeux de données complexes. Ainsi, nous avons travaillé sur l'optimisation des méta-paramètres C et G pour éviter tout risque de surentrainement du modèle et une complexité des frontières trop importante. Les premiers modèles de classification calculés à partir des spectres des quatre espèces étudiées (*Staphylococcus epidermidis*, *Pseudomonas fluorescens*, *Pseudomonas syringae* et *Escherichia coli*) ont permis d'atteindre un taux de classification moyen de plus de 99%. Grâce au protocole d'isolement de souches environnementales remises en culture, nous avons aussi isolé huit souches différentes. L'acquisition des spectres, suivi d'une étude des prétraitements puis d'une optimisation des paramètres C et G ont permis de construire un modèle de classification. Cependant pour obtenir une vision plus représentative, nous avons employé la méthode *bootstrap* pour une estimation globale des performances du modèle de classification. Ainsi nous avons obtenu un taux supérieur à 98% de bonne classification ce qui est tout à fait pertinent face aux faibles différences spectrales observées entre les bactéries.

Cette thèse a aussi été l'occasion de mettre en place, avec notre partenaire du laboratoire XLIM (Université de Limoges), quelques expériences de spectroscopie M-CARS sur des bactéries. La vitesse d'acquisition en spectroscopie Raman conventionnel ne nous permettrait effectivement pas d'assurer une classification de plusieurs milliers de particules par seconde, cadence nécessaire au déclenchement d'une alerte pertinente. Ces quelques expériences ont montré que des contributions spectrales étaient observables mais que la qualité spectrale était insuffisante pour créer une discrimination. Nous avons aussi montré que les bactéries isolées supportaient assez difficilement l'énergie des lasers pulsés. Le développement d'un futur appareil de classification passe ainsi par une optimisation de la mesure M-CARS.

Même si l'objectif d'analyse génomique sur des bactéries environnementales non cultivées n'est pas atteint, un grand nombre de verrous ont pu être levés pendant cette thèse.

En plus des difficultés liées aux échantillons naturels ou à l'adhérence des bactéries, des difficultés instrumentales sont venues s'ajouter. Nous avons cependant pu résoudre la majeure partie des problèmes auxquels nous avons été confrontés et avons tout de même démontré que les modèles de classification calculés présentent des taux de classification très satisfaisants. Néanmoins nous devons poursuivre nos efforts sur la procédure d'identification par analyse génomique. La tâche est difficile mais nous pensons qu'il est possible de l'améliorer pour permettre une identification des cellules extraites par microdissection. Lorsque cette étape sera réellement optimisée, il sera possible de réaliser un développement chimiométrique plus poussé sur les futurs échantillons. Nous pouvons très facilement imaginer augmenter le nombre d'espèces à inclure dans le modèle de classification, évaluer les performances du modèle face à des espèces qui n'ont pas encore été rencontrées. Il serait possible de développer un logiciel permettant de signaler la présence d'un individu alors inconnu du modèle durant la phase d'acquisition des spectres Raman sur microspectromètre afin d'opérer en priorité une microdissection laser pour obtenir son identité par analyse génomique. Avant la construction d'un premier prototype de détecteur en temps réel d'agents bactériologiques dans l'air, il reste énormément de travail, mais ces premiers résultats présentés dans ces travaux de thèse prouvent que l'analyse par spectroscopie Raman couplée aux outils chimiométriques peut potentiellement répondre à la problématique de la biosurveillance de l'air.



# Bibliographie

- [1] Y. Lévi, « Inquiétudes sur la présence d'antibiotiques et de bactéries antibiorésistantes dans les eaux », *Environ. Risques Santé*, vol. 5, n° 4, p. 261-265, juill. 2006.
- [2] D. L. Hawksworth et F. Rose, « Qualitative Scale for estimating Sulphur Dioxide Air Pollution in England and Wales using Epiphytic Lichens », *Nature*, vol. 227, n° 5254, p. 145-148, juill. 1970.
- [3] D. A. Shea et S. A. Lister, « The BioWatch Program : Detection of Bioterrorism », Congressional Research Service Report No. RL 32152, nov. 2003.
- [4] C. V. Raman et K. S. Krishnan, « A new type of secondary radiation (Nature, 1928, 121, 501-502) », *Curr. Sci. -BANGALORE-*, vol. 74, n° 4, p. 381, 1998.
- [5] G. Landsberg et L. Mandelstam, « Eine neue Erscheinung bei der Lichtzerstreuung in Krystallen », *Naturwissenschaften*, vol. 16, n° 28, p. 557-558, juill. 1928.
- [6] M. Delhaye, P. Dhamelinourt, « Raman microprobe and microscope with laser excitation » J. Raman, spectrosc., vol. 3, p. 33-42, 1975.
- [7] P. Dhamelinourt, F. Wallart, M. Leclercq, A.T. N'Guyen et D.O. Landon, « Laser Raman Molecular Microprobe (Mole) » *An. chem.*, vol. 51, p.414-419A, 1979.
- [8] M. Emmert-Buck *et al.*, « Laser Capture Microdissection », *Science*, vol. 274, p. 998, 1996.
- [9] M. BRENNAN, « Laser dissection captures cells for early cancer detection », *Chem. Eng. News Arch.*, vol. 74, n° 47, p. 33-34, nov. 1996.
- [10] Y. Kang, I. McMillan, M. H. Norris, et T. T. Hoang, « Single prokaryotic cell isolation and total transcript amplification protocol for transcriptomic analysis », *Nat. Protoc.*, vol. 10, n° 7, p. 974-984, juill. 2015.

- [11] D. L. Massart, B. G. Vandeginste, L. M. C. Buydens, P. J. Lewi, J. Smeyers-Verbeke, et S. D. Jong, *Handbook of Chemometrics and Qualimetrics: Part A*. New York, NY, USA: Elsevier Science Inc., 1997.
- [12] A. Savitzky et M. J. E. Golay, « Smoothing and Differentiation of Data by Simplified Least Squares Procedures. », *Anal. Chem.*, vol. 36, n° 8, p. 1627-1639, juill. 1964.
- [13] S. Wold, K. Esbensen, et P. Geladi, « Principal component analysis », *Chemom. Intell. Lab. Syst.*, vol. 2, n° 1, p. 37-52, août 1987.
- [14] K. Pearson, « On lines and planes of closest fit to systems of points in space », *Philos. Mag.*, vol. 2, n° 11, p. 559-572, nov. 1901.
- [15] H. Hotelling, *Analysis of a complex of statistical variables into principal components*. Baltimore: Warwick & York, 1933.
- [16] H. Wold, « Causal flows with latent variables: Partings of the ways in the light of NIPALS modelling », *Eur. Econ. Rev.*, vol. 5, n° 1, p. 67-86, 1974.
- [17] C. Cortes et V. Vapnik, « Support-Vector Networks », *Mach. Learn.*, vol. 20, n° 3, p. 273-297, sept. 1995.
- [18] M. S. Bazaraa, H. D. Sherali, et C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3 edition. Hoboken, N.J: Wiley-Interscience, 2006.
- [19] O. Devos, C. Ruckebusch, A. Durand, L. Duponchel, et J.-P. Huvenne, « Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation », *Chemom. Intell. Lab. Syst.*, vol. 96, p. 27-33, mars 2009.
- [20] J. E. S. Greenstreet et K. P. Norris, « The existence of differences between the infra-red absorption spectra of bacteria », *Spectrochim. Acta*, vol. 9, n° 3, p. 177-198, janv. 1957.
- [21] H. M. Randall, D. W. Smith, A. C. Colm, et W. J. Nungester, « Correlation of Biologic Properties of Strains of Mycobacterium with Infra-Red Spectrums. I. Reproductibility of Extracts of *M. tuberculosis* as determined by Infra-Red Spectroscopy. », *Am. Rev. Tuberc. Pulm. Dis.*, vol. 63, n° 4, p. 372-80, 1951.

- [22] H. M. Randall, D. W. Smith, et W. J. Nungester, « Correlation of Biologic Properties of Strains of Mycobacterium with their Infrared Spectrums. II. The Differentiation of Two Strains, H37Rv and H37Ra, of *M. tuberculosis* by means of their Infrared Spectrums. », *Am. Rev. Tuberc. Pulm. Dis.*, vol. 65, n° 4, p. 477-80, 1952.
- [23] J. W. Riddle, P. W. Kabler, B. A. Kenner, R. H. Bordner, S. W. Rockwood, et H. J. R. Stevenson, « BACTERIAL IDENTIFICATION BY INFRARED SPECTROPHOTOMETRY1 », *J. Bacteriol.*, vol. 72, n° 5, p. 593-603, nov. 1956.
- [24] R. D. Holland *et al.*, « Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry », *Rapid Commun. Mass Spectrom. RCM*, vol. 10, n° 10, p. 1227-1232, 1996.
- [25] K. Oliveira, G. W. Procop, D. Wilson, J. Coull, et H. Stender, « Rapid identification of *Staphylococcus aureus* directly from blood cultures by fluorescence in situ hybridization with peptide nucleic acid probes », *J. Clin. Microbiol.*, vol. 40, n° 1, p. 247-251, janv. 2002.
- [26] D. Naumann, D. Helm, et H. Labischinski, « Microbiological characterizations by FT-IR spectroscopy », *Nature*, vol. 351, n° 6321, p. 81-82, mai 1991.
- [27] Y. Feng, G. Downey, D.-W. Sun, D. Walsh, et J. Xu, « Towards improvement in classification of *Escherichia coli*, *Listeria innocua* and their strains in isolated systems based on chemometric analysis of visible and near-infrared spectroscopic data », *J. Food Eng.*, vol. 149, p. 87-96, mars 2015.
- [28] G. J. Puppels *et al.*, « Studying single living cells and chromosomes by confocal Raman microspectroscopy », *Nature*, vol. 347, n° 6290, p. 301-303, sept. 1990.
- [29] R. M. Jarvis et R. Goodacre, « Characterisation and identification of bacteria using SERS », *Chem. Soc. Rev.*, vol. 37, n° 5, p. 931-936, mai 2008.
- [30] K. Maquelin, L.-P. Choo-Smith, H. P. Endtz, H. A. Bruining, et G. J. Puppels, « Rapid Identification of *Candida* Species by Confocal Raman Microspectroscopy », *J. Clin. Microbiol.*, vol. 40, n° 2, p. 594-600, févr. 2002.

- [31] K. Maquelin *et al.*, « Identification of medically relevant microorganisms by vibrational spectroscopy », *J. Microbiol. Methods*, vol. 51, n° 3, p. 255-271, nov. 2002.
- [32] L. P. Choo-Smith, K. Maquelin, H. P. Endtz, H. A. Bruining, et G. J. Puppels, « A novel method for rapid identification of micro-organisms using confocal Raman microspectroscopy », in *Spectroscopy of Biological Molecules: New Directions*, Springer, Dordrecht, 1999, p. 537-540.
- [33] M. Harz, P. Rösch, et J. Popp, « Vibrational spectroscopy--a powerful tool for the rapid identification of microbial cells at the single-cell level », *Cytom. Part J. Int. Soc. Anal. Cytol.*, vol. 75, n° 2, p. 104-113, févr. 2009.
- [34] P. Rösch *et al.*, « On-Line Monitoring and Identification of Bioaerosols », *Anal. Chem.*, vol. 78, n° 7, p. 2163-2170, avr. 2006.
- [35] P. Rösch, M. Harz, K.-D. Peschke, O. Ronneberger, H. Burkhardt, et J. Popp, « Identification of single eukaryotic cells with micro-Raman spectroscopy », *Biopolymers*, vol. 82, n° 4, p. 312-316, juill. 2006.
- [36] P. Rösch *et al.*, « Chemotaxonomic Identification of Single Bacteria by Micro-Raman Spectroscopy: Application to Clean-Room-Relevant Biological Contaminations », *Appl. Environ. Microbiol.*, vol. 71, n° 3, p. 1626-1637, janv. 2005.
- [37] M. Harz, P. Rösch, K.-D. Peschke, O. Ronneberger, H. Burkhardt, et J. Popp, « Micro-Raman spectroscopic identification of bacterial cells of the genus *Staphylococcus* and dependence on their cultivation conditions », *The Analyst*, vol. 130, n° 11, p. 1543-1550, nov. 2005.
- [38] U. Schmid, P. Rösch, M. Krause, M. Harz, J. Popp, et K. Baumann, « Gaussian mixture discriminant analysis for the single-cell differentiation of bacteria using micro-Raman spectroscopy », *Chemom. Intell. Lab. Syst.*, vol. 96, n° 2, p. 159-171, avr. 2009.
- [39] C. R. Woese, O. Kandler, et M. L. Wheelis, « Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. », *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, n° 12, p. 4576-4579, juin 1990.

- [40] B. Vivien *et al.*, « Prise en charge d'une victime NRBC », 2012.
- [41] J. D. CAVALLO *et al.*, « Surveillance microbiologique de l'environnement dans les établissements de santé », 2002.
- [42] « Bacterial Pathogens of Humans ». [En ligne]. Disponible sur: <http://textbookofbacteriology.net/medical.html>. [Consulté le: 17-oct-2017].
- [43] V. A. Marple, « History of Impactors—The First 110 Years », *Aerosol Sci. Technol.*, vol. 38, n° 3, p. 247-292, mars 2004.
- [44] P. S. Thorne, M. S. Kiekhaefer, P. Whitten, et K. J. Donham, « Comparison of bioaerosol sampling methods in barns housing swine », *Appl. Environ. Microbiol.*, vol. 58, n° 8, p. 2543-2551, août 1992.
- [45] « College of Agricultural, Consumer and Environmental Sciences | New Mexico State University ». [En ligne]. Disponible sur: <http://aces.nmsu.edu/>. [Consulté le: 04-oct-2017].
- [46] H. Zhen, T. Han, D. E. Fennell, et G. Mainelis, « Release of Free DNA by Membrane-Impaired Bacterial Aerosols Due to Aerosolization and Air Sampling », *Appl. Environ. Microbiol.*, vol. 79, n° 24, p. 7780-7789, déc. 2013.
- [47] J. D. Hader, T. P. Wright, et M. D. Petters, « Contribution of pollen to atmospheric ice nuclei concentrations », *Atmos Chem Phys*, vol. 14, n° 11, p. 5433-5449, juin 2014.
- [48] T. Han, M. Wren, K. DuBois, J. Therkorn, et G. Mainelis, « Application of ATP-based bioluminescence for bioaerosol quantification: Effect of sampling method », *J. Aerosol Sci.*, vol. 90, n° Supplement C, p. 114-123, déc. 2015.
- [49] K. willeke, X. Lin, et S. A. Grinshpun, « Improved Aerosol Collection by Combined Impaction and Centrifugal Motion », *Aerosol Sci. Technol.*, vol. 28, n° 5, p. 439-456, janv. 1998.

- [50] J. R. Marchesi *et al.*, « Design and Evaluation of Useful Bacterium-Specific PCR Primers That Amplify Genes Coding for Bacterial 16S rRNA », *Appl. Environ. Microbiol.*, vol. 64, n° 2, p. 795-799, févr. 1998.
- [51] « National Center for Biotechnology Information ». [En ligne]. Disponible sur: <https://www.ncbi.nlm.nih.gov/>. [Consulté le: 05-oct-2017].
- [52] H. P. Horz, M. E. Vianna, B. P. F. A. Gomes, et G. Conrads, « Evaluation of Universal Probes and Primer Sets for Assessing Total Bacterial Load in Clinical Samples: General Implications and Practical Use in Endodontic Antimicrobial Therapy », *J. Clin. Microbiol.*, vol. 43, n° 10, p. 5332-5337, oct. 2005.
- [53] J. D. Driskell, Y. Zhu, C. D. Kirkwood, Y. Zhao, R. A. Dluhy, et R. A. Tripp, « Rapid and Sensitive Detection of Rotavirus Molecular Signatures Using Surface Enhanced Raman Spectroscopy », *PLoS ONE*, vol. 5, n° 4, avr. 2010.
- [54] K. Hamasha, Q. I. Mohaidat, R. A. Putnam, R. C. Woodman, S. Palchaudhuri, et S. J. Rehse, « Sensitive and specific discrimination of pathogenic and nonpathogenic *Escherichia coli* using Raman spectroscopy—a comparison of two multivariate analysis techniques », *Biomed. Opt. Express*, vol. 4, n° 4, p. 481-489, mars 2013.
- [55] K. C. Schuster, E. Urlaub, et J. R. Gapes, « Single-cell analysis of bacteria by Raman microscopy: spectral information on the chemical composition of cells and on the heterogeneity in a culture », *J. Microbiol. Methods*, vol. 42, n° 1, p. 29-38, sept. 2000.
- [56] M. Li *et al.*, « Rapid resonance Raman microspectroscopy to probe carbon dioxide fixation by single cells in microbial communities », *ISME J.*, vol. 6, n° 4, p. 875-885, avr. 2012.
- [57] Y. Wang *et al.*, « Raman Activated Cell Ejection for Isolation of Single Cells », *Anal. Chem.*, vol. 85, n° 22, p. 10697-10701, nov. 2013.
- [58] W. E. Huang, M. Li, R. M. Jarvis, R. Goodacre, et S. A. Banwart, « Shining light on the microbial world the application of Raman microspectroscopy », *Adv. Appl. Microbiol.*, vol. 70, p. 153-186, 2010.

- [59] S. A. Strola *et al.*, « Single bacteria identification by Raman spectroscopy », *J. Biomed. Opt.*, vol. 19, n° 11, p. 111610, 2014.
- [60] J.-X. Cheng, Y. K. Jia, G. Zheng, et X. S. Xie, « Laser-scanning coherent anti-Stokes Raman scattering microscopy and applications to cell biology », *Biophys. J.*, vol. 83, n° 1, p. 502-509, juill. 2002.
- [61] M. Okuno *et al.*, « Quantitative CARS Molecular Fingerprinting of Single Living Cells with the Use of the Maximum Entropy Method », *Angew. Chem. Int. Ed.*, vol. 49, n° 38, p. 6773-6777, sept. 2010.
- [62] C.-C. Chang et C.-J. Lin, « LIBSVM: a library for support vector machines », *ACM Trans. Intell. Syst. Technol. TIST*, vol. 2, n° 3, p. 27, 2011.
- [63] R. L. Mancinelli et W. A. Shulls, « Airborne bacteria in an urban environment », *Appl. Environ. Microbiol.*, vol. 35, n° 6, p. 1095-1101, juin 1978.
- [64] X. Dou *et al.*, « Quantitative analysis of double-stranded DNA amplified by a polymerase chain reaction employing surface-enhanced Raman spectroscopy », *Appl. Opt.*, vol. 37, n° 4, p. 759-763, févr. 1998.
- [65] P. Stiefel, S. Schmidt-Emrich, K. Maniura-Weber, et Q. Ren, « Critical aspects of using bacterial cell viability assays with the fluorophores SYTO9 and propidium iodide », *BMC Microbiol.*, vol. 15, p. 36, févr. 2015.

