UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Programa de Pós-Graduação em Química

Espectroscopia e cromatografia líquida com espectrometria de massa associadas à quimiometria na classificação e avaliação de perfil lipidômico de classes bacterianas

Aline de Sousa Marques

Tese de Doutorado

Natal/RN, agosto de 2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE CIÊNCIAS EXATAS E DA TERRA

INSTITUTO DE QUÍMICA

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

ALINE DE SOUSA MARQUES

ESPECTROSCOPIA E CROMATOGRAFIA LÍQUIDA COM ESPECTROMETRIA DE MASSA ASSOCIADAS À QUIMIOMETRIA NA CLASSIFICAÇÃO E AVALIAÇÃO DE PERFIL LIPIDÔMICO DE CLASSES BACTERIANAS.

Tese submetida ao Programa de Pós-Graduação em Química da Universidade Federal do Rio Grande do Norte (PPGQ/UFRN), como parte dos requisitos necessários para obtenção do título de Doutor em Química.

Orientador: Prof. Dr. Kássio Michell Gomes de Lima

NATAL – RN

2017

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Marques, Aline de Sousa.
Espectroscopia e cromatografia líquida com espectrometria de massa associadas à quimiometria na classificação e avaliação de perfil lipidômico de classes bacterianas / Aline de Sousa Marques. - 2017.
105 f.: il.

Tese (doutorado) - Universidade Federal do Rio Grande do Norte. Centro de Ciências Exatas e da Terra. Instituto de Química. Programa de Pós-Graduação em Química. Natal, RN, 2017.
Orientador: Kássio Michell Gomes de Lima.

1. Bioespectroscopia - Tese. 2. SPA-LDA - Tese. 3. GA-LDA - Tese. 4. NIR - Tese. 5. ATR-FTIR - Tese. 6. ROI - Tese. 7. LC-MS - Tese. I. Lima, Kássio Michell Gomes de. II. Título.

RN/UF/CCET CDU 543.429.9

Aline de Sousa Marques

ESPECTROSCOPIA E CROMATOGRAFIA LÍQUIDA COM ESPECTROMETRIA DE MASSA ASSOCIADAS À QUIMIOMETRIA NA CLASSIFICAÇÃO E AVALIAÇÃO DE PERFIL LIPIDÔMICO DE CLASSES BACTERIANAS

Tese apresentada ao Programa de Pós-graduação em Química da Universidade Federal do Rio Grande do Norte, em cumprimento às exigências para obtenção do título de Doutora em Química.

Aprovada em: 18 de agosto de 2017.

Comissão Examinadora:

Dr. Kássio Michell Gomes de Lima – UFRN (orientador)

Dr. Fabricio Gava Menezes – UFRN

Dra. Maria Celeste Nunes de Melo – UFRN

Dra. Márcia Miguel Castro Ferreira – UNICAMP

Dr. Márcio José Coelho de Pontes – UFPB

Á Deus, por ter me confiado este caminho, mesmo eu não acreditando que seria possível.

Aos meus pais, Wilma e João, por terem estado ao meu lado por toda caminhada.

Com todo AMOR e GRATIDÃO,

Eu dedico.

AGRADECIMENTOS

Este trabalho foi o resultado de uma longa jornada pessoal e profissional, que só foi possível ser realizado com a ajuda de pessoas e instituições que se fizeram presentes ao longo desses anos, e me cabe reservar um espaço para agradecê-los.

Primeiramente, agradeço aos meus pais João e Wilma por terem sido fundamentais na minha formação, sempre priorizando a educação como uma pedra fundamental para a construção sólida do meu futuro.

Às minhas irmãs, que sempre acreditaram em mim e sempre me incentivaram.

Ao professor Kássio Lima, pela orientação e, acima de tudo, por confiar no meu potencial e me mostrar que é possível ir mais além.

Aos integrantes do grupo de química biológica e quimiometria (QBQ) por estarem sempre presentes, ajudando a dar sentido quando nada fazia sentido, em especial a Ana Carolina, Fernanda, Raquel, Jábine e Karine, que estiveram presentes em toda a caminhada.

Aos professores Celeste e Renato com seus orientandos, por fornecer amostras e acreditar no trabalho, dando todo apoio científico.

Ao grupo CHEMAGEB e o professor Romà Tauler, por terem me acolhido, me auxiliado e terem sido fundamentais em uma etapa muito importante da minha vida, por terem sido minha família durante todo um ano e me ajudado a dar um passo enorme de crescimento pessoal e profissional.

À CAPES pelo apoio financeiro que é de grande importância.

E a todos que, direta ou indiretamente, me ajudaram durante a jornada.

CURRICULUM VITAE

Formação acadêmica/ titulação

2013 – atual	Doutorado em Química – UFRN (Natal /RN)
2011 – 2013	Mestrado em Química – UFRN (Natal/RN)
2007 – 2010	Graduação em Química (Bacharel) – UFRN (Natal/RN)

Projetos de pesquisa

2016 – atual	MCR-ALS (*Multivariate Curve Resolution – Alternating Least Squares*) aplicado em imagens hiperspectrais obtidas por Raman para avaliar os efeitos de contaminação por Arsênio em espécies de cianobactérias Instituição de fomento: CAPES Coordenador: Kássio Michell Gomes de Lima IQ-UFRN (Natal/RN)
2014 – 2015	Utilização de um novo método de seleção de variáveis (ROI) em dados de LC/MS na análise lipidômica de duas espécies de cianobactérias. Instituição de fomento: CAPES

Coordenador: Kássio Michell Gomes de Lima IQ-UFRN (Natal/RN)

2011 – 2013 Detecção de bactérias multirresistentes utilizando a bioespectroscopia e análise multivariada como ferramenta diagnóstica.

Instituição de fomento: CAPES

Coordenador: Kássio Michell Gomes de Lima IQ-UFRN (Natal/RN)

Principais produções bibliográficas

.

MARQUES, ALINE S.; BEDIA, CARMEN; DE LIMA, KÁSSIO M. G.; TAULES, ROMÀ. Assessment of the effects of As(III) treatment on cyanobacteria lipidomic profiles by LC-MS and MCR-ALS. Analytical and Bioanalytical Chemistry, v. 408, p. 5829-5841, 2016.

MARQUES, ALINE S.; CASTRO, JENIELLY N. F.; COSTA, FAGNER J. M. D.; NETO, RENATO M.; DE LIMA, KÁSSIO M. G. Near- infrared spectroscopy and variable selection techniques to discrimitate Pseudomonas aeruginosa strais in clinical samples. Microchemical Journal, v. 124, p. 306-310, 2016.

MARQUES, ALINE S.; MORAES, EDGAR P.; JÚNIOR, MIGUEL A. A.; MOURA, ANDREW D.; NETO, WALTER F.; NETO, RENATO M.; DE LIMA, KÁSSIO M. G. Rapid discrimination of Klebsciella pneumonieae carbapenemase 2 – producing and non-producing Klebsciella peneumonieae strais using near-infrared spectroscopy (NIRS) and multivariate curve resolutions. Talanta, v. 134, p. 126-131, 2015.

MARQUES, ALINE S.; DE MELO, MARIA C. N.; CIDRAL, THIAGO A.; DE LIMA, KÁSSIO M. G. Feature selection stratategies for identification of Staphylococcus aureus recovered in blood cultures using FT-IR spectroscopy

successive projections algorithm for variable selection: a case study. Journal of Microbiological Methods, v. 98, p. 26-30, 2013.

MARQUES, ALINE S.; NICÁCIO, JÁBINE T. N.; CIDRAL, THIAGO A.; DE MELO, MARIA C. N.; DE LIMA, KÁSSIO, M. G. The use of near infrared spectroscopy and multivariate technique to differentiate Escherichia coli and Salmonella Enteritidis inoculated into pulp juice. Journal of Microbiological Methods, v. 93, p. 90-94, 2013.

RESUMO

ESPECTROSCOPIA E CROMATOGRAFIA LÍQUIDA COM ESPECTROMETRIA DE MASSA ASSOCIADAS À QUIMIOMETRIA NA CLASSIFICAÇÃO E AVALIAÇÃO DE PERFIL LIPIDÔMICO DE CLASSES BACTERIANAS.

Esta tese de doutorado é um aporte teórico-prática para o desenvolvimento de estudos que utilizem a bioanalítica, particulamente materiais biológicos provenientes de bactérias, podendo estes ser isolados, DNA, entre outros, em conjunto com ferramentas quimiométicas de análise. Para isso, buscou-se identificar diferenças bacterianas quando submetidas a uma fonte de estresse a partir de diferentes técnicas analíticas. A primeira abordagem foi realizada partindo da bioespectroscopia, utilizando-se de dados espectroscópicos obtidos na região do infravermelho. A bioespectroscopia na região do infravermelho é descrita como uma técnica não invasiva, de alto rendimento, baixo custo (quando comparado com técnica padrões de análise) e objetivas, e que possui um enorme potencial na análise de bactérias, complementando ou mesmo substituindo métodos de diagnóstico de doenças convencionalmente conduzidos por especialistas através de métodos padrões de análises de alto custo e que necessitam de reagentes específicos. Os dados obtidos a partir da bioespectroscopia em amostras bacterianas são complexos e apresentam muitas bandas de sobreposição sendo necessária a aplicação de ferramentas matemáticas para superar estas dificuldades. Para isso, algumas ferramentas matemáticas, como os métodos de seleção de variáveis, que utilizam a análise discriminante linear com Algoritmo de Projeção Sucessiva (SPA-LDA) e Algoritmo Genético (GA-LDA), geralmente são utilizadas com a finalidade de facilitando a extração de informações relevantes. A espectroscopia na região do infravermelho, em específico infravermelho próximo (NIR) e infravermelho com trasformata de Fourier e reflectância total atenuada (ATR-FTIR), em conjunto com métodos de seleção de variáveis (SPA-LDA e GA-LDA) foram utilizadas na discriminação de amostras de bactérias (*Sthaphylococcus aureus*, *Klebsiella pneumoneae* e *Pseudomonas aeruginosa*). Foram identificados prováveis biomarcadores como lipídeos e proteínas em ~1550 cm-1 e 1400 cm-1 e vibrações de DNA em ~1080 cm-1. Valores de sensibilidade de 75% e 95% para modelos de SPA-LDA e 100% e 93% para modelos GA-LDA foram encontrados. Com base nesses resultados, pode-se concluir que o SPA-LDA e GA-LDA em conjunto com a espectroscopia na região do infravermelho mostraram-se ferramentas eficientes melhorando o tempo e custo de diagnóstico possibilitando o tratamento mais rápido em relação aos métodos padrões de diagnóstico e, consequentemente, sendo possível evitar a evolução de uma possível infecção. A segunda abordagem foi avaliar possíveis mudanças no perfil lipidômico de bactérias resultante de sua exposição a uma fonte de estresse externa (Arsênio (III)), utilizando as cianobactérias *Anabaena* sp. e *Planktothrix agardhii*. Os dados foram obtidos a partir a Cromatografia Líquida- Espectrometria de Massas (LC-MS) que por gerar uma matriz de dados muito extensa foi necessária a utilização de uma estratégia de seleção proposta recentemente, definida como ROI (do inglês *regions of interests*) que diminui significativamente o tamanho da matriz de dados obtidas por LC-MS. Resolução Multivariada de Curvas com Mínimos Quadrados

Alternantes (MCR-ALS) foi utilizado como método de resolução das fontes de variação, recuperando as informações de seus componentes puros que se encontravam misturadas. As massas majoritárias encontradas, sendo algumas delas 766.54, 565.40 e 871.56 (m/z), determinam que as cianobactérias estudadas, ao serem submetidas a As(III), sofrem mudanças relacionadas a estruturas que compõem os processos fotossintéticos das mesmas.

Palavras-chave: Bioespectroscopia. SPA-LDA. GA-LDA. NIR. ATR-FTIR. ROI. MCR-ALS. LC-MS.

ABSTRACT

SPECTROSCOPY AND LIQUID CHROMATOGRAPHY WITH SPECTROMETRY OF MASS ASSOCIATED TO CHEMOMETRY IN THE CLASSIFICATION AND EVALUATION OF LIPIDOMIC PROFILE OF BACTERIAL CLASSES.

This doctoral thesis is a theoretical-practical contribution for the development of studies that use bioanalytical, particularly biological materials from bacteria, which can be isolated, DNA, among others, in conjunction with chemistry analysis tools. For this, it was sought to identify bacterial differences when submitted to a source of stress from different analytical techniques. The first approach was based on biospectroscopy, using spectroscopic data obtained in the infrared region. Biospectroscopy in the infrared region is described as a non-invasive, high-throughput, low-cost (when compared with standard analytical techniques) and objective techniques, and has a huge potential in the analysis of bacteria, complementing or even replacing diagnostic methods of diseases conventionally conducted by skilled persons by standard methods of expensive analyzes and requiring specific reagents. The data obtained from biospectroscopy in bacterial samples are complex and have many overlapping bands and it is necessary to apply mathematical tools to overcome these difficulties. For this, some mathematical tools, such as variable selection methods, using Linear Discriminant Analysis with Successive Projection Algorithm (SPA-LDA) and Genetic Algorithm (GA-LDA), are generally used for the purpose of solving these data, facilitating the extraction of information. Infrared spectroscopy, in specific Near Infrared (NIR) and infrared spectroscopy with Fourier transform and Attenuated Total Reflectance (ATR- FTIR), in conjunction with variable selection methods (SPA-LDA and GA-LDA) was used in the discrimination of bacterial samples (*Sthaphylococcus aureus*, *Klebsiella pneumoneae* and *Pseudomonas aeruginosa*). Possible biomarkers such as lipids and proteins were identified at ~ 1550 cm -1 and 1400 cm -1 and DNA vibrations at ~ 1080 cm -1. Sensitivity values of 75% and 95% for SPA-LDA models and 100% and 93% for GA-LDA models were found. Based on these results, it can be concluded that the SPA-LDA and GA- LDA in conjunction with the infrared spectroscopy showed efficient tools improving the time and cost of diagnosis allowing the treatment faster than the standard methods of diagnosis, and consequently, it is possible to avoid the evolution of a possible infection**.** The second approach was to evaluate possible changes in the lipid profile of bacteria resulting from its exposure to an external stress source (Arsenic (III)), using the cyanobacteria *Anabaena sp.* and *Planktothrix agardhii*. The data were obtained from Liquid Chromatography-Mass Spectrometry (LC-MS), which, in order to generate a very extensive data matrix, required the use of a recent selection strategy, defined as ROI (regions of interest), which significantly decreased the Size of the data matrix obtained by LC-MS. Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS) was used as a method to solve variation sources, retrieving the information of its pure components that were mixed. The majority masses found, such as 766.54, 565.40 and 871.56 (m/z), determine that the studied cyanobacteria, when subjected to As (III), undergo

changes related to structures that make up the photosynthetic processes of the same.

Keywords: Bioespectroscopy. SPA-LDA. GA-LDA. NIR. ATR-FTIR. ROI. MCR-ALS. LC-MS

ABREVIAÇÕES

ALS	Assimetric Least Squares
ATR	Atenuated total reflectance
DNA	Desoxyribonucleic acid
FT	Fourier-transform
FTIR	Fourier-transform infrared
GA	Genetic algoritmo
IR	Infrared
KP	*Klebsiela pneumonieae*
KPC	*Klebsiela pneumonieae* Carbapenemase
LC-MS	Líquid chromatography-mass spectrometry
LDA	Linear discriminant analysis
MCR	Multivariate curve resolution
MS	Mass spectrometry
MSC	Multiplicative signal correction
NIR	Near infrared
PCA	Principal componente analysis
ROI	Regions of interests

PREFÁCIO

Esta Tese foi desenvolvida a partir da parceria entre Instituto de Química (IQ), Departamento de Microbiologia e Parasitologia (DMP), Laboratório de Micobactéria (LABMIC) da Universidade Federal do Rio Grande do Norte (UFRN – Natal/RN) e com o grupo de pesquisa CHEMAGEB (Chemometric and high-throughput omic analytical methods for assessment of global change effects on envinronmental and biological systems) do Consejo Superior de Investigaciones Científicas (CSIC – Barcelona/ES).

Este estudo teve os seguintes colaboradores: Prof. Kássio Michell G. Lima (IQ / UFRN) sendo o coordenador do estudo e fornecendo orientações relacionadas à análise computacional e experimental; Profª. Maria Celeste Nunes de Melo e Thiago André Cidral (DMP/UFRN) fornecendo amostras biológicas, disponibilizando espaço para realização de experimentos e orientação científica; Prof. Renato Mota Neto e seus alunos Jennielly Castro e Fagner Costa (LABMIC/UFRN) realizando preparo de amostras e auxiliando na obtenção espectral, bem como fundamentação científica; O professor Romà Tauler e Camen Bedia (CHEMAGEB/CSIC) auxiliando na fundamentação teórica e prática de análises computacionais e experimentais, bem como no desenvolvimento de novos algoritmos; E eu, Aline Marques, preparei as amostras, cultivando, replicando, extraindo as amostras para a análise espectral e cromatográfica, fiz a aquisição dos dados, o pré-processamento, construí os modelos quimiométricos e escrevi os manuscritos, este último com a contribuição dos co-autores. Abaixo está listada uma breve biografia acadêmica dos autores e co-autores.

Aline de Sousa Marques. Concluiu o Bacharelado em Química na Universidade Federal do Rio Grande do Norte (UFRN/Brasil) em 2010. Recebeu o título de mestre também pela UFRN em 2013. Atualmente ela é estudante de doutorado no grupo de Química Biológica e Quimiometria (QBQ/UFRN), trabalhando com técnicas espectroscópicas e cromatográficas, aplicando métodos de seleção e classificação em estudos bacterianos. (http://lattes.cnpq.br/0793041287437113)

Kássio Michel Gomes de Lima. É um professor adjunto de química analítica em Instituto de Química da Universidade Federal do Rio Grande do Norte (UFRN/Brasil). Recebeu o título de doutorado em ciências na Universidade Estadual de Campinas (UNICAMP/Brasil) em 2007 e foi pós-doutorado no período de Julho 2013-Julho 2014 (CSIC/Barcelona, Espanha) no Grupo CHEMAGEB, liderado pelo Dr. RomaTauler. Ele também trabalhou como pesquisador visitante na Universidade de Lancaster (2014), Center for Biophotonics, liderado pelo Prof. Francis L. Martin. Ele é o coordenador do grupo de pesquisa Química Biológica e Quimiometria (QBQ/UFRN) desde 2011 e seus atuais interesses de pesquisa estão de técnicas multivariadas de calibração e classificação em sistemas biológicos. (http://lattes.cnpq.br/6928918856031880)

Maria Celeste Nunes de Melo. Mestre e doutora pela Universidade Federal do Rio de Janeiro (UFRJ/Brasil). Professora associada em ciências (microbiologia) do Departamento de Microbiologia e Parasitologia (DMP/UFRN). Sua principal área de atuação é a microbiologia médica. Está envolvida em projetos relacionados a virulência e resistência bacteriana. (http://lattes.cnpq.br/0580551464788795)

Thiago André Cidral. Possui graduação em Biomedicina e Mestrado em Biologia Parasitária pela Universidade Federal do Rio Grande do Norte (UFRN/Brasil). Atualmente é Doutorando em Ciências da Saúde pela Universidade Federal do Rio Grande do Norte (UFRN/Brasil) e tem experiência na área de Microbiologia, com ênfase em Bacteriologia, atuando principalmente nos seguintes temas: resistência bacteriana, *Staphylococcus coagulase* negativo, enterococcus, hemocultura e multirresistência aos antimicrobianos. (http://lattes.cnpq.br/9503814899080091)

Renato Motta Neto. Farmacêutico – bioquímico analista, mestre em patologia e doutor em cirurgia pela Universidade Federal do Ceará (UFCE/Brasil). É professor do Departamento de Microbiologia e Parasitologia (DMP/UFRN), atuante nas áreas de diagnóstico e epidemiologia das infecções bacterianas e o estudo genotípico e fenotípico do gênero Mycobacterium. (http://lattes.cnpq.br/6909091962347443)

Jenielly de Noronha Ferreira de Castro. Possui graduação em Biomedicina e especialização em Microbiologia Clínica e laboratorial. É mestre em Ciências Biológicas com ênfase em Interações microbiológicas e parasitárias, atualmente cursa Doutorado em Bioquímica. Tem experiência na área de microbiologia, com ênfase em bacteriologia médica. (http://lattes.cnpq.br/0987509634213465)

Fagner James Martins Dantas Costa. Biomédico e aluno do Programa de Pós-Graduação em Ciências Biológicas (PPGCB) na categoria de Mestrado Acadêmico na Universidade Federal do Rio Grande do Norte (UFRN/Brasil). Atualmente faz parte do grupo de pesquisas do Laboratório de Micobactérias (LABMIC), atuando na área de Bacteriologia Clínica, com ênfase na Epidemiologia de Bactérias Gram-negativas Multirresistentes. (http://lattes.cnpq.br/984566792410946)

Romà Tauler. Químico, mestre e doutor em Química Analítica pela Universidade de Barcelona (UB/Espanha), professor pesquisador no Departamento de Química Ambiental no *Institute of Environmental Assessment and Water Research* (IDAEA/CSIC/Espanha). Ganhou o prêmio de melhor quimiometrista do ano em 2009 no *Eastern Analytical Symposium* (USA). Editor chefe do jornal *Chemometrics and International Laboratory Systems*. Atua como quimiometrista desenvolvendo métodos multivariados de análises, em especial métodos resolução multivariada de curvas (MCR) e ciências analíticas e Omics.

Carmen Bedia. Farmacêutica pela Universidade de Barcelona (UB/Espanha) e, atualmente, pós-doutoranda no *Institute of Environmental Assessment and Water Research* (IDAEA/CSIC/Espanha) especialista em metabolômica, lipidômica e culturas de células.

SUMÁRIO

CAPÍTULO 1 INTRODUÇÃO GERAL..... 19

CAPÍTULO 2 Feature selection strategies for identification of *Staphylococcus aureus* recovered in blood cultures using FT-IR spectroscopy successive projections algorithm for variable selection: A case study.

Aline de Sousa Marques, Maria Celeste Nunes de Melo**,** Thiago André Cidral**,** Kássio Michell Gomes de Lima,

Journal of Microbiological Methods, 2014..... 44

CAPÍTULO 3 Rapid discrimination of *Klebsiella pneumoniae* carbapenemase 2 – producing and non-producing *Klebsiella pneumoniae* strains using near-infrared spectroscopy (NIRS) and multivariate analysis.

Aline S. Marques, Edgar Perim Moraes**,** Miguel A. A. Júnior**,** Andrew D. Moura**,** Valter F. A. Neto**,** Renato M. Neto**,** Kássio M. G. Lima.

Talanta, 2014..... 49

CAPÍTULO 4 Near-infrared spectroscopy and variable selection techniques to discriminate *Pseudomonas aeruginosa* strains in clinical samples.

Aline S. Marques, Jenielly F. Castro**,** Fagner J. M. D. Costa, Renato M. Neto**,** Kássio M. G. Lima.

Microchemical Journal, 2016..... 56

CAPÍTULO 5 Assessment of the effects of As(III) treatment on cyanobacteria lipidomic profiles by LC-MS and MCR-ALS.

Aline S. Marques, Carmen Bedia**,** Romà Tauler **,** Kássio

M. G. Lima.

Analytica Bioanalytica Chemistry, 2016..... 62

CAPÍTULO 6 CONCLUSÕES E PERSPECTIVAS..... 76

APÊNDICE A Hyperspectral Raman spectroscopy image analysis of Cyanobacteria exposed to Arsenic(III) using multivariate curve resolution-alternating least squares (MCR-ALS).

Aline S. Marques, Carmen Bedia, Roma Tauler, Kássio M. G. Lima

Manuscrito em fase de correção..... 79

CAPÍTULO 1 – INTRODUÇÃO GERAL

1. INTRODUÇÃO..... 19

2. PRINCIPAIS OBJETIVOS..... 31

3. COMPOSIÇÃO DA TESE..... 32

4. METODOLOGIA..... 33

REFERÊNCIAS..... 36

1. INTRODUÇÃO

Esta tese de doutorado é um aporte teórico-prático para o desenvolvimento de estudos que utilizem a bioanalítica, particulamente materiais biológicos provenientes de bactérias, podendo estes ser isolados, DNA, entre outros, em conjunto com ferramentas quimiométicas de análise. Os trabalhos desenvolvidos incluem duas técnicas analíticas, a espectroscopia na região do infravermelho e a cromatografia líquida com espectrometria de massa.

A primeira técnica utilizada foi a espectroscopia na região do infravermelho (IR), especificamente no infravermelho médio com reflexão total atenuada (ATR- FTIR) e infravermelho próximo (NIR) em conjunto com as técnicas de seleção de variáveis: algoritmo de projeção sucessiva e algoritmo genético, ambos com análise discriminante linear (SPA-LDA e GA-LDA), para avaliar a eficiência destes na distinção de amostras de bactérias patogênicas sensíveis e resistentes a antibióticos em sangue, isolado bacteriano e em seus constituintes (DNA).

A segunda técnica analítica utilizada foi a cromatografia líquida acoplada com espectrometria de massa, utilizando uma estratégia de seleção de variáveis (ROI) e resolução multivariada de curvas com mínimos quadrados alternantes (MCR-ALS) para determinar variações no perfil lipidômico em

amostras de bactérias ambientais quando submetidas ao estresse de um agente externo.

1.1. BACTÉRIAS PATOGÊNICAS E RESISTÊNCIA BACTERIANA

Definem-se patogênicas, aquelas bactérias que são passíveis a causar doenças infecciosas, tais como tuberculose, pneumonia, meningite, dentre outras. [1] De acordo com classificação da Anvisa de 2016, estão entre as 10 bactérias mais comuns causadoras de infecções sanguíneas em UTIs brasileiras a *Klebsciela pneumoniae*, *Staphylococcus aureus* e *Pseudomonas aeruginosa*. [2] Infecções bacterianas representam uma importante causa na morbidade e mortalidade humana. [2] Estas morbidade e mortalidade aumentam significativamente quando as bactérias causadoras destas infecções desenvolvem algum tipo de resistência. [4-7] Segundo a lista publicada pela Organização Mundial de Saúde (OMS), onde é foi listado os "patógenos prioritários" resistentes a antibióticos para os quais é urgente o desenvolvimento de novos antibióticos. Essa lista está dividida em categorias de prioridade crítica, alta e média para destacar o nível de urgência da necessidade de novos antibióticos. A "crítica" inclui bactérias com resistência a múltiplas drogas, elas incluem *Acinetobacter*, *Pseudomonas*, e diversas *Enterobacteriaceae* (incluindo *Klebsiella, E. coli, Serratia, Proteus*), que podem causar infecções graves e frequentemente letais, como infecções sanguíneas e pneumonia. Essas bactérias se tornaram resistentes a um grande número de antibióticos, incluindo carbapenêmicos e cefalosporinas de terceira geração. A segunda e terceira partes da lista, a de prioridade alta e média, contêm outras bactérias que estão aumentando a resistência aos antibióticos existentes e que causam doenças mais comuns, como gonorreia e intoxicação alimentar.[8]

Vários métodos de diagnóstico de microrganismos estão descritos na literatura e em protocolos oficiais. [7] Entretanto, muitos deles, apresentam ineficiência quanto a rapidez e precisão [7,9] e, também,

envolvem muitas etapas de preparação e pré-enriquecimento. [10,11] Com o objetivo de melhorar o tempo de diagnóstico e início de uma terapia eficiente e, consequentemente, evitar a evolução de uma possível infecção, métodos mais rápidos e precisos estão sendo a desenvolvidos. [12]

Como uma alternativa rápida, as técnicas espectroscópicas têm sido utilizadas. [13,14] A aplicação da espectroscopia IR na análise de microrganismos tem sido um alvo considerável nos últimos anos. [15] ATR- FTIR e NIR, quando associadas à um tratamento estatístico, podem ser utilizadas para determinar a impressão digital de metabólitos em microrganismos, tais como DNA, RNA, proteínas e compostos da parede celular. [16,17]

1.2. ESPECTROSCOPIA

1.2.1. Espectroscopia no infravermelho

Os métodos espectroscópicos são baseados nas interações da radiação com a matéria para obter informações sobre uma amostra. O analito encontra-se, inicialmente, em seu estado fundamental e, a o receber a radiação, algumas espécies do analito sofrem uma transição para um estado excitado podendo essa transição ser eletrônica, vibracional e rotacional. [18] Cada espécie molecular possui a capacidade de absorver frequências características. [19]

A espectroscopia na região do infravermelho engloba, em sua maior parte, às transições vibracionais das moléculas. Isto porque a radiação, geralmente, não é suficiente para causar transições eletrônicas. [20] Ela possui uma ampla faixa espectral que é dividida em Infravermelho Próximo (NIR, 780 – 2500 nm), Infravermelho Médio (MIR, 2500 – 5x104 nm) e Infravermelho Distante (FIR, 5x104 – 1x106 nm). [21] As vibrações moleculares podem se incluir em estiramentos ou deformações angulares.

A vibração de estiramento ocorre com uma variação constante na distância entre átomos ao longo do eixo da ligação, baseado na lei de Hooke [22], enquanto que a vibração de deformação angular é caracterizada pela variação no ângulo entre duas ligações.[23]

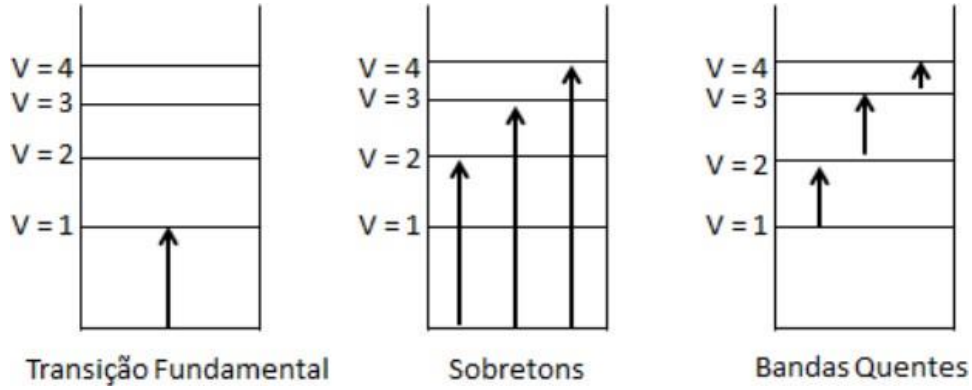
É possível explicar as bandas de absorção observadas no infravermelho geradas a partir dos modos fundamentais de vibração molecular, porém não consegue prever a presença de forças repulsivas entre átomos, a dissociação quando o comprimento da ligação excede a sua extensão e a presença de sobretons nos espectros, como por exemplo na região do infravermelho próximo. [24]

1.2.1.1. NIR

A região da espectroscopia no infravermelho próximo, denominada NIR, origina espectros que apresentam sobreposições e bandas de combinação das ligações CH, NH e OH de vários grupos funcionais [25], podendo outras ligações, como CC, CO, CN e PO, contribuir para o espectro NIR, embora possuam sinal fraco quando comparados com as anteriormente citadas, resultando em uma contribuição pouco significativa. [26]

O NIR baseia-se na absorção de energia por parte das ligações existentes nas moléculas de uma amostra que são causadas por três mecanismos diferentes: sobreposição de vibrações fundamentais, combinação de vibrações fundamentais e ainda absorções eletrônicas, gerando bandas de baixa seletividade [18,27]. Essas bandas são originadas de transições com $\Delta v=\pm 2$, $\Delta v=\pm 3$... (ver referência 28), não obedecendo à regra de seleção, como pode ser observado na Figura 1.

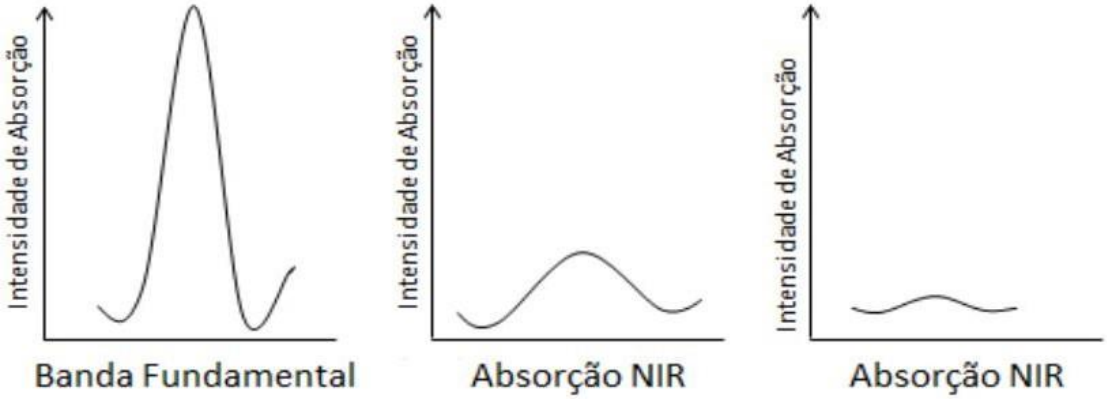
Figura 1 – Transição entre níveis vibracionais.



(Fonte: MARQUES, Aline S., 2013)

As bandas de absorção do NIR são largas, sobrepostas e mais fracas que suas bandas fundamentais, como pode ser visto na Figura 2. Apesar da baixa absortividade, é possível uma maior profundidade de penetração, permitindo análises diretas de fortes absorventes e nivelando amostras com espalhamento, favorecendo a realização de análises químicas e físicas em uma única medida. [18]

Figura 2- Representação de bandas de absorção na região do infravermelho próximo.



(Fonte: MARQUES, Aline S. 2013)

1.2.1.2. ATR-FTIR

1.2.1.2.1. Reflectância total atenuada - ATR

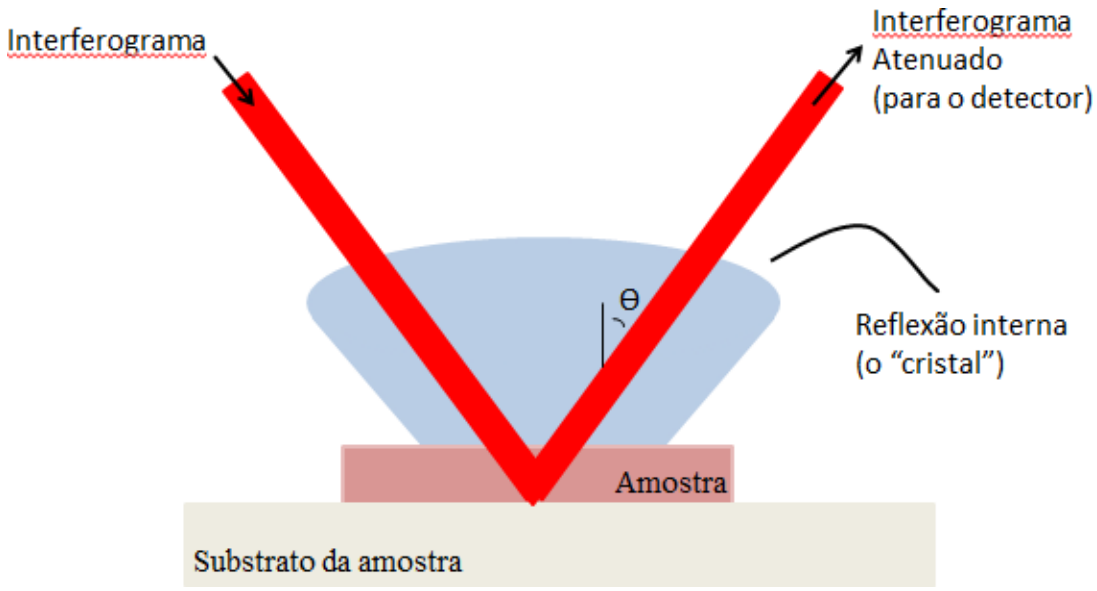
A espectroscopia de Reflexão Interna ou Refletância Total Atenuada (ATR) possui vantagens diante de outros modos da espectroscopia no infravermelho, como exemplo os modos de transmissão e transflecção [29], e requer pouco ou nenhum preparo de amostra.

O princípio do ATR baseia-se na reflexão de quando o feixe de radiação passa do cristal ATR para a amostra. Como ilustrado na Figura 3, a fração do feixe de luz incidente que é refletida é proporcional ao aumento do ângulo de incidência, atingindo uma reflexão completa quando excede um determinado ângulo crítico. No ponto de reflexão o feixe atua penetrando a uma pequena distância dentro da amostra. A profundidade dessa penetração é dada pela equação 1:

$$d(\lambda) = \frac{\lambda}{2\pi n\ (\sin 2\theta - (\frac{n_2}{1})^2 \frac{1}{n_1}))} \quad (1)$$

Onde d(λ) é a profundidade da penetração, λ é o comprimento de onda, n_1 o índice de refração (cristal ATR), n_2 o índice de refração (amostra) e θ é o ângulo de incidência.[29,30]

Figura 3- Ilustração da espectroscopia de reflexão total atenuada (ATR).



(Fonte: adaptado da referência 31)

1.2.1.2.2. Espectroscopia FTIR

A espectroscopia FTIR é um método realizado com espectrômetros de infravermelho mais modernos, e trabalham sob um princípio diferente.

O caminho ótico produz um padrão em forma de ondas que contém todas as frequências que formam o espectro do infravermelho, esse padrão é chamado de interferograma, que é, em sua essência, um gráfico de intensidade versus tempo. [31] Esse gráfico é, então, convertido em um gráfico de intensidade versus frequência com uma operação matemática conhecida como transformata de Fourier (FT). [32]

A vantagem de utilizar a transformata de Fourier com o infravermelho é que é possível obter dezenas de interferogramas para uma mesma amostras em menos de um segundo, realizar o procedimento matemático com a soma

dos interferogramas guardados e gerar um espectro com melhor razão sinal/ruído.[33]

Os dados obtidos a partir da espectroscopia ATR-FTIR e NIR com amostras biológicas são complexos e apresentam muitas bandas de sobreposição sendo necessária a aplicação de ferramentas matemáticas para sua resolução. Entretanto, ao utilizarmos o espectro todo em um modelo matemático, algumas variáveis são redundantes e não contém informação efetiva alterando o desempenho final do modelo. [34] Para isso, os métodos de seleção de variáveis, SPA-LDA e GA-LDA [35,36], podem ser utilizados, reduzindo o conjunto de dados a variáveis predominantes.

1.3. MÉTODOS LINEARES DE CONSTRUÇÃO DE CARACTERÍSTICAS

Os métodos lineares de construção de características são definidos como um conjunto de ferramentas estatísticas que consiste na análise simultânea de múltiplas variáveis com o objetivo de avaliar a interrelação entre elas encontrando características em comum.[37]

1.3.1. Análise de componentes principais (PCA)

PCA é um método não supervisionado que transforma ortogonalmente um conjunto de observação em um conjunto de valores, chamados componentes principais. A matriz original é decomposta como o produto de duas outras matrizes, os escores e os loadings, como na Equação 2.

$$X = TP^t + \mathrm{E} \qquad (2)$$

Onde X é a matriz de dados (IxJ), T a matriz de escores (IxA), P a matriz de loadings (JxA) e E a matriz residual, com j sendo o numero de variáveis, i o numero de objetos ou amostras e a número de componentes que contém as informações significativas retidas.[38]

O PCA é comumente utilizado com o objetivo de reduzir o número de variáveis, sem perder a informação original. [18]

1.3.2. Análise discriminante linear (LDA)

A análise discriminante linear, diferentemente do PCA, é um método supervisionado largamente utilizado para classificar um grupo em classes pré-definidas. Foi proposta, inicialmente por Fisher [39], tem o objetivo de resolver problemas de diferenciação entre dois ou mais grupos, visando a posterior classificação destes. Um vetor é construído com a informação de cada grupo, identificando os grupos com valores de 0 a N por grupo. [40]

1.3.3. Métodos de seleção de variáveis

Os métodos de seleção de variáveis têm sido amplamente utilizados em conjunto com métodos de classificação. Destacam-se, entre elas o algoritmo de projeções sucessivas (SPA) [41] e o algoritmo genético (GA). [42]

1.3.3.1. SPA-LDA

O SPA é utilizado na busca de subconjuntos de variáveis que tenhas informações redundantes, diminuindo a colinearidade dos dados. Ele inicia com uma variável e incorpora outra variável a cada interação, até que a quantidade de número de variáveis desejadas seja atingida. São geradas cadeias de variáveis e avaliadas por meio de uma equação de risco (G). O subconjunto de variáveis que apresentar menor valor de G será selecionado. [38]

1.3.3.2. GA-LDA

O GA trata-se de uma técnica de otimização que baseia-se no princípio da biologia da evolução e seleção natural. Correlaciona-se a codificação de variáveis de forma binária, por meio de um gerador aleatório, onde cada

variável recebe o valor 0 ou 1, significando, respectivamente, não incluído ou incluído no modelo, gerando um conjunto de variáveis que não incorpora tendências externas. [43,45]

1.4. CIANOBACTÉRIAS E CONTAMINAÇÃO AMBIENTAL

As cianobactérias são os únicos procariontes já estudados que possuem a capacidade de realizar fotossíntese. Ela é encontrada em larga escala em diversos ambientes aquáticos, mas também podem existir em habitats terrestres. [46] É atribuído a elas 50% da fotossíntese terrestre, sendo um dos principais contribuintes do ciclo do oxigênio global. [47]

Quando submetidas a fontes de contaminação, como exemplo metais pesados, elas podem sofrer alterações significativas em seu processo fotossintético, [48] que ocorre em sua na membrana tilacóide que possui lipídeos como componentes estruturais, majoritariamente algumas espécies de galactolipideos. [49]

1.5. CROMATOGAFIA LÍQUIDA E ESPECTROMETRIA DE MASSA

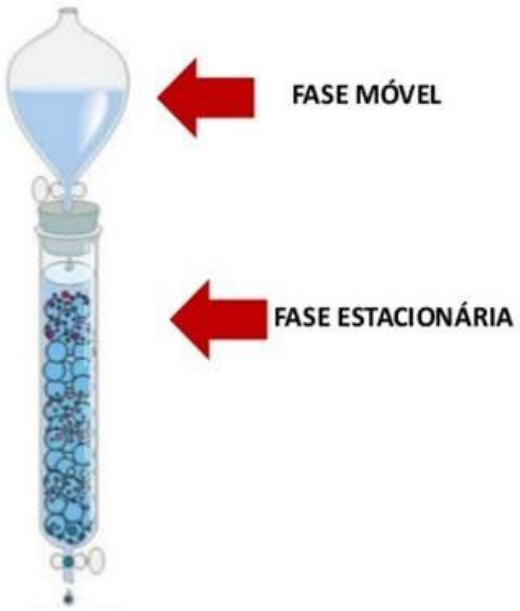
1.5.1. Cromatografia líquida acoplada com espectrometria de massa

A cromatografia líquida, atualmente conhecida como HPLC (High Performance Liquid Chromatography) é um método analítico que objetiva a separação de diferentes espécies químicas presentes em uma amostra. Essa separação ocorre por meio de uma interação seletiva entre as moléculas da amostra (solvente) e as fases estacionária e móvel. A fase estacionária está relacionada com a coluna cromatográfica (Figura 4), que possui em sua composição pequenas partículas. A fase móvel (solvente), flui continuamente pelo sistema arrastando a amostra injetada através da coluna até o detector.[50]

A separação fundamenta-se no fato de que, as substâncias presentes na amostra, devido as suas diferentes estruturas moleculares, dispões e distintos graus de afinidade com a coluna e o solvente apresentando suas velocidades de migração igualmente diferentes. As substância com maior afinidade com a fase estacionária é aquela que elui por último, sendo o oposto válido.[51]

Essa técnica é comumente utilizada para separar, identificar e quantificar substâncias em diferentes tipos de produtos.

Figura 4. Esquema de coluna cromatográfica



(Fonte: Adaptado da referência 51)

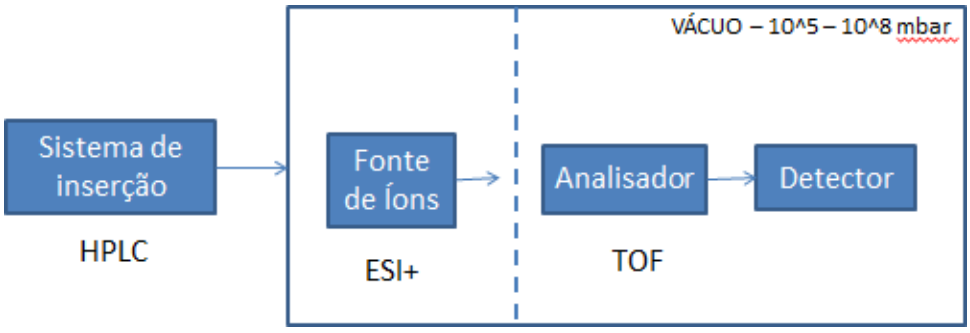
A Espectrometria de Massas (MS) é definida como o estudo da matéria pela formação de íons em fase gasosa e posteriormente caracterizados por um Espectrômetro de Massas de acordo com sua massa, carga, estrutura ou propriedades fisico-químicas.[52] O resultado de uma análise por MS se dá pela forma de um espectro, onde a abscissa corresponde à razão entre a

massa e o número de cargas do íon (m/z) e a ordenada está relacionada à sua intensidade. [53]

A cromatografia líquida quando acoplada com a espectrometria de massa combina o poder de separação da HPLC com a capacidade da espectrometria de massa (MS) de detectar de forma seletiva e confirmar a identidade de uma molécula. LC-MS usa um sistema HPLC e à medida que as fases líquidas móveis saem da coluna, a amostra líquida é pulverizada para produzir microgotas, que evaporam rapidamente, liberando as moléculas de analito ionizadas que podem, então, ser separadas via MS. [54]

A Figura 5, representa o esquema de um espectrômetro de massas para a obtenção de dados LC-MS. Em abordagens ômicas, é especialmente usado um sistema de ionização eletrospray (ESI) com um analisador de massa (TOF), assim as moléculas que chegam à sofrem um processo de ionização suave chamado ionização por electrospray [55,56] que, em geral, não quebra as moléculas da amostra, apenas as ioniza, permitindo a criação de agregados com ganho ou perda de átomos ou moléculas, se este agregado apresentar massa maior que a molécula original é chamado de aducto e se menor, fragmento.[57]

Figura 5. Esquema de um espectrômetro de massas



(Fonte: adaptado da referência 54)

1.5.2. MCR-ALS

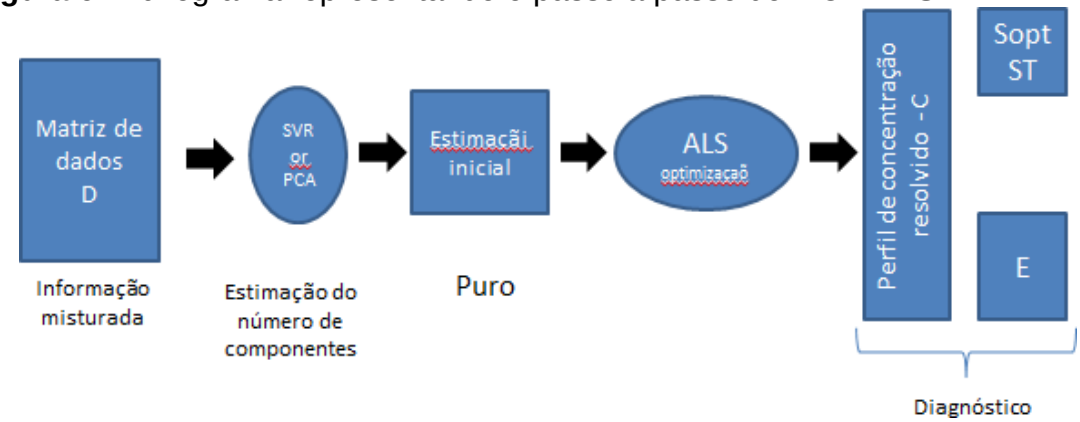
As análises realizadas por LC-MS geram uma grande quantidade de dados para uma única amostra biológica, para isso estratégias de seleção de massas (ROI) pode ser utilizado para reduzir significativamente o tamanho do conjunto de dados e posteriormente terem suas informações isoladas e resolvidas pelo método de resolução multivariada de curvas (MCR-ALS).

O método de Resolução Multivariada de Curvas (MCR, do inglês *Multivariate Curve Resolution*) é um método de processamento de sinais analíticos que tem o objetivo de isolamento, resolução e quantificação das fontes de variação em um determinado conjunto de dados. [58,59] Ele recupera informações misturadas, não supervisionadas, ou seja, não é necessário informar nenhuma hipótese sobre a contribuição de diferentes fatores, provenientes de um instrumento (D) em contribuições reais de seus componente puros, representados pelos perfis de concentração (C) e perfis espectrais (ST). [60]

A equação que define o modelo geral do MCR é dado por D = CST, em que D é a matriz da resposta instrumental, C é a matriz de concentração e S a matriz de espectros puros.[61]

O método MCR com Mínimos Quadrados Alternantes (ALS, do inglês *Alternation Leasts Squares*) é utilizado para buscar o resultado que apresenta o melhor ajuste, chamado de otimização, e permite recuperar perfis de concentração individuais e sinais (espectros, imagens) de espécies que explicam a variância observada dos dados. [60,61]

Figura 6. Fluxograma representando o passo a passo do MCR-ALS



2. PRINCIPAIS OBJETIVOS

- Fornecer um complemento potencial e/ou método alternativo para detecção e classificação de bactérias patogênicas (*Staphylococcus aureus*, *Klebsiella pneumonieae* e *Pseudomonas aeruginosas)* aos métodos tradicionais.

- Desenvolver um método para detectar mudanças no perfil lipidômico de cianobactérias (*Anabaena* sp*.* e *Planktothrix agardhii*) quando expostas a um agente contaminante (AS(III)), podendo ser utilizado como indicativo de contaminação ambiental.

- Avaliar a eficiência dos métodos de seleção de variáveis SPA-LDA e GA-LDA em estudos biológicos partindo de dados obtidos por espectroscopia no infravermelho:
    - Na detecção de diferentes concentrações da bactéria *Staphylococcus aureus* contidas em amostras de sangue.
    - Na diferenciação de amostras de DNA de *Klebsiella pneumonieae* produtora e não produtora de Carbapenemase.
    - Na discriminação de amostras resistentes e sensíveis de isolados bacterianos de *Pseudomonas aeruginosas* a partir de amostras clínicas.

3. COMPOSIÇÃO DA TESE

A presente tese foi organizada por ordem cronológica de desenvolvimento e está composta de publicações em que participei como primeira autora e que foram essenciais para o desenvolvimento dos meus argumentos. Esta está organizada em capítulos como segue:

Capítulo 1 é composto pela introdução geral.

Capítulo 2 – *Feature selection strategies for identification of Staphylococcus aureus recovered in blood cultures using FT-IR spectroscopy successive projections algorithm for variable selection: A case study* – este trabalho introduz o uso dos métodos de seleção de variáveis SPA-LDA e GA- LDA a dados obtidos por espectroscopia no infravermelho (FT-IR) a partir de amostras de sangue contaminadas com *Staphylococcus aureus,* em diferentes concentrações, para avaliar a eficiência destes métodos e encontrar marcadores biológicos associados a estas diferenças.

Capítulo 3 – *Rapid discrimination of Klebsiella pneumoniae carbapenemase 2 – producing and non-producing Klebsiella pneumoniae strains using near-infrared spectroscopy (NIRS) and multivariate analysis* – traz os métodos de seleção de variáveis SPA-LDA e GA-LDA associados à espectroscopia no infravermelho próximo na diferenciação de amostras de DNA de cepas de *Klebsiella pneumoniae* produtora e não produtora de carbapenemase 2.

Capítulo 4 – *Near-infrared spectroscopy and variable selection techniques to discriminate Pseudomonas aeruginosa strains in clinical samples*
– aplica os métodos, SPA-LDA e GA-LDA com espectroscopia no infravermelho próximo, para avaliar sua eficiência na diferenciação de amostras clínicas de *Pseudomonas aeruginosa*.

Capítulo 5 – *Assessment of the effects of As(III) treatment on cyanobacteria lipidomic profiles by LC-MS and MCR-ALS* – trabalho realizado em parceria com o Consejo Superior de Investigaciones Científicas – CSIC – IDAEA –Espanha, apresenta um novo método (Regions of Interests), desenvolvido pelo professor pesquisador Romà Tauler, que tem como objetivo

reduzir dados obtidos por LC-MS. Neste trabalho, foi avaliado o perfil lipidômico de duas classes de cianobactérias (*Anabaena* sp*.* e *Planktothrix agardhii*) após serem expostas a diferentes concentrações de Arsênio(III).

Capítulo 6 – Conclusões e perspectivas que sumariza os resultados encontrados, apresentando uma conclusão geral e fornece sugestões para trabalhos futuros.

Apêndice 1 – *Hyperspectral Raman spectroscopy image analysis of Cyanobacteria exposed to Arsenic(III) using multivariate curve resolution-alternating least squares (MCR-ALS).* – trabalho realizado em parceria com o Consejo Superior de Investigaciones Científicas – CSIC – IDAEA –Espanha. Neste trabalho, foram utilizadas imagens hiperespectrais obtidas por espectroscopia Raman de duas classes de cianobactérias (*Anabaena sp.* e *Planktothrix agardhii*) após serem expostas a diferentes concentrações de Arsênio(III), com o objetivo de avaliar as mudanças estruturais que ocorrem nas mesmas.

4. METODOLOGIA

Os estudos realizados com espectroscopia foram desevolvidos com a parceria entre o Instituto de Química e Departamento de Microbiologia e Parasitologia da Universidade Federal do Rio Grande do Norte, Natal,Brasil. Enquanto que o estudo com a cromatografia líquica acoplada a espectrometria de massa foi desenvolvido em parceria do Instituto de Química da Universidade Federal do Rio Grande do Norte, Natal, Brasil com o grupo de pesquisa CHEMAGEB do Conselho Superior de Investigações Científicas, Barcelona, Espanha, em projeto financiado pela CNPq/Capes (070/2012).

4.1. Obtenção e preparo das amostras

4.1.1. Staphylococcus aureus

As cepas de *Staphylococcus aureus*, ATCC 29213 foram fornecidas pelo Instituto Nacional de Controle de Qualidade em Saúde, Fundação Oswaldo

Cruz, Rio de Janeiro, Brasil, e em seguida foram cultivadas em caldo BHI (do inglês *Brain Heart Infusion*, Oxoid, Hampshire, Inglaterra) e diluídas em solução salina em diferentes concentrações (1x107, 1x106, 1x105, 1x104 e 1x103 UFC/mL).

4.1.2. Klebsiella pneumoniae

As cepas de *Klebsiella pneumoniae* foram fornecidas pelo Laboratório de Pesquisa em Infecção Hospitalar, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil, foram cultivadas em ágar sangue, MacConkey e BHI (HIMEDIA) e tiveram o seu DNA extraído com QIAamp DNA Mini Kit.

4.1.3. Pseudomonas aeruginosa

As amostras de *Pseudomonas aeruginosa* foram fornecidas pelo Laboratório de Micobactéria do Departamento de Microbiologia e Parasitologia da Universidade Federal do Rio Grande do Norte, Natal, Brasil.

4.1.4. Anabaena sp*. e Planktothrix agardhii*

As culturas de *Anabaena sp*. e *Planktothrix agardhii* foram obtidas da Coleção de Culturas de Algaes e Protozoários (CCAP1403/21 e CCAP1459/11) da Scottish Association for Marine Science, Reino Unido, foram replicadas e cultivadas em meio contaminado com Arsênio(III) e depois tiveram seus lipídeos extraídos.

4.2. Espectroscopia IR

4.2.1. Espectroscopia ATR-FTIR

Os espectros ART-FTR foram coletados no modo de transmissão na faixa de 600 a 4000 cm-1 usando o Bruker ALPHA FT-IR espectrômetro, equipado com o acessório ATR. Os espectros (32 varreduras, resolução espectral de 8 cm-1) foram convertidos em absorbância com o software Bruker OPUS.

4.2.2. Espectroscopia NIR

Cada espectro NIR (8 cm-1 de resolução espectral e 32 varreduras, em triplicata), foi adquirido no modo de reflectância com utilizando espectrômetro Mini Fourier-transform da ARCspectro ANIR.

4.3. LC-MS

As análises foram realizadas utilizando o sistema Waters Acquity UPLC conectado ao Waters LCT Premier espectrômetro de massa, operando em modo de ionização eletrospray positiva de 50 a 1800 Da em intervalos de aquisição de 0,2s. A fase estacionária utilizada foi a C8 Acquity UPLC e duas fases móveis, fase A: H2O 2 mM de formato de amônio e fase B: MeOH 1 mM de formato de amônio, ambas contendo 0,2% de ácido fórmico. A taxa de fluxo foi de 0,3 mL min-1 e o gradiente fase A/fase B iniciou com 20:80, mudou para 10:90 em 3 min, de 3 a 6 min manteve-se em 10:90, mudando para 1:99 até 18 min retornando a condição inicial até o minuto 20.

4.4. Análises computacionais

4.4.1. Pré-processamentos

Foram utilizados a correção multiplicativa de sinal (MSC) [referencia], derivadas e suavizações com o método Savitzky-Golay. [referencia]

4.4.2. Métodos

Análise de Componentes Principais (PCA), métodos de seleção de variáveis: Algoritmos de Projeção Sucessiva (SPA) e Algorítmo Genético (GA), seguidos de Análise Linear Discriminante (LDA) foram utilizados para classificar amostras das bactérias patogênicas descritas nos Capítulos 2, 3 e 4. Na elucidação de variações no perfil lipidômico das cianobactérias descritas no Capítulo 5 foi utilizado a estratégia de seleção de variáveis Regiões de Interesse (ROI) e Resolução Multivariada de Curvas com Mínimos Quadrados Alternantes (MCR-ALS).

4.4.3. Figuras de mérito

Sensibilidade (SENS) e Especificidade (SPEC) foram utilizados para avaliar os modelos de classificação multivariada. [referencia]

REFERÊNCIAS

[1] Foodborne Pathogenic Microrganisms and Natural Toxins Handbook: <**http://www.cfsan.fda.gov/~mow/chap4.html**>

[2] **<**http://www1.folha.uol.com.br/equilibrioesaude**>**, acessado em 14 de maio de 2016.

[3] DOERN, G.V., VAUTOUR, R., GAUDET, M., LEVY, B., 1994. Clinical impact of rapid in vitro suscep-tibility testing and bacterial identification. **J. Clin. Microbiol.** 32, 1757–1762.

[4] P. NORDMANN, G. CUZON, T. NAAS, **Lancet Infect. Dis**. v. 9, p. 228–236, 2009.

[5] C.P. THOMAS, L.S.P. MOORE, N. ELAMIN, M. DOUMITH, J. ZHANG, S. MAHARJAN, M. WARNERC, C. PERRYC, J.F. TURTONC, C. JOHNSTONEA, A. JEPSONA, N.D.C. DUNCANA, Forecasting carbapenem resistance from antimicrobial consumption surveillance: Lessons learnt from an OXA-48-producing *Klebsiella pneumoniae* outbreak in a West London renal unit. International Journal of Antimicrobial Agents. v. 46, p. 150-156, 2015.

[6] M.P. FINK, D.R. SNYDMAN, M.S. NIEDERMAN, K.V. LEEPER, R.H. JOHNSON, S. HEARD, ET AL., Treatment of severe pneumonia in hospitalized patients: results of a multicenter, randomized, double-blind trial comparing intravenous ciprofloxacin with imipenem-cilastatin, antimicrob. **Agents Chemother.** v. 38, p. 547–557, 1994.

[7] PENCE, M. A, MCELVANIA TEKIPPE, E., BURNHAM, C.-A.D. Diagnostic assays for identifi-cation of microorganisms and antimicrobial resistance determinants directly from , 2013.

[8] MS lista patógenos resistentes a antibióticos prioritários para pesquisa – Medscape. < https://portugues.medscape.com> acessado em 3 de março de 2017.

[9] WATERER, G.W., WUNDERINK, R.G., 2001. The influence of the severity of community-acquired pneumonia on the usefulness of blood cultures. **Respir. Med.** 95, 78–82.

[10] B.F. WOOLFREY, J.M.K. FOX, R.T. LALLY, C. QUALL, Broth microdilution testing of Pseudomonas aeruginosa and aminoglycosides: need for employing dilutions differing by small arithmetic increments**, J. Clin. Microbiol.** v. 16, p. 663–667, 1982.

[11] F. PASTERAN, O. VELIZ, M. RAPOPORT, L. GUERRIERO, A. CORSO, Sensitive and specific mod-ified Hodge test for KPC and metallo-beta-lactamase detection in Pseudomonas aeruginosa by use of a novel indicator strain, Klebsiella pneumoniae ATCC 700603, **Clin. Microbiol**. v. 49, p. 4301–4303, 2011.

[12] PENCE, M.; MCELVANIA TEKIPPE, E.; BURNHAM, C. A. D. Diagnostic assays for identification of microorganisms and antimicrobial resistance determinants directly from positive culture blood culture broth. **Clin Lab Med**. v. 33, p. 4950-4959.

[13] A.S. MARQUES, J.T.N. NICÁCIO, T.A. CIDRAL, M.C.N. DE MELO, K.M.G. DE LIMA. **Microbiol Methods.** v. 93, p. 90–94, 2013.

[14] MARTIN, F.L., KELLY, J.G., LLABJANI, V., MARTIN-HIRSCH, P.L., PATEL, I.I., TREVISAN, J., FULLWOOD, N.J., WALSH,M.J. Distinguishing cell types or populations based on the computation-al analysis of their infrared spectra. **Nat. Protoc**. v5, p. 1748–1760, 2010.

[15] KOSOGLU, M. A.; TATA, B. D.; LLEV, L. K.; HASSUN, M. Developing methodology to identify intrisec biomarkers in biological models using Fourier transform infrared (FTIR) spectroscopy. **Journal of Selected Topics in Quantum Eletronics**. v. 23, issue 2.

[16] MARQUES, A. N. L.; MENDES FILHO, J.; FREIRE, P. T. C.; SANTOS, H. S.; ALBUQUERQUE, M. R. J. R.; BANDEIRA, P. N.; LEITE, R. V.; BRAZ-FILHO, R.; GUSMÃO, G. O. M.; NOGUEIRA, C. E. S.; TEIXEIRA, A. M. R. Vibrational spectroscopy and DFT calculations of flavonoid

darriobtusone A. **Journal of molecurar structure**. v. 130, p. 231-237.

[17] RIDING, M.J.; MARTIN, F.L.; TREVISAN, J.; LLABJANI, V.; PATEL, I.I.; JONES, K.C.; SEMPLE, K.T. Concentration-dependent effects of carbon nanoparticles in gram-negative bacteria determined by infrared spectroscopy with multivariate analysis. **Environ. Pollut**. v. 163, p. 226–234, 2012.

[18] MARQUES, A. S. Uso da Espectroscopia do Infravermelho Próximo e técnicas multivariadas para diferenciar Escherichia coli e Salmonella Enteritidis inoculadas em polpa de fruta (abacaxi). Natal, 2013.

[19] GEORGE, W. O.; MCINTYRE, P.S. Infrared Spectroscopy. David J. Mowthorpe. Chichester : John Wiley, 1987.

[20] SKOOG, D.A. Princípios de Análise Instrumental, 5 ed. Porto Alegre: **Bookman**, 2002.

[21] DIEM, M., CHIRIBOGA, L. & YEE, H. Infrared spectroscopy of human cells and tissue. VIII. Strategies for analysis of infrared tissue mapping data and pplications to iver tissue. **Biopolymers**, v. 57, p. 282-290.

[22] BAUER, W.; WESTFALL, G. D.; DIAS, H. Física para Universitários - Mecânica, McGraw Hill Brasil, 2012

[23] PASQUINI, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. **J. Brasilian Chemical Society**. v.14 , p.198-219, 2003.

[24] BURNS, D.A.; CIURCZAK, E.W. Handbook of Near-infrared analysis, 3 ed. [S.l]: **CRC Press**, 2007

[25] KAEWTATHIP, T.; CHAROENREIN, S. Changes in volatile aroma compounds of pineapple (Ananas comosus) during freezing and thawing. **International Journal of Food Science & Technology**. v. 47,p. 985 – 990, 2012.

[26] BARKER, M.; RYENS, W. Partial leat squares for discrimintion. **Journal**

of Chemometrics. v. 17, p. 166-173, 2003.

[27] NAES, T.; ISAKSSON, T.; FERN, T.; DAVIES, T. A user-friendly guide to multivariate calibration and classification, UK: **NIR Publications**, 2002.

[28] COLTHRUP, N. B. Introduction to infrared and raman spectroscopy. 3 ed. Nova York: **Academic press**, 1990.

[29] TREVISAN, J. Chemical screening with the Syrian hamster embryo (SHE) assay coupled with attenuated total reflection (ATR) Fourier- transfom irfrared (FTIR) spectroscopy: from a case development towards frameworks for vibrational bioespectroscopy data analysis. 2012

[30] SKOOG, D. A.; HOLLER, F; J.; CROUCH, S. R. Principles of instrumental approach. 2 ed. Belmont: **Books-cole Thomson**, 2007.

[31] MARTIN, F. L.; KELLY, J. G.; LLABJANI, V.; MARTIN-HIRSCH, P. L., PATEL, I. I.; TREVISAN, J.; FULLWOOD, N. J.; WALSH, M. J. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. **Nature protocols**. v. 5, p. 1748-1760, 2010.

[32] AMIALI, N. M.; GOLDING, G. R.; SEDMAN, J.; SIMOR, A. E., ISMAIL, A. A. Rapid identification of community-associated methicilin-resistant Staphylococcus aureus by Fourier transform infrared spectroscopy. **Diagn. Microbiol. Infect. Dis**. v. 70, p. 157-166.

[33] PAVIA, D. L. Introduction to organic laboratory techniques: a small scale approach. 2 ed. Belmont: **Brooks-cole** Thomson, 2007.

[34] MARQUES, A.S.; MORAES, E. P.; JUNIOR, M. A. A.; MOURA, A. D.; NETO, V. F. A. Rapid discrimination of klebsiella pneumoniae carbapenemase 2 – producing and non-producing klebsiella pneumoniae strains using near-infrared spectroscopy (NIRS) and multivariate analysis, **Talanta**. v. 134, p.126–131, 2015.

[35] PONTES, M.J.C.; GALVÃO, R.K.H.; ARAÚJO, M.C.U.; MOREIRA, P.N.T.; NETO, O.D.P.; JOSÉ, G.E.; SALDANHA, T.C.B. The successive projections

algorithm for spectral variable se-lection in classification problems. **Chemom. Intell. Lab. Syst**. v. 78, p. 11–18, 2005

[36] TAPP, H.S.; DEFERNEZ, M.; KEMSLEY, E.K. FTIR spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils. **J. Agric. Food Chem**. v. 51, p. 6110–6115, 2003.

[37] JÚNIOR, C. A. M.; ANZANELLO, M. J. Sistemática de seleção de variáveis para classificação de produtos em categorias de modelos de reprodução. **Gest. Prod**. v. 22, p. 201-212, 2015.

[38] ANDRADE, M. C.; PINTO, L. C. M. Classificação de folhar por tamanho e forma através de descritores geométricos e análise de componentes principais**. IV Workshop em Tratamento de Imagens**, UFMG, p. 54-61 2003.

[39] FISHER, R. The use of multiple measurement in taxonomic problems. **Ann. Eugen**. v. 7, p. 179-189.

[40] DUDA, R. O.; HART, P. E.; Pattern Classification. **John Wilwy & Sons**, New York, 2 ed.

[41] SOARES, S. F. C.; GALVÃO, R. K.H.; PONTES, M. J. C.; ARAÚJO, M. C.U. A new validation criterion for guiding the selection of variable by the successive projections algorithm in classification problems. **J. Braz. Chem. Soc.** v. 25, p. 176-181, 2014.

[42] COSTA FILHO, P. A.; POPPI, R. J. Algorítmo genético em química. **Química Nova**. v. 22, p. 405-411, 1998.

[43] ARAÚJO, M. C. U.; SALDANHA, C. A.; GALVÇÃO, R. K. H. The sucessive projections algorithm for variable selection in espectroscopy multicomponente analysis. **Chemometrics and Intelligent Laboratory Systems**. v. 57, p. 65-73, 2001.

[44] BANDYOPADHYAY, S.; MURTHY, C. A.; PAL, S. K, Pattern classification with genetic algorithms. **Patterns Recognition Letters.** v. 6, p. 801-808, 1995.

[45] DE OLIVEIRA, V. E.; MIRANDA, M. A.C.; SOARES, M. S. C.;

EDWARDS, H. G. M.; DE OLIBVEIRA, L. F. C. Study of carotenoids in cyanobacterya by Raman spectroscopy. **Spectrochim. Acta A Mol. Biomol. Spectrosc.** v. 150, p. 373-380, 2015.

[46] YIN, X. X.; CHEN, J.; QIN, J.; SUN, G. X.; ROSEN, B. P.; ZHU, Y. G. Biotransformation and volatilization of arsenic by three photosyntetic cyanobacteria. **Plant Physiol**. v. 153(3), p. 1631-1638, 2011.

[47] CASSIER-CHAUVAT, C.; CHAUVAT, F. Responses to oxidative and heavy metal stresses in cyanobacteria: recente advances. **Int J Mol Sci**. v. 16, p. 871.

[48] BANKOVA, V., POPOV, S., MAREKOV, N. High performance liquid chromatographic analysis of flavonoids.

[49] CAMERON A E., EGGERS D F, An íon velocitron. **Review of Scientific Instruments**, v.19, p. 605, 1948.

[50] FENN, J. B., MANN, M., MENG, C.K., WONG, S. F. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. **Science**, v. 246, p. 64-71, 1989.

[51] STEPHENS, W. A Pulsed Mass Spectrometer with Time dispersion. **Physical Reviews**, v. 69, p. 691, 1946.

[52] MIDORIKAWA, K.; BANSKOTA, A. J.; TEZUKA, Y. NAGAOKA, T., MATSUSHIGE, K., MESSAGE, D., HUERTAS, A. A. G., KADOTA, S. Liquid chromatography Mass Spectrometry Analysis of Propolis. **Phytochemical Analysis**. v. 12, p. 366- 373, 2001.

[53] YAMASHITA, M., FENN, J. B. Electrospray Ions Source – Another Variation on the Free-jet Theme. **Journal of Physical Chemistry**, v. 88, p. 4451-4459, 1984.

[54] PEREIRA, A. S., PINTO, A. C., CARDOSO, J. N., AQUINO NETO, F. R., RAMOS, M. F. S., DELLAMORA-ORTIZ, G. M., SANTOS, E., P., Application of High Temperature High Resolution Gas Chromatography to Crude Extracts of Propolis. **Journal of High Resolution**

Chromatography. v. 21, n. 7, p. 396-400, 1998.

[55] HOFFMAN E., STROOBANT V. Mass Spectrometry. **Principles and Aplications**. 2 ed. England: Wiley and Sons,2003 , 407 pp.

[56] DADADASHI, M.; ABDOLLAHI, H.; TAULER, R. Error propagation along the different regions of multivariate curve resolution feasible solutions. **Chemometrics and Intelligent Laboratory Systems**. v. 162, p. 203-213, 2017.

[57] MALIK, A.; TAULER, R.Exploring the interaction between O3 and NOx pollution patterns in the atmosphere of Barcelona, Spain using the MCR–ALS method. **Science of The Total Environment**. v. 517, p. 151- 161, 2015.

[58] TAULER, R.; DE JUAN, A. Chapter 5: Multivariate Curve Resolution for Quantitative Analysis. **Data Handling in Science and Technology**. v. 29, p. 247-292, 2015.

[59] TAULER, R. 'OMICS' special issue for Chemometrics and Intelligent Laboratory Systems. **Chemometrics and Intelligent Laboratory Systems**. v. 104, p. 1, 2010.

[60] ZHANG, X.; TAULER, R. Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging. **Analytica Chimica Acta**. v. 762, p. 25-38, 2013.

Capítulo 2

Feature selection strategies for identification of *Staphylococcus aureus* recovered in blood cultures using FT-IR spectroscopy successive projections algorithm for variable selection: A case study

Aline de Sousa Marques

Maria Celeste Nunes de Melo

Thiago André Cidral

Kássio Michell Gomes de Lima

Journal of Microbiological Methods, 2014

Contribuição:

- Eu cultivei e repliquei as cepas bacterianas utilizadas neste trabalho.
- Eu desenvolvi o método para simulação do sangue contaminado.
- Eu coordenei a coleta dos espectros.
- Eu realizei o tratamento dos dados adquiridos.
- Eu escrevi a primeira versão do manuscrito.

Aline de Sousa Marques Prof. Dr. Kássio Michell Gomes de Lima

Journal of Microbiological Methods 98 (2014) 26–30

Contents lists available at ScienceDirect

Journal of Microbiological Methods

j our nal homepage: www.elsevier. com/l ocate/ jmicme th

Feature selection strategies for identification of *Staphylococcus aureus* recovered in blood cultures using FT-IR spectroscopy successive projections algorithm for variable selection: A case study

Aline de Sousa Marques [a], Maria Celeste Nunes de Melo [b], Thiago André Cidral [b], Kássio Michell Gomes de Lima [a,*]

[a] Universidade Federal do Rio Grande do Norte, Instituto de Química, Programa de Pós-Graduação em Química, Grupo de Pesquisa em Química Biológica e Quimiometria, CEP 59072-970 Natal, RN, Brazil
[b] Departamento de Microbiologia e Parasitologia, UFRN, Natal 59072-970, Brazil

a r t i c l e i n f o

Article history:
Received 17 November 2013
Received in revised form 19 December 2013
Accepted 21 December 2013
Available online 31 December 2013

Keywords:
Blood cultures
ATR-FTIR spectroscopy
S. aureus
LDA
GA–LDA
SPA–LDA

a b s t r a c t

Staphylococcus aureus is one of the leading causes of bacteremia, with high levels of accompanying morbidity and mortality. Current gold standard for the detection of *S. aureus* is very time-consuming, typically taking 24 h or longer. We set out to determine whether Fourier-transform infrared spectroscopy (FT-IR) combined with variable selection techniques, such as, genetic algorithm–linear discriminant analysis (GA–LDA) and successive projection algorithm–linear discriminant analysis (SPA–LDA) could be applied to detect this pathogen of bloodstream infection in samples based on the unique spectral "fingerprints" of their biochemical composition. Thirty real blood samples from healthy volunteers were contaminated with five different concentrations (10^7 until 10^3 CFU/mL) of microorganism and it analyzed by IR spectroscopy. The resulting GA–LDA model successfully classified all test samples with respect to their concentration in contaminated blood using only 18 wavenumbers. Discriminant functions revealed that GA–LDA clearly segregated different microorganism concentrations and the variable selected confirmed the chemical entities associated with the microorganism. The current study indicates that IR spectroscopy with feature selection techniques have the potential to provide one rapid approach for whole-organism fingerprint diagnostic microbial directly in blood culture.

1. Introduction

Bloodstream infections (BSIs) represent an important cause of human morbidity and mortality accounting for 30–40% of all cases of severe sepsis and septic shock (Doern et al., 1994). Diagnostic assays for identification of microorganisms and antimicrobial resistance determinants directly from positive blood culture broth are reported. (Pence et al., 2013). Prompt detection of microorganisms circulating in the bloodstream of patients is imperative as it allows clinicians to make decisions on possible therapeutic interventions (Dellinger et al., 2008; Weinstein and Doern, 2011). Automated blood culture systems are the most sensitive approach for detection of the bacteremia causative agent. There are some automated blood cul- ture systems commercially available, such as BACTEC FX (BD, Franklin Lakes, NJ, USA) and VersaTREK (ThermoFisher Scientific, Waltham, MA, USA). However, this procedure typically takes 24 h (e.g., for *Staphylococcus aureus* up to 5 days for *Candida* species) to generate the results (Pence et al., 2013). Moreover none of the currently available molec- ular methods is sufficiently rapid and accurate (Pence et al., 2013; Waterer and Wunderink, 2001). Because time is essential in preventing the evolution of BSI to severe sepsis or septic shock, faster detection methods are needed.

In the past 10 years, molecular techniques have been explored as tools for the identification of microbial species and subspecies (Wenning and Scherer, 2013). In particular, attenuated total reflection Fourier-transform infrared spectroscopy (ATR-FTIR) can be utilized to determine the microorganism metabolic fingerprint (DNA, RNA, proteins, cell-wall components), emerging as an interesting alternative for a rapid and cost-effective identification of microorganisms (Riding et al., 2012). ATR-FTIR is also characterized by a minimum of sample handling. It requires no extractions and is non-destructive. Moreover, amplifications, labeling, or staining steps of any kind are needless. The metabolic fingerprint generated by ATR-FTIR spectroscopy reflects the balance of some factors such as compositional and quantitative differences of biochemical compounds in microbial cells (Martin et al., 2010).

The application of IR spectrometry for *S. aureus* microorganism analysis (Amiali et al., 2011; Grunert et al., 2013; Maquelin et al., 2003) has been a target in the past years. Amiali et al. (2011) determined an FTIR spectral region or combination of regions reflecting a specific biochemical feature of a community-associated methicillin-resistant *S. aureus* (CA-MRSA). The authors provided a substitute for descriptive epidemiology in the definition of CA-MRSA strain types. Grunert et al. (2013) studied the potentiality of FTIR spectroscopy for

* Corresponding author. Tel.: +55 84 3215 3828; fax: +55 83 3211 9224.
E-mail address: kassio@ufrnet.br (K.M.G. de Lima).

0167-7012/$– see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.mimet.2013.12.015

differential diagnostic of the most clinically relevant *S. aureus* capsular polysaccharide types. Maquelin et al. (2003) realized a first prospective clinical study in which the causative pathogens (*Staphylococcus aureus*, *Enterococcus faecalis*, *Escherichia coli* and *Pseudomonas aeruginosa*) of blood infections were identified by FTIR spectroscopy.

For the analysis of *S. aureus* bacteria with IR spectroscopy, most of the reports are based on principal component analysis (PCA) for initial data reduction (de Sousa Marques et al., 2013), hierarchical cluster analysis (HCA) for analyzed groups in a set of data on the basis of spectral similarities (Martin et al., 2011), and linear discriminant analysis (LDA) for classify unknown samples into predetermined groups (Cheung et al., 2011). However, when employing full spectrum in the construction of these mathematic models, many variables are redundant and/or non-informative, and their inclusion may affect the performance of the final model. A well-succeeded approach to overcome this drawback is the successive projections algorithm (SPA) (Pontes et al., 2005) in conjunction with LDA and genetic algorithm (GA) (Tapp et al., 2003).

The present paper proposes the determination of an FTIR spectral region, or combination of variables, that reflects a specific biochemical feature of *S. aureus* in blood samples. We employed SPA and GA to select an appropriate subset of wavenumbers for LDA. Other goals were the elucidation of the altered variables using different concentrations of bacteria in the blood and the identification of the altered biochemical-bacteria fingerprint. This novel approach envisions rapid microbial identifications in clinical diagnostic assays.

2. Material and Thethods

2.1. Bacterial strain

Strain of *S. aureus* ATCC 29213 was cultivated in 2 mL of Brain Heart Infusion (BHI) broth (BHI, Oxoid, Ltd., Basingstoke, Hampshire, England) for 24 h at 35 °C. A microbiological strain suspension was standardized to 0.5 McFarland scale (~ 10^8 CFU/mL) in sterile saline.

2.2. Sample preparation

For IR measurements blood samples from healthy volunteers were contaminated with *S. aureus* in a microwell plate at five dilutions (1 × 10^7, 1 × 10^6, 1 × 10^5, 1 × 10^4 and 1 × 10^3 CFU/mL). The data set consisted of 36 samples that were divided into five for each dilution (30 samples, bacteria group) and six for the control group (uncontaminated blood).

2.3. ATR-FTIR spectroscopy

ATR-FTIR spectroscopy was performed using a Bruker ALPHA FT-IR spectrometer equipped with an ATR accessory. Spectra (8 cm^{-1} spectral resolution giving 4 cm^{-1} data spacing equivalent to 258 wavenumbers, co added for 32 scans) were converted into absorbance by Bruker OPUS software. The time measurement was of 26 s (32 scans) per spectrum. Absorbance spectra of bacterial samples were obtained against the spectrum of sterile blood used as background. Immediately following collection of each background, approximately 0.1 mL of each sample was applied to the ATR crystal using a transfer pipet, ensuring that no air bubbles were trapped on the crystal surface. After each measurement the ATR plate was washed with ethanol (70% v/v) and dried using tissue paper. Cleanliness of the ATR plate was verified by collecting an absorbance spectrum of the crystal using the most recently collected background as a reference. Before and between spectral acquisitions, samples were stored in the dark at ambient temperature. The ATR-FTIR spectrometer was placed in an air-conditioned room (21 °C), and samples were allowed to equilibrate to this temperature before analysis.

2.4. Chemometric methods: PCA, LDA, SPA–LDA and GA–LDA

A data set with many variables can be simplified by performing data reduction which makes the system more easily interpretable. Principal component analysis (PCA) is a well-known way to reduce the number of variables, in which the spectral matrix X is decomposed as:

$$X = TP^t + E \quad (1)$$

where X is the $I \times J$ data matrix, T is the $I \times A$ matrix of score vectors, the score vectors t_a are orthogonal (i.e., $T^tT = \mathrm{diag}(\lambda_a)$ and λ_a are eigenvalues of the matrix X^tX), P is the $J \times A$ matrix of loadings vectors, E is the $I \times J$ residual matrix, I is the number of objects, J is the number of variables, and A is the number of calculated components.

LDA is a supervised linear transformation that projects the variables (wavenumbers, for example) into a variable-reduced space which is optimal for discrimination between treatment classes. An LDA seeks for a projection matrix such that Fisher criterion (i.e. the ratio of the between-variance scatter to the within-class variance) is maximized after the projection. The variables created through LDA (factors) are linear combinations of the wavenumber-absorbance intensity values (Martin et al., 2007). Thus, the use of LDA for identification or classification of spectral data generally requires appropriate variable selection procedures (Pontes et al., 2005; Silva et al., 2013). In the present study, the SPA and GA were adopted for this function. In the SPA–LDA and GA–LDA models, the validation set was used to guide the variable selection, a strategy to avoid overfitting. The optimum number of variables for SPA–LDA and GA–LDA was determinate from the minimum of the cost function G calculated for a given validation data set as:

$$G = \frac{1}{N_V}\sum_{n=1}^{N_V} g_n; \quad (2)$$

where g_n is defined as

$$g_n = \frac{r^2\left(x_n; m_{I(n)}\right)}{\min_{I \neq I(n)} r^2\left(x_n; m_I\right)} \quad (3)$$

where $I(n)$ is the index of the true class for the nth validation object x_n.

In GA–LDA model, the mutation and reproduction probabilities were kept constant, 10 and 80%, respectively. The initial population was 120 individuals, with 50 generations. The best solution resulting from the three realizations of the GA was kept.

For this study, LDA scores, loading, and discriminant function (DF) values were derived for the biochemical-bacteria fingerprint region. The first LDA factor (LD1) was used to visualize the alterations of the blood sample in 1-dimensional (D) score plots that represented the main chemical alterations. SPA–LDA and GA–LDA were used to detect the biochemical alterations relative to the corresponding vehicle control (uncontaminated blood).

2.5. Software

The data import, pre-treatment, and the construction of chemometric classification models (LDA, SPA–LDA and GA–LDA) were implemented in the MATLAB version 6.5 (Math-Works, Natick, USA). Different preprocessing methods were used, including the baseline correction, derivative, smoothing Savitzky–Golay methods by using a first and second-order polynomial, and varying the number of window points (3, 5, 7 and 15). For SPA–LDA and GA–LDA models, the samples were divided into training (25), validation (6) and test sets (5) by applying the classic Kennard-Stone (KS) uniform sampling algorithm (Kennard and Stone, 1969) to the IR spectra.

3. Results and discussion

3.1. IR spectra and interpretation

Fig. 1a shows ATR-FTIR spectra of blood (in this case the air used as background) and bacterial samples at five dilutions (1×10^7, 1×10^6, 1×10^5, 1×10^4 and 1×10^3 CFU/mL). The IR absorption bands of bacterial cells, which are masked by the intense water signals in the spectrum of blood, reflect vibrations of molecules present in cap- sule, cell wall, membrane and cytoplasm. Spectra of the *S. aureus* (blood spectrum as background) at different concentrations are shown in Fig. 1b. This strategy also avoids the interference of water in the measurements. As can be seen, the ATR-FTIR spectrum of this biological system is quite complex and consists of some broad bands that arise from the superimposition of absorption by various macromolecules. This spectral region contains some blocks of relevant biochemical information: (i) 900 to 1200 cm^{-1}, the polysaccharide region (dominated by

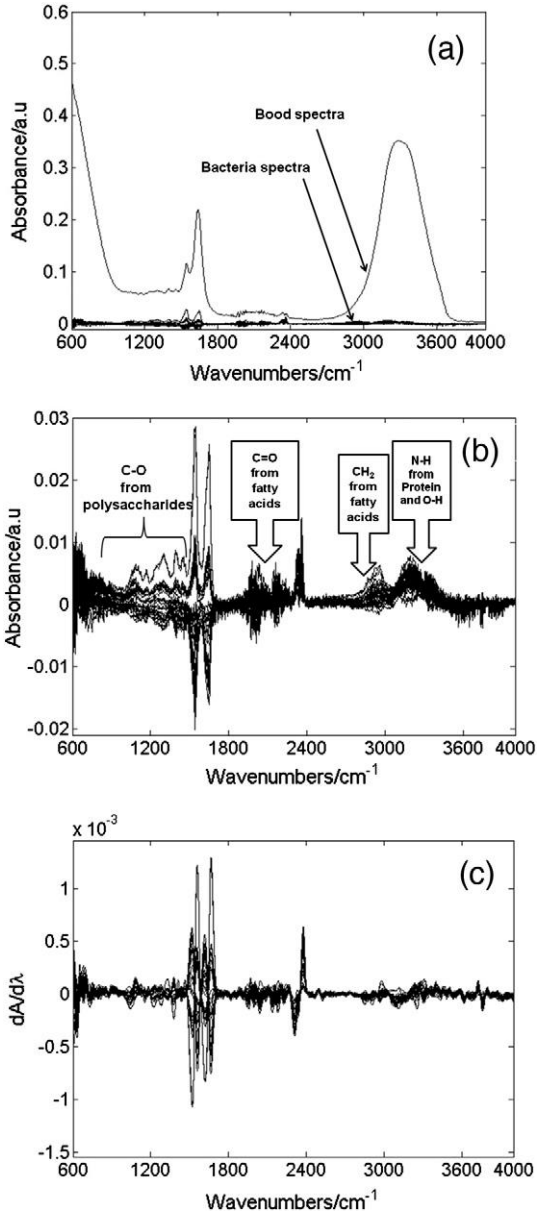


Fig. 1. Original spectrum of ATR-FTIR measurement of: (a) blood sample and *S. aureus* strain after inoculation in blood sample; (b) absorbance spectra of the biochemical-bacteria fingerprint regions for IR spectroscopy; (c) IR spectra with first derivative of the Savitzky–Golay using a window of fifteen points.

C\O\C and C\O\P stretching vibrations); (ii) 1300 to 1700 cm^{-1}, the region associated with lipids, proteins [Amide I(~1650 cm^{-1}), Amide II(~1550 cm^{-1}) and at ~1400 cm^{-1} (C__O stretching of amino acids)]; (iii) 1750–1850 cm^{-1}, C__O from lipids esters, carboxyl groups, and nucleic acids; (iv) 2800–3000 cm^{-1}, C\H of CH_2 in fatty acids and CH_3 in lipids and proteins; (v) 3100–3400 cm^{-1}, N\H of amide in proteins.

In our study, the differences among the spectra are in the quantity and distribution of the different functional groups. The identification of bacteria based on analysis of complex spectral signatures by evaluating peak intensities or half-width differences from a few bands that are resolvable by some means is very hard. These problems were circumvented by applying the Savitzky–Golay first derivative procedure with a first-order polynomial and a 15-point window, as shown in Fig. 1c. A first-derivative transformation makes unique spectral features of the different bacterial concentration more prominent. First-derivative transformation is often used to process spectral data because it separates overlapping absorption bands, removes baseline shifts, and increases apparent spectral resolution. Thus, the discriminatory power of ATR-FTIR spectroscopy can be significantly improved by combining it with appropriate chemometric, especially by employing supervised through multivariate techniques such as PCA and LDA that highlight the differences between the spectra. Furthermore, in this study, we have evaluated the improvement in the capability of the LDA in discriminating among the different concentrations of the bacteria in blood by incorporating two different variable selection methods, SPA and GA.

3.2. Principal component analysis

PCA is commonly employed to reduce the dimensionality of spectral data and to obtain preliminary information about data distribution. After processing, the range of 600-4000 cm^{-1} was submitted to PCA analysis. A PCA model was built from the calibration set using 2 PC, explaining together 91.12% of the variance in the data after applying the pre-processing (first derivative Savitzky–Golay). As seen in Fig. 2a, there is a separation with overlapping of the concentrations addressed in this case study. It is possible to observe, based on Fig. 2a, that the 1×10^5 and 1×10^6 classes are overlapped. On the other hand, since PCA is an unsupervised technique, this weak separation was obtained without using the information that the samples belong to 5 different groups and is indicative of features spectral differences. Based on this feasible study and preliminary results, the 1478-1721 cm^{-1} and 2282-2431 cm^{-1} regions should be considered optimal and more appropriate for the identification of *S. aureus* in blood, as shown in the Fig. 2b.

3.3. SPA–LDA and GA–LDA

SPA was applied to the data set and resulted in the selection of 10 variables, namely 598, 616, 657, 628, 698, 1665, 1961, 1998, 2145 and 3501 cm^{-1}. Using the 10 selected wavelengths, it was obtained the Fisher scores for all the samples of the data set (Fig. 3). Fig. 3 presents the first two (no-standardized) discriminant functions (DF1, DF2) for the overall data set. The coefficients of these functions were calculated by using the training-set statistics (class means and pooled covariance matrix) for the 10 selected variables. As can be seen, DF1 × DF2 does not discriminate among bacteria concentra- tion samples. In the SPA–LDA solution some of the wavenumbers se- lected by SPA are located in regions where no information is apparent. Hence, the inclusion of such variables has an adverse effect on the sensitivity of the LDA model to instrumental noise.

The GA employed for comparison resulted in the selection of 17 wavenumbers (among 1666 available), namely 618, 690, 824, 945, 1371, 1402, 1545, 1861, 2221, 2474, 2621, 2754, 3254, 3346, 3625, 3637 and 3762 cm^{-1}. Fig. 4a shows the screen plot associated with

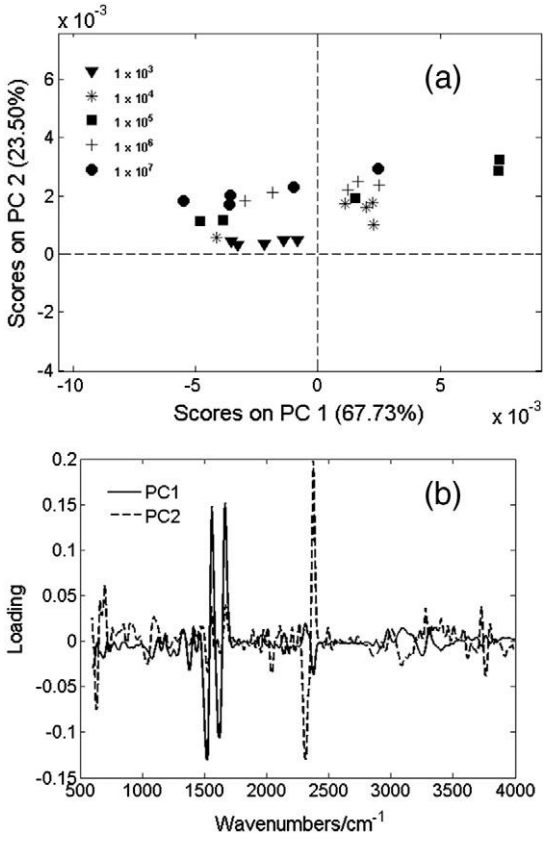


Fig. 2. PCA results: (a) score plot of PC1 versus PC2 (1 × 10³: ▼, 1 × 10⁴: *, 1 × 10⁵: ■, 1 × 10⁶:+, 1 × 10⁷: ●, CFU/mL) and (b) loading plot.

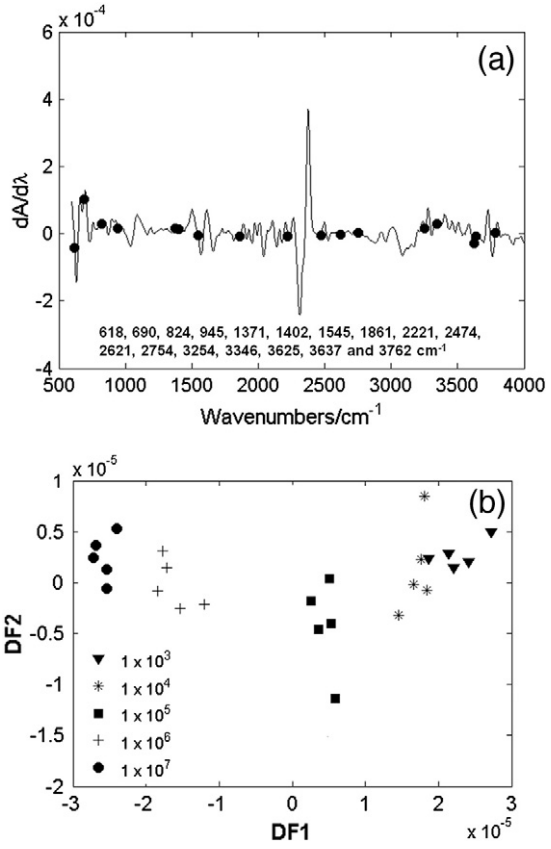


Fig. 4. (a) IR spectrum (first derivative of the Savitzky–Golay using a window of fifteen points). The arrows indicate the position in the spectra of the 17 wavenumbers variables selected by the GA–LDA; (b) DF1 × DF2 discriminant function values calculated by using the variables selected by GA–LDA of the data set (1 × 10³:▼, 1 × 10⁴: *,1 × 10⁵: ■, 1 × 10⁶:+, 1 × 10⁷: ●, CFU/mL).

the variable selection with GA–LDA, whose cost function minimum point was obtained with 17 wavenumbers. Using the selected wavenumbers by GA–LDA, it was obtained DF1 × DF2 for all the samples of the data set (Fig. 4b). As can be seen, there is a greater effect of homogeneity among classes, being obtained no misclassification, using only the 17 wavenumbers selected by GA in the LDA modeling. Examination of the selected wavenumbers following GA–LDA (Fig. 4a) indicated that the main biochemical alterations induced by *S. aureus* were on lipids, proteins [Amide II (~1550 cm^{-1}), and at ~1400 cm^{-1} (C__O stretching of amino acids)], and to a lesser extent at DNA vibrations (~1080 cm^{-1}). Several selected wavenumbers appear to be of particular interest, namely, the variables at 3254 and 3346, representing the peptide bonds from proteins. The variables at 3625, 3637 and 3762 represent the spectral region of fatty acid region and the variables between 900 and 600 cm^{-1} correspond to the fingerprint region.

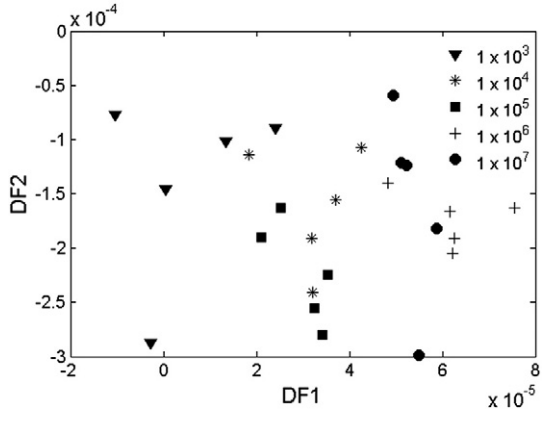


Fig. 3. DF1 × DF2 discriminant function values calculated with the variables selected by SPA–LDA of the data set (1 × 10³:▼, 1 × 10⁴: *,1 × 10⁵: ■, 1 × 10⁶:+, 1 × 10⁷: ●, CFU/mL).

4. Conclusions

In conclusion, the joint use of ATR-FTIR and discrimination analysis is a powerful means for routine identification of *S. aureus* in blood culture and hence a preferential choice for clinical diagnosis. This study proposed a methodology for bacteria identification employing ATR-FTIR spectrometry and LDA analysis coupled with the SPA for wavenumber selection. Variable selection techniques (SPA and GA) using LDA was subsequently performed in an attempt to gain more information regarding potential differences among the bacteria concentrations from the ATR-FTIR spectra. In a case study involving 5 different concentration bacteria samples, the resulting GA–LDA model successfully detected the biochemical alterations for the bacteria using only 17 wavenumbers. In contrast, traditional full-spectrum PCA model was not able to achieve weak separation. This study suggests that IR spectroscopy with computational analysis can be applied as a novel approach to investigate chemical dose-response relationships in blood cultures as well as to obtain biomarkers (selected wavenumbers). Although these results are encouraging, much larger databases of vibrational spectra of a wider range of microbial species, as well as a larger number of isolates per species, must be established. In the future, the miniaturization of instruments for field measurements which employed LED that emit radiation at wavelengths previously selected (e.g., UV, Vis, NIR and IR), the information generated by a multi-LED photometer may be sufficient for use in analysis or discriminatory simultaneous determinations based on multivariate analysis (de Lima, 2012).

AcknowledgThents

The authors would like to acknowledge the financial support from the CAPES for a fellowship for Marques, A.S., the Graduate Program in Chemistry (PPGQ) of UFRN and the INCQS-FIOCRUZ–RJ for providing the bacterial samples. We are grateful to Fabio Godoy (Bruker Optics Ltd.) for excellent technical assistance in this study by Bruker ALPHA FT-IR spectrometer. K.M.G. Lima acknowledges the CNPq/Capes Project (Grant 070/2012) and FAPERN (Grant 005/2012) for financial support.

References

Amiali, N.M., Golding, G.R., Sedman, J., Simor, A.E., Ismail, A.A., 2011. Rapid identification of community-associated methicillin-resistant *Staphylococcus aureus* by Fourier transform infrared spectroscopy. Diagn. Microbiol. Infect. Dis. 70, 157–166.

Cheung, K.T., Trevisan, J., Kelly, J.G., Ashton, K.M., Stringfellow, H.F., Taylor, S.E., Singh, M.N., Martin-Hirsch, P.L., Martin, F.L., 2011. Fourier-transform infrared spectroscopy discriminates a spectral signature of endometriosis independent of inter-individual variation. Analyst 136, 2047–2055.

De Lima, K.M.G., 2012. A portable photometer based on LED for the determination of aromatic hydrocarbons in water. Microchem. J. 103, 62—67.

De Sousa Marques, A., Nicácio, J.T.N., Cidral, T.A., de Melo, M.C.N., de Lima, K.M.G., 2013. The use of near infrared spectroscopy and multivariate techniques to differentiate *Escherichia coli* and *Salmonella* Enteritidis inoculated into pulp juice. J. Microbiol. Methods 93, 90—94.

Dellinger, R.P., Levy, M.M., Carlet, J.M., Bion, J., Parker, M.M., Jaeschke, R., Reinhart, K., Angus, D.C., Brun-Buisson, C., Beale, R., Calandra, T., Dhainaut, J.-F., Gerlach, H., Harvey, M., Marini, J.J., Marshall, J., Ranieri, M., Ramsay, G., Sevransky, J., Thompson, B.T., Townsend, S., Vender, J.S., Zimmerman, J.L., Vincent, J.-L., 2008. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. Intensive Care Med. 34, 17—60.

Doern, G.V., Vautour, R., Gaudet, M., Levy, B., 1994. Clinical impact of rapid in vitro susceptibility testing and bacterial identification. J. Clin. Microbiol. 32, 1757—1762.

Grunert, T., Wenning, M., Barbagelata, M.S., Fricker, M., Sordelli, D.O., Buzzola, F.R., Ehling-Schulz, M., 2013. Rapid and reliable identification of *Staphylococcus aureus* capsular serotypes by means of artificial neural network-assisted Fourier transform infrared spectroscopy. J. Clin. Microbiol. 51, 2261—2266.

Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148.

Maquelin, K., Kirschner, C., Vreeswijk, T. Van, Stämmler, M., Endtz, H.P., Bruining, H.A., Naumann, D., Puppels, G.J., 2003. Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. J. Clin. Microbiol. 41, 324–329.

Martin, F.L., German, M.J., Wit, E., Fearn, T., Ragavan, N., Pollock, H.M., 2007. Identifying variables responsible for clustering in discriminant analysis of data from infrared microspectroscopy of a biological sample. J. Comput. Biol. 14, 1176–1184.

Martin, F.L., Kelly, J.G., Llabjani, V., Martin-Hirsch, P.L., Patel, I.I., Trevisan, J., Fullwood, N.J., Walsh, M.J., 2010. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. Nat. Protoc. 5, 1748—1760.

Martin, F.L., Llabjani, V., Evans, G., Trevisan, J., Martin-Hirsch, P.L., Patel, I.I., Stringfellow, H.F., 2011. High contrast images of uterine tissue derived using Raman microspectroscopy with the empty modelling approach of multivariate curve resolution-alternating least squares. Analyst 136, 4950–4959.

Pence, M. a, McElvania TeKippe, E., Burnham, C.-A.D., 2013. Diagnostic assays for identification of microorganisms and antimicrobial resistance determinants directly from positive blood culture broth. Clin. Lab. Med. 33, 651—684.

Pontes, M.J.C., Galvão, R.K.H., Araújo, M.C.U., Moreira, P.N.T., Neto, O.D.P., José, G.E., Saldanha, T.C.B., 2005. The successive projections algorithm for spectral variable selection in classification problems. Chemom. Intell. Lab. Syst. 78, 11—18.

Riding, M.J., Martin, F.L., Trevisan, J., Llabjani, V., Patel, I.I., Jones, K.C., Semple, K.T., 2012. Concentration-dependent effects of carbon nanoparticles in gram-negative bacteria determined by infrared spectroscopy with multivariate analysis. Environ. Pollut. 163, 226–234.

Silva, C.S., Borba, F.D.S.L., Pimentel, M.F., Pontes, M.J.C., Honorato, R.S., Pasquini, C., 2013. Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. Microchem. J. 109, 122–127.

Tapp, H.S., Defernez, M., Kemsley, E.K., 2003. FTIR spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils. J. Agric. Food Chem. 51, 6110—6115.

Waterer, G.W., Wunderink, R.G., 2001. The influence of the severity of community-acquired pneumonia on the usefulness of blood cultures. Respir. Med. 95, 78—82.

Weinstein, M.P., Doern, G.V., 2011. A critical appraisal of the role of the clinical microbiology laboratory in the diagnosis of bloodstream infections. J. Clin. Microbiol. 49, S26–S29.

Wenning, M., Scherer, S., 2013. Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method. Appl. Microbiol. Biotechnol. 97, 7111–7120.

Capítulo 3

Rapid discrimination of Klebsiella pneumoniae carbapenemase 2 – producing and non-producing Klebsiella pneumoniae strains using near-infrared spectroscopy (NIRS) and multivariate analysis

Aline S. Marques

Edgar Perim Moraes

Miguel A. A. Júnior

Andrew D. Moura

Valter F. A. Neto

Renato M. Neto

Kássio M. G. Lima

Talanta, 2014

Contribuição:

- Eu coordenei a coleta dos espectros.
- Eu realizei o tratamento dos dados adquiridos.
- Eu escrevi a primeira versão do manuscrito.

Aline de Sousa Marques Prof. Dr. Kássio Michell Gomes de Lima

Talanta 134 (2015) 126–131

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Rapid discrimination of *klebsiella pneumoniae* carbapenemase 2 – producing and non-producing *klebsiella pneumoniae* strains using near-infrared spectroscopy (NIRS) and multivariate analysis

Aline S. Marques [a], Edgar P. Moraes [a], Miguel A.A. Júnior [b], Andrew D. Moura [b], Valter F.A. Neto [b], Renato M. Neto [b], Kássio M.G. Lima [a,n]

[a] Biological Chemistry and Chemometrics, Institute of Chemistry, Federal University of Rio Grande do Norte, Natal 59072-970, RN-Brazil
[b] Laboratory of Mycobateria, Department of Microbiology and Parasitology, Federal University of Rio Grande do Norte, Natal 59072-970, RN-Brazil

a r t i c l e i n f o

Article history:
Received 8 September 2014
Received in revised form
1 November 2014
Accepted 3 November 2014
Available online 12 November 2014

Keywords:
Carbapenemase
Klebsiella pneumoniae
Near infrared spectroscopy
SPA-LDA
GA-LDA

a b s t r a c t

Klebsiella pneumoniae Carbapenemase (KPC–2)-producing and non-producing *Klebsiella pneumoniae* (KP) have rapidly disseminated worldwide, challenging the diagnostics of Gram-negative infections. We evaluate the potential of a novel non-destructive and rapid method based on Near-Infrared Spectroscopic (NIRS) and multivariate analysis for distinguishing KPC-2 – producing and non-producing KP. Thirty-nine NIRS spectra (24 KPC-2-producing KP, 15 KPC-2 non-producing KP) were acquired; different pre-processing methods such as baseline correction, derivative and Savitzky–Golay smoothing were performed. A spectral region fingerprint was achieved after using genetic algorithm–linear discriminant analysis (GA–LDA) and successive projection algorithm (SPA–LDA) algorithms for variable selection. The variables selected were then used for discriminating the microorganisms. Accuracy test results including sensitivity and specificity were determined. Sensitivity in KPC-2 producing and non-producing KP categories was 66.7% and 75%, respectively, using a SPA-LDA model with 66 wavenumbers. The resulting GA-LDA model successfully classified both microorganisms with respect to their "fingerprints" using only 39 wavelengths. Sensitivity in KPC-2 producing category was moderate(≤66.7%) using a GA-LDA model. However, sensitivity in KPC-2 non-producing category using GA-LDA accurately predicted the correct class (with 100% accuracy). As100% accuracy was achieved, this novel approach identifies potential biochemical markers that may have a relation with microbial functional roles and means of rapid identification of KPC-2 producing and non-producing KP strains.

1. Introduction

Gram-negative bacilli (GNB) of clinical importance can be divided into two major groups. Glucose-fermenting, oxidase-negative, and catalase-positive members constitute one group, called Enterobacteriaceae [1]. Resistance to carbapenem among members of the Enterobacteriaceae family has become a major health care concern worldwide [2]. *Klebsiella pneumoniae* carbapenemase (KPC)-producing bacteria are an emerging group of highly drug-resistant bacteria causing infections associated with significant morbidity and mortality [3]. Among the most common clinical syndromes associated with KPC-producing bacteria presence are: pneumonia, urinary tract infections and manifestations of wounds, bacteremia, chronic atrophic rhinitis, arthritis, dysen- tery, meningitis, and sepsis in children, particularly those acquired in hospital [4]. The genes for the 10 known KPC variants (KPC 2–11) are carried on large plasmids [5]. Detection of KPC-2 producing bacteria may be a challenge for clinical laboratories because in this study it was associated with positive extended-spectrum β-lactamase (ESBL) confirmation tests (clavulanate-potentiated activities of ceftriaxone, ceftazidime, cefepime, and aztreonam) [6].

Traditionally, the tests for detection of KPC-type producing bacteria and other microorganisms are combined in a series of solid and/or liquid media which are inoculated with bacteria and interpreted/analyzed after a certain incubation period. The classical approaches are based on agar diffusion methods [7], broth microdilution (BMD) [8], modified Hodge test [9], polymerase chain reaction (PCR) – based assays [10], among others. Although these assays have achieved good sensitivity and specificity with favorable positive and negative predictive values for these micro- organisms, they often require multiple steps with additional time needed for the clarification/discernment of species and/or detec- tion of antimicrobial resistance. Moreover, slow multistep culture- based assays are time consuming, labor-intensive and require

[n] Corresponding author. Tel.: þ 55 84 3342 2323; fax: þ55 83 3211 9224.
E-mail address: kassiolima@gmail.com (K.M.G. Lima).

http://dx.doi.org/10.1016/j.talanta.2014.11.006
0039-9140/

skilled clinical microbiologists working in the lab. There is a need for a quicker, non-destructive, sensitive, and specific means of detecting and differentiating KPC-type producing bacteria strains.

Over the past decades, advanced molecular techniques in diagnostic microbiology have been revolutionizing the practice of clinical microbiology [11]. In particular, Near-Infrared Spectroscopy (NIRS) provides the ability to quickly detect analytes and has been explored as an interesting alternative tool for the identification of microbial species and subspecies [12–15]. Its sensitivity to the CH, NH, and OH absorptions related to microbial components, its speedy response time, the simplicity of sample preparation involved, the fact that the measurement is non-destructive, and its low instrumentation cost have fixed its position alongside other spectroscopies, including ultraviolet, visible, mid-infrared, Raman, and others.

Marques et al. [12] (2013) described the usefulness of NIRS in the identification and classification of Escherichia coli and Salmonella *enteritidis* from commercial fruit pulp (pineapple). The authors obtained good performance achieving prediction ability of 87.5% for *E. coli* and 88.3% for *S. enteritidis*, respectively. Rodriguez-Saona et al. [13] investigated the feasibility of NIRS in microbiology and the development of methodology for the quick identification of bacterial strains such as Escherichia coli HB101, non-virulent strain of Escherichia coli ATCC 43888, Escherichia coli ATCC 1224, Pseudomonas aeruginosa, Bacillus amyloliquifaciens, Bacillus cereus, and Listeria innocua. The use of NIRS spectral information and multivariate techniques in this study showed potential for the identification and subtyping of different bacterial species. Arnold et al. [14] used NIRS to monitor a submerged filamentous bacterial (Streptomyces *fradiae*) bioprocess. The present work is original/new in fully reporting how NIRS can be used to simultaneously measure the concentration of key analytes at-line in an antibiotic production process involving a filamentous bacterium, and in detailing the actual modeling process and subsequent critical assessment of model quality and performance.

However, the method of data analysis is a critical aspect of any diagnostic assay, particularly for NIR. The major difficulties in the analysis of microbial species and subspecies are the weakness of the NIR signals from strains components and the complexity of overlapping bands. To overcome these difficulties, many chemometric algorithms have been applied to NIR data such as principal component analysis (PCA) for initial data reduction and exploratory data analysis [16], hierarchical cluster analysis (HCA) for analyzing groups in a set of data on the basis of spectral similarities [17], and linear discriminant analysis (LDA) for classi- fying unknown samples into predetermined groups [18]. Further, as part of the computational methodology, variable selection methods such as successive projections algorithm (SPA) [19] in conjunction with LDA and genetic algorithm (GA) [20] improve the model performance compared with the full spectrum model. These algorithms eliminate potential interferents and variables that generate a lower signal/noise ratio.

Although these studies have shown that NIR together with chemometrics analysis have been explored as alternative tools for the identification of microbial species, little research has been directed toward the use of NIRS and variable selection to be used in pathogenic microbiology studies. Herein, we have attempted to evaluate the potential of a quicker method for identification of KPC-2-producing and non-producing *Klebsiella pneumoniae* strains. For this, the present article set out to determine whether biochemical intra-individual differences or "fingerprint" features between KPC-2-producing and non-producing bacteria could be identified using NIR spectroscopy with subsequent variable selec- tion methods. We employed SPA and GA to select an appropriate subset of wavenumbers for LDA. This approach can lead to more selective and specific microbial detection for medically relevant microorganisms by vibrational spectroscopy. Nevertheless, KPC-2-producing and non-producing *Klebsiella pneumoniae* was never discriminated by NIRS using wavelength selection to elucidate the altered biochemical-microbial fingerprint.

2. Material and methods

2.1. Bacteria strains

Specimens of Enterobacteriaceae family from different biological sites were recruited from three health centers in the city of Natal / Rio Grande do Norte, Brazil from April 2012 to August 2013. The specimens were cultured on blood agar (HIMEDIA), Mac Conkey Agar (HIMEDIA) and Brain Heart Infusion Broth (BHI/ HIMEDIA) followed by inoculation in conventional atmosphere at 35 1C for a period of 24 h. The phenotypic identification was confirmed by the Vitek automation system (BioMérieux Vitek, St. Louis). The antimicrobial susceptibility testing, as well as confirming the production of ESBL and KPC-producing tests were determined by the disk - agar diffusion (Kirby - Bauer), as recommended by the CLSI 2013.

2.2. Analysis of molecular pattern multi-resistance Enterobacteriaceae

Enterobacteriaceae with phenotypic standard for ESBL [21] and KPC genotypic analysis of the genes *bla* TEM, *bla* SHV, *bla* CTX -M and *bla* KPC were sent. For extraction and purification of total DNA, 250 QIAamp DNA Mini Kit (Cat. no. 51306) from Qiagen kit was used. The types of ESBL and KPC were determined through custom protocol to Polymerase Chain Reaction (PCR) using specific primers. The primer pair 5'–ATTCTTGAAGACGAAAGGGC-3' (forward) and 5'–ACGCT-CAGTGGAACGAAAAC-3' (reverse) was used for amplification of a sequence of 1150 base pairs (pb) from the TEM family [22]. For the SHV family, the primer pair 5'–GGGTTATTCTTATTTGTCGC-3' (forward) and 5'–TTAGCGTTGCCAGTGCTC-3' (reverse) (947pb) was used, for the CTX-M family, 5'–TTTGCGATGTGCAGTACCAGTAA-3' (forward) and 5'-CGATATCGTTGGTGGTGCCATA-3 ' (reverse) (544pb) [23] and KPC-2,5- TGTCACTGTATCGCCGTC-3 (forward) and 5- CTCAGTGCTCT ACAGAAAAACC-3 (reverse) (1100pb) [24] were used. The PCR was performed using an appropriate protocol in a final volume of 25 mL containing 1 mL of DNA, 16.25 mL of nuclease-free water, 2.5 mL of Taq buffer 10X, 2 mL of deoxynucleotide mixture, 0.75 mL of $MgCl_2$ (50 mmol L^{-1}), 1 mL (10 pmol) of each "primer" and 0.3 mL of Taq DNA polymerase (5U μL^{-1}, Ludwig, Alvorada /RS/Brazil). For the amplification conditions of the *bla* gene CTX-M, *bla* SHV, *bla* TEM and *bla* KPC were used at 94 1C pre-denaturation, 30 cycles of 1 min at 94 1C (denaturation), 2 min at 57.2 1C for annealing, 2 min 72 1C for extension, succeeded by a final extension of 8 min at 72 1 C. After the PCR reaction, the visualization of the amplified fragments was performed by agarose gel electrophoresis (Ludwig) 1% in 1X TAE (Tris-acetate).

2.2.1. NIR spectroscopy

Each NIR spectra (8 cm^{-1} spectral resolution, co-added for 32 scans and in triplicate) were directly acquired, in reflectance mode on a miniature scanning Fourier-Transform spectrometer from ARCspectro ANIR (Neuchâtel, Switzerland), which is based on a lamellar grating interferometer (35 mm 35 mm 65 mm) and uses a micro-mechanical actuator. The portable NIR device uses an InGaAs photodiode (900 nm to 2600 nm) and the reflected light was directed to the spectrometer through a bundle of optical fibers (model R600-7-VIS-125F, Ocean Optics, USA) linked to the probe end and the data acquisition and analyses were carried out by ARCspectro ANIR 1.64 software. Absorbance spectra of DNA

samples were obtained against the spectrum of reflectance standard (Labsphere, 8750) used as background. A disposable syringe (1 mL) was used to place the samples (0.1 mL) on an aluminum-plated backing plate (0.1 mm sample thickness). The transflectance probe was positioned on the sample surface. The transflectance probe was washed with ethanol (70% v/v) and dried using tissue paper after each sample. Cleanliness of the transflectance probe was verified by collecting an absorbance spectrum of the probe using the most recently collected background as a reference. Spectral measurements were done in an acclimatized room under controlled temperature of 22 1C, 60% relative air humidity, and samples were allowed to equilibrate to this temperature before analysis. Sample positioning, data collection, and storage took less than one minute per sample. Ninety spectra (45 KPC-2-producing *Klebsiella pneumoniae* and 45 KPC-2-non-producing *Klebsiella pneumoniae*) were randomized. Fig. 1 shows the experi- mental arrangement for sampling discrimination of KPC-2-producing and non-producing *Klebsiella pneumonia* using NIR spectroscopy.

2.3. Chemometrics methods: PCA, SPA-LDA and GA-LDA

For microbial identification, two basic approaches can be applied in the chemometric techniques: unsupervised and super- vised techniques. The objective of unsupervised methods, also called exploratory methods, is to depict the spectral data, without prior knowledge about the microorganism studied. Principal component analysis (PCA) [25] is a well-known unsupervised way to reduce the number of variables, in which the spectral matrix X is decomposed as:

$$X = TP^T + E; \quad (1)$$

where X is the $I \times J$ data matrix, T is the $I \times A$ matrix of score vectors, the score vectors t_a are orthogonal (i.e., $T^TT = \mathrm{diag}(\lambda_a)$ and λ_a are eigenvalues of the matrix X^TX), P is the $J \times A$ matrix of loadings vectors, superscript ^{T}, as usual, indicates the transpose of a matrix, E is the $I \times J$ residual matrix, I is the number of objects, J is the number of variables, and A is the number of components calculated. In other words, PCA is often used for reducing the dimensionality of the data without decreasing their variance, and each spectrum is then compared to the others in order to make homogeneous groups.

The second approach is based on supervised techniques coupled with variable selection methods such as genetic algorithm–linear discriminant analysis (GA–LDA) and successive projection algorithm (SPA–LDA). In this report, LDA will refer to the canonical discriminant procedure developed by Fisher in 1936 [26] and designed to maximize between-groups variability relative to a measure of pooled within-groups variability. The variables created through LDA (factors) are linear combinations of the wavenumber-absorbance intensity values [27]. Thus, the use of LDA for identification or classification of spectral data generally requires appropriate variable selection procedures [28]. In the present study, the validation samples were performed in SPA and GA to select the best optimum number of variables by minimizing a cost function calculated as:

$$G = \frac{1}{N_V}\sum_{n=1}^{N_V} g_n; \quad (2)$$

where g_n is defined as

$$g_n = \frac{r^2(x_n; m_{l(n)})}{\min_{l(m)\neq l(n)} r^2(x_n; m_{l(m)})} \quad (3)$$

where $l(n)$ is the index of the true class for the nth validation object x_n.

For this study, LDA scores, loading, and discriminant function (DF) values were derived for the biochemical-bacteria fingerprint region. The first LDA factor (LD1) was used to visualize the alterations of the bacteria sample in 1-dimensional (D) score plots that represented the main biochemical alterations.

2.3.1. Software

Data loading, pre-processing (mean-centering, Savitzky–Golay smoothing with different windows (3, 5, 7 and 15), first polynomial order, derivatization of first and second derivatives) and PCA were implemented in a MATLAB® version 7.10 environment (Math-Works, Natick, USA) with the PLS-toolbox version 7.5.2 (Eigenvector Research, Inc., Wenatchee, WA). The KS, GA–LDA and SPA–LDA classification routines were implemented in

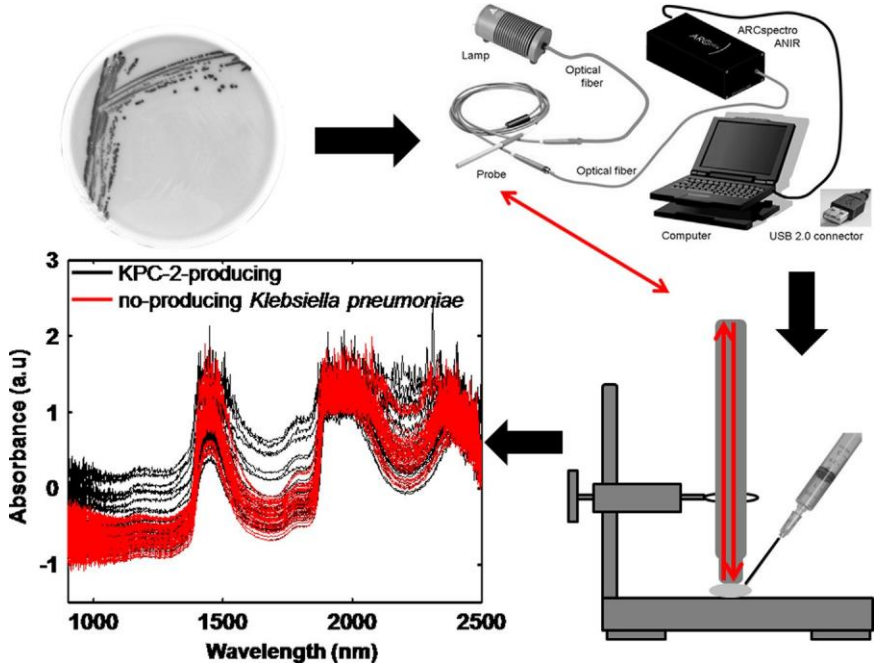


Fig. 1. Experimental arrangement for sampling discrimination of KPC-2-producing and no-producing *Klebsiella pneumoniae* using NIR spectroscopy.

Matlabs 7.10. For SPA-LDA and GA-LDA models, each class was treated separately. The Kennard Stone (KS) algorithm [29] was employed to select the samples which would compose the training (60% of samples), validation (20% of samples) and test (20% of samples) sets to the NIR spectra. Mean centering was applied to all spectra before performing variable subset selection and calibra- tion. The training and validation samples were used in the modeling procedures (including variable selection for LDA), whereas the test samples were only used in the final evaluation and comparison of the classification models.

For the GA routine, the initial population was 120 individuals with 60 generations each. The mutation and reproduction probabilities were kept constant at 10 and 60%, respectively. The best solution (in terms of the fitness value) resulting from the three realizations of the GA was kept. The final results (SPA-LDA and GA-LDA models) were expressed in terms of classification rates for the validation set.

Receiver-operating characteristic (ROC) analysis is recommended standard practice for test evaluation studies for non- binary tests [30]. For this study, measures of test accuracy, such as sensitivity (probability that a test result will be positive when the disease is present), specificity (probability that a test result will be negative when the disease is not present) were calculated as important quality standards in test evaluation.

3. Results and discussion

3.1. NIR spectra

As can be seen in Fig. 1, NIR spectra acquired from two classes (KPC-2-producing and non-producing *Klebsiella pneumoniae*) present a consistent baseline offset and bias. Although these are quite common features in NIR spectra acquired by diffuse reflectance techniques [31], some pre-treatments need to be performed. Among the pre-processing techniques tested, the one showing better separation of the classes employing PCA, SPA-LDA and GA- LDA was the combination of Savitzky-Golay first derivative (15 points window, first degree polynomial) and Savitzky-Golay smoothing (15 points window). In all cases, mean-centering was effectively applied to calculate the average spectrum of the data set and subtract the average from each spectrum. A first-derivative transformation makes unique spectral features of the different bacteria strains more prominent. In addition, first-derivative transformation is often applied to process spectral data since it separates overlapping absorption bands, removes baseline shifts and increases apparent spectral resolution. These effects are shown in Fig. 2.

As seen in Fig. 2, the consistent baseline offset and bias after processing were corrected and now it is possible to assign some

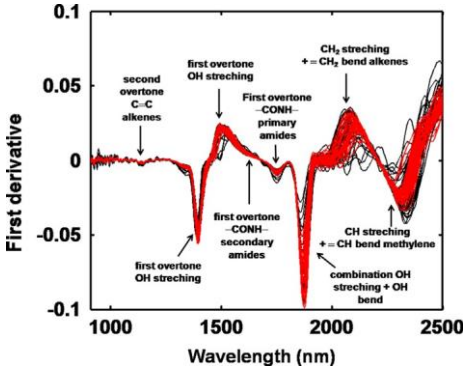


Fig. 2. NIR spectra with first derivative of the Savitzky-Golay using a window of fifteen points. (-) KPC-2-producing; (-) no-producing *Klebsiella pneumoniae*.

overtones and combination bands evidenced in the spectrum, including the following: a weak band at approximately 1140 nm was influenced by absorption exerted by the second overtone- related alkenes; absorption peaks were recorded at approximately 1398 and 1485, both associated with the first overtone OH stretching (water); a band at approximately 1632 nn is associated with the first overtone CONH secondary amides, a probable biochemical signature from DNA/RNA (cytosine, guanine and uracil), while at 1756 nm there was influence from absorption exerted by the first overtone —CONH— primary amides; a strong band at approximately 1900 nm was related to OH combination (stretching and bend modes); and the regions from 2110 to 2205 nm and from 2250 to 2300 nm were assigned to alkenes (CH_2 stretching and bend) and methylene (CH stretching and bend), respectively. Tentative assignments to KPC-2-producing and non-producing *Klebsiellapneumoniae* were based on systema- tic comparison of these major regions with the band known [32].

3.2. Principal component analysis

After processing, the range of 900-2600 nm was submitted to PCA analysis. The PCA model was built from the calibration set using 2 PC explaining together 65.34% of the variance in the data after applying the pre-processing (Savitzky-Golay smoothing and first derivative). As can be seen in Fig. 3a, the classes are over- lapped. This weak separation was obtained without using the information that the samples belong to two different groups and is indicative of features of/with spectral differences. The major loadings (Fig. 3b) are in the 1400–1410 nm (first overtone OH stretching) region with other contributing regions at 1890–1910 nm (OH combi- nation) and 2150–2400 nm (alkenes and methylene bands).

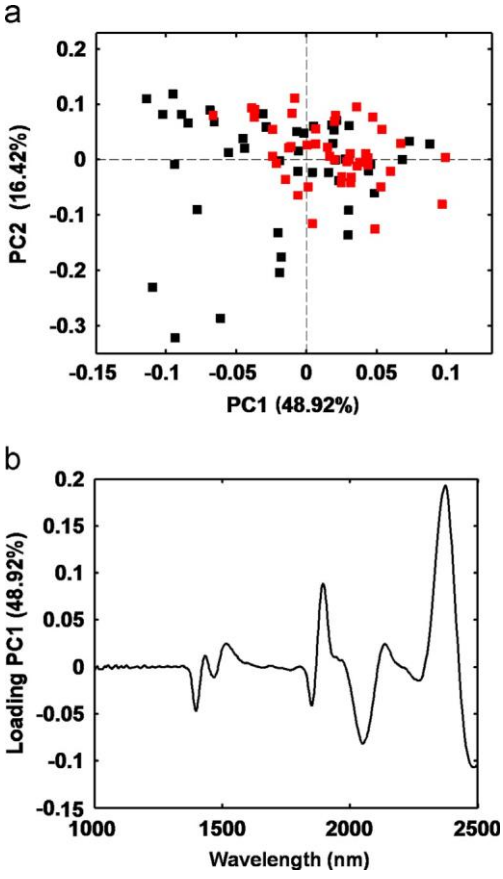


Fig. 3. PCA results: (a) score plot of PC1 *versus* PC2 (KPC-2-producing: ■, no- producing *Klebsiella pneumoniae*: and (b) loading plot (PC1, 48.92%).

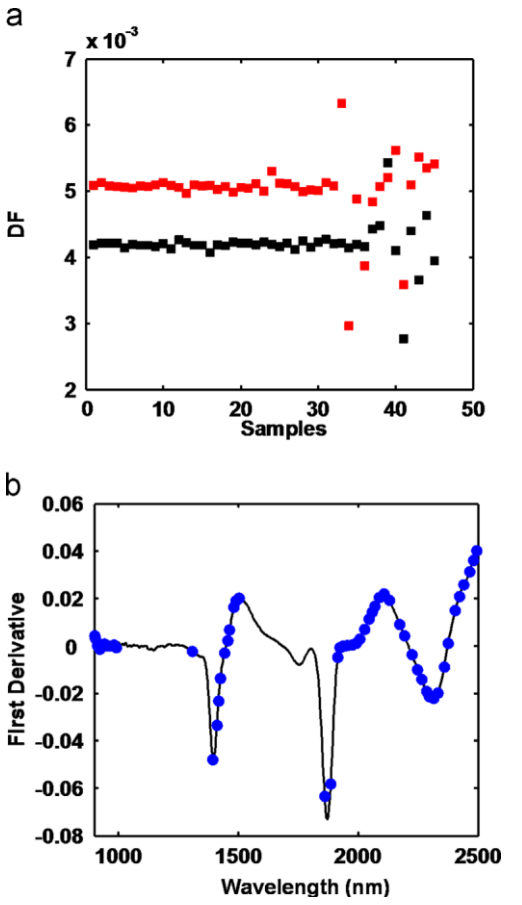


Fig. 4. (a) DF1 ~ samples discriminant function values calculated by using the variables selected by SPA–LDA of the data set (KPC-2-producing: ■,no-producing *Klebsiella pneumoniae*: ; (b) the blue full circles indicate the position in the spectra of the 66 wavelengths variables by using the variables selected by SPA–LDA of the data set.

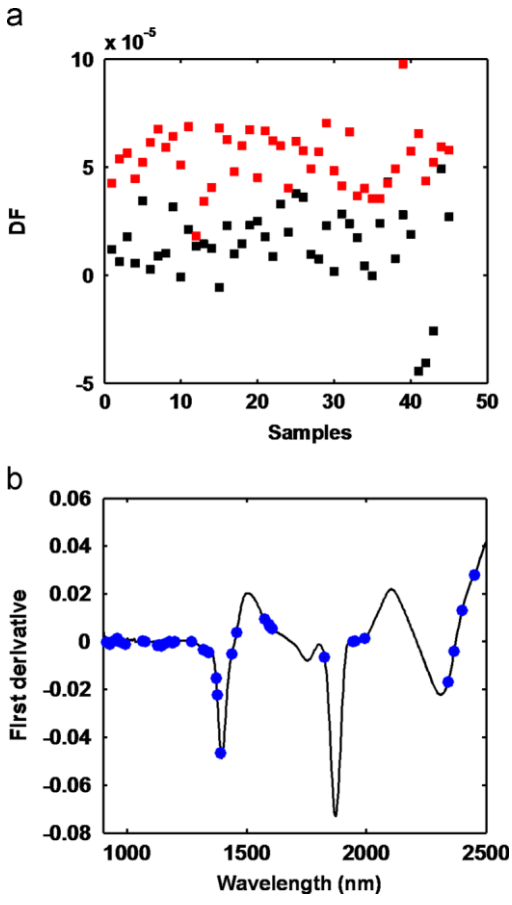


Fig. 5. (a) DF1 ~ samples discriminant function values calculated by using the variables selected by GA–LDA of the data set (KPC-2-producing: ■, no-producing *Klebsiella pneumoniae*: ; (b) the blue full circles indicate the position in the spectra of the 39 wavelengths variables by using the variables selected by GA–LDA of the data set.

3.3. *SPA-LDA and GA-LDA results*

SPA was applied to the data set (KPC-2-producing versus non-producing *Klebsiella pneumoniae*) and resulted in the selection of 66 variables, namely 900, 903, 911, 921, 930, 934, 942, 949, 963, 981, 990, 1308, 1393, 1410, 1417, 1425, 1443, 1455, 1463, 1481, 1489, 1502, 1860, 1885, 1914, 1923, 1937, 1943, 1956, 1974, 1990, 2006, 2025, 2045, 2059, 2069, 2086, 2107, 2129, 2173, 2192, 2224, 2248, 2264, 2285, 2297, 2314, 2332, 2358, 2376, 2403, 2422, 2441, 2465, 2480, 2494, 2520, 2530, 2540, 2556, 2561, 2577, 2587, 2593, 2598 and 2600 nm. Using the 66 selected wavelengths, the Fisher scores were obtained and there was also good segregation from each class. Fig. 4a shows the Fisher scores and Fig. 4b indicates the wavelength selected by SPA–LDA. Upon examination of the selected wavenumbers following SPA-LDA (Fig. 4b) indicate that the main biochemical alterations were on CH_2 stretchingþ CH_2 bending of vinyl alkenes (2192 nm) and CH stretchingþ C¼ O stretching of CHO aldehydes (2224 nm).The SPA-LDA model achieved a sensitivity of 66.7% and 75.0%, respectively, for KPC- 2-producing versus non-producing KP.

GA-LDA model for comparison achieved an improvement in segregation between KPC-2-producing versus non-producing KP. The GA resulted in the selection of 39 wavenumbers (of 909 available); these were 914, 927, 939, 958, 973, 993, 1065, 1075, 1128, 1134, 1144, 1157, 1178, 1190, 1200, 1270, 1318, 1322, 1341, 1342, 1372, 1376, 1391, 1438, 1458, 1458, 1576, 1592, 1596, 1600, 1606, 1824, 1943, 1953, 1993, 2340, 2367, 2399, 2450 and 2545 nm. Using the 39 selected wavelengths (Fig. 5a),the Fisher scores for all the samples of the data set (Fig. 5b) were obtained whose minimum point cost function was achieved with 39 wavenumbers. As can be seen, there was again an excellent separation from each category. The accuracy of GA-LDA for KPC-2-producing was 66.7% for sensitivity. For non-producing KP, GA-LDA model achieved a sensibility of 100.0%. Several selected wavenumbers appear to be of particular interest, namely the variables at 1318 and 2340 nm, representing the CH stretching and bending of methylene and NH stretching and bending of amide primary, respectively.

4. Conclusion

As was demonstrated, NIR spectroscopy facilitates the derivation of an integrated biochemical-bacteria fingerprint based on the NIR absorbing properties of the constituent chemical vibrations. The potential of NIR spectroscopy to discriminate KPC-2- producing and non-producing bacteria has been examined in this work. The study clearly demonstrates that different phenotypes of bacteria can be clearly segregated using NIR spectroscopy with subsequent PCA, SPA–LDA and GA–LDA algorithms. These findings have at least two important implications. First, NIRS and computational analysis can be applied as a novel approach to investigate these microbial pathogens because the timely and accurate detection/identification is critical for patient treatment decisions and outcomes for millions of patients each year. Second, with the miniaturization of instruments for field measurements employing light emission diode (LED) that emit radiation at wavelengths

previously selected (e.g., NIR), this approach may be used in analysis or discriminatory simultaneous determinations based on multivariate analysis [33].

Acknowledgments

The authors would like to acknowledge the financial support from the CAPES for a fellowship for Marques, A.S, the Graduate Program in Chemistry (PPGQ) of UFRN and mycobacteria, malaria and toxoplasmosis Laboratories. The authors thank Dra. Marise Dutra Asensi (Laboratório de Pesquisa em Infecção Hospitalar/LAPIH-Fio Cruz/RJ) for providing the bacterial strains control. The work was funded by grants from CNPq/Capes project (Grant 070/2012) and FAPERN (Grant 005/2012).

References

[1] D.L. Paterson, Am. J. Med. 119 (2006) S20–S28.
[2] P. Nordmann, G. Cuzon, T. Naas, Lancet Infect. Dis. 9 (2009) 228–236.
[3] C.P. Thomas, L.S.P. Moore, N. Elamin, M. Doumith, J. Zhang, S. Maharjan, M. Warnerc, C. Perryc, J.F. Turtonc, C. Johnstonea, A. Jepsona, N.D.C. Duncana, A.H. Holmesa, Antimicrob. Agents 42 (2013) 531–536.
[4] I. Pena, J.J. Picazo, C. Rodríguez-Avial, I. Rodríguez-Avial, Int. J. Antimicrob. Agents 43 (5) (2014) 460–464.
[5] F. Perez, A. Endimiani, A.J. Ray, B.K. Decker, C.J. Wallace, K.M. Hujer, D.J. Ecker, M.D. Adams, P. Toltzis, M.J. Dul, A. Windau, S. Bajaksouzian, M.R. Jacobs, R.A. Salata, R.A. Bonomo, J. Antimicrob. Chemoth. 65 (2010) 1807–1818.
[6] Clinical Laboratory Standards Institute. Performance Standards for Antimicrobial Susceptibility Testing: Twenty-third Informational Supplement M100-S23, CLSI, Wayne, PA, USA, 2013.
[7] S. Bratu, D. Landman, R. Haag, R. Recco, A. Ermo, M. Alam, J. Quale., Arch. Intern. Med. 165 (2005) 1430–1435.
[8] J.H. Jorgensen, J.C. Lee, Microdilution Technique for Antimicrobial Susceptibility Testing of Haemophilus influenzae 8 (1975) 610–612.
[9] K. Lee, Y. Chong, H.B. Shin, Y.A. Kim, D. Yong, J.H. Yum, Clin. Microbiol. Infect. 7 (2001) 88–91.
[10] J.M. Cole, A.N. Schuetz, C.E. Hill, F.S. Nolte, J. Clin. Microbiol. 47 (2009) 322–326.
[11] K. Maquelin, C. Kirschner, L.-P. Choo-Smith, N. van den Braak, H.P. Endtz, D. Naumann, G.J. Puppels, J. Microbiol. Methods 51 (2002) 255–271.
[12] A.S. Marques, J.T.N. Nicácio, T.A. Cidral, M.C.N. de Melo, K.M.G. de Lima, J. Microbiol. Methods 93 (2013) 90–94.
[13] L.E. Rodriguez-Saona, F.M. Khambaty, F.S. Fry, E.M. Calvey, J. Agric. Food Chem. 49 (2001) 574–579.
[14] S. Arnold, J. Crowley, S. Vaidyanathan, L. Matheson, P. Mohan, J. Hall, L.M. Harvey, B. McNeil, Enzyme Microb. Technol. 27 (2000) 691–697.
[15] D. Alexandrakis, G. Downey, A.G.M. Scannel, J. Agric. Food Chem. 56 (2008) 3431–3437.
[16] Y. He, X. Li, X. Deng, J. Food Eng. 79 (2007) 1238–1242.
[17] M.T. Bona, J.M. Andrés, Talanta 72 (2007) 1423–1431.
[18] H. Yang, J. Irudayaraj, M. Paradkar, Food Chem. 93 (2005) 25–32.
[19] M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, P.N.T. Moreira, O.D.P. Neto, G.E. José, T.C.B. Saldanha, Chemometr. Intell. Lab. Syst 78 (2005) 11–18.
[20] H.S. Tapp, M. Defernez, E.K. Kemsley, J. Agric. Food Chem. 51 (2003) 6110–6115.
[21] V. Jarlier, M.H. Nicolas, G. Fournier, A. Philippon, Rev. Infect. Dis. 10 (1988) 867–878.
[22] C.D. Steward, J.K. Rasheed, K. Susannah, J.W. Biddle, P.M. Raney, G.J. Anderson, P.P. Williams, K.L. Brittain, A. Oliver, J.E. McGowan Jr., F.C. Tenover, J. Clin. Microbiol. 39 (2001) 2864–2872.
[23] J.D.D. Pitout, D.L. Church, D.B. Gregson, B.L. Chow, M. Mccracken, M.R. Mulvey, K.B. Laupland, Antimicrob. Agents Chemother. 51 (2007) 1281–1286.
[24] H. Yigit, A.M. Queenan, G.J. Anderson, A. Domenech-sanchez, J.W. Biddle, C.D. Steward, S. Alberti, K. Bush, F.C. Tenover, Antimicrob. Agents Chemoth. 45 (2001) 1151–1161.
[25] B.M. Wise, N.L. Ricker, D.F. Veltkamp, B.R. Kowalski, Process Contr. Qual 1 (1990) 41–51.
[26] R.A. Fisher, Ann Eugen 7 (1936) 179–188.
[27] F.L. Martin, M.J. German, E. Wit, T. Fearn, N. Ragavan, H.M. Pollock, J. Comput. Biol. 14 (2007) 1176–1184.
[28] A.S. Marques, M.C.N. de Melo, T.A. Cidral, K.M.G. de Lima, J. Microbiol. Methods 98 (2013) 26–30.
[29] R.K.H. Galvão, M.C.U. Araujo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, Talanta 67 (2005) 736–740.
[30] K.T. Cheung, J. Trevisan, J.G. Kelly, K.M. Ashton, H.F. Stringfellow, S.E. Taylor, M.N. Singh, P.L. Martin-Hirsch, F.L. Martin, Analyst 136 (2011) 2047–2055.
[31] E.W. Ciurczak, Principles of near-infrared spectroscopy, in: D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, Marcel Dekker, New York, 2001, pp. 7–18.
[32] H.W. Siesler, Y. Ozaki, K. Satoshi, H.M. Heise, Near-Infrared Spectroscopy: Principles, Instruments, Applications, Wiley-VCH, Weinheim, 2002.
[33] K.M.G. de Lima, Microchem. J. 103 (2012) 62–67.

Capítulo 4

Near-infrared spectroscopy and variable selection techniques to discriminate Pseudomonas aeruginosa strains in clinical samples

Aline S. Marques

Jenielly F. Castro

Fagner J. M. D. Costa

Renato M. Neto

Kássio M. G. Lima

Microchemical Journal, 2016

Contribuição:

- Eu coordenei a coleta dos espectros.
- Eu realizei o tratamento dos dados adquiridos.
- Eu escrevi a primeira versão do manuscrito.

Aline de Sousa Marques Prof. Dr. Kássio Michell Gomes de Lima

Microchemical Journal 124 (2016) 306–310

Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc

Near-infrared spectroscopy and variable selection techniques to discriminate *Pseudomonas aeruginosa* strains in clinical samples

Aline S. Marques [a], Jenielly N.F. Castro [b], Fagner J.M.D. Costa [b], Renato M. Neto [b], Kássio M.G. Lima [a,*]

[a] Biological Chemistry and Chemometrics, Institute of Chemistry, Federal University of Rio Grande of Norte, Natal 59072-970, RN-Brazil
[b] Laboratory of Mycobateria, Department of Microbiology and Parasitology, Federal University of Rio Grande do Norte, Natal 59072-970, RN-Brazil

article info

Article history:
Received 16 July 2015
Received in revised form 9 September 2015
Accepted 10 September 2015
Available online 21 September 2015

Keywords:
Pseudomonas aeruginosa
Multi-resistant Sensitive
Near-infrared spectroscopy
SPA-LDA
GA-LDA

abstract

Pseudomonas aeruginosa is a leading cause of nosocomial infections, ranking second among the negative Gram pathogens reported to the National Nosocomial Infection Surveillance System. *P. aeruginosa* may develop resistant during prolonged therapy with all antimicrobial agents. Therefore, isolates that are initially susceptible may become resistant within 3–4 days after initiation of therapy. Testing of repeat isolates may be warranted. There is a need for sensitive and specific tests. We set out to determine whether near-infrared spectroscopy (NIR) combined with variable selection techniques employing successive projection algorithm – linear discriminant analysis (SPA-LDA) or genetic algorithm – (GA-LDA) could discriminate *P. aeruginosa* strains according to resistant vs. sensitive. The variables selected were then used for discriminating the strains. The influence of various spectral pre-treatments (Savitzky–Golay smoothing, multiplicative scatter correction (MSC), and Savitzky–Golay derivatives) was calculated. In addition, accuracy test results including sensitivity and specificity were determined. Sensitivity in the resistant category was 95% using a SPA-LDA model with 70 wavelengths. Sensitivity and specificity in both categories was 93% using a GA-LDA model with 32 wavelengths. We show that NIR spectroscopy of *P. aeruginosa* combined with variable selection techniques is a powerful tool for resistant vs. sensitive strains based on the unique spectral "fingerprints" of their biochemical microbial identification, emerging as an alternative for rapid and cost-effective identification of strains.

1. Introduction

Some bacteria are intrinsically resistant to antimicrobial agents because they lack the target site for that drug, or the drug is unable to reach the site of action, or the organism contains a chromosomally encoded resistant mechanism. For example, isolates of *Pseudomonas aeruginosa* may become resistant to the antibiotic being used to treat the infection [1] and that prior use of a particular antibiotic pre- dicts that *P. aeruginosa* will develop resistant to that antibiotic [2].

P. aeruginosa is the most prevalent species of non-fermentative Gram-negative bacilli (NFGNB) in clinical isolates of nosocomial nature, the leading cause of nosocomial bloodstream infections, ranking third among Gram-negative bacteria after *Escherichia coli* and *Klebsiella* species [3].

Detection of *P.aeruginosa* in the clinical laboratory is of major importance for the determination of appropriate therapeutic schemes and the implementation of infection control measures. In order to detect the presence of *P.aeruginosa* in clinical samples, the polymerase chain reaction (PCR), [4] disk diffusion or dilution methods, [5] broth microdilution (BMD), [6] and enzyme block test [7] are standard microbiological assays. However, these methods are time consuming, expensive, and involve numerical preparation steps like indispensable selective pre-enrichment. There is a need for sensitive and specific tests which are faster, non-destructive, and have/are specific means of detecting and differentiating *P. aeruginosa* strains.

Alternatively, near-infrared spectroscopy (NIRS) can be used on various microbiologic species for direct detection and identification [8—10], emerging as an interesting alternative for quick and cost-effective identification of living specimens. NIRS is also characterized by a minimum of sample handling. It requires no extractions and is non-destructive. Each bacterial species has a complex cell wall/membrane composition which gives a unique NIR fingerprint, due to the stretching and bending vibrations of molecular bonds or functional groups present in its proteins, nucleic acids, lipids, sugars, and lipopolysaccharides, among others. Therefore, each bacterium will have a unique and characteristic spectrum, and single microorganisms could be identified from an NIR spectrum. For example, Marques et al. [9] recently determined the biochemical intra-individual differences or "fingerprint" features between *Klebsiella pneumoniae* Carbapenemase (KPC–2)-producing and non-producing *Klebsiella pneumoniae* (KP) using NIR spectroscopy with subsequent variable selection methods. The authors concluded that NIRS

* Corresponding author. Tel.: +55 84 3342 2323.
E-mail address: kassiolima@gmail.com (K.M.G. Lima).

http://dx.doi.org/10.1016/j.microc.2015.09.006
0026-265X/

may provide a novel approach to identify potential biochemical markers that may have a relation with microbial functional roles between KPC-2 producing and non-producing KP strains.

Several circumstances have contributed to the successful development of NIR spectroscopy of microbiologic species. The use of multivariate statistical approaches, which allow for the extraction of qualitative and quantitative information from complex spectra for bacterial characterization and are largely responsible for advancing the NIR technique. For example, principal component analysis (PCA) for reducing the multidimensionality of the data set into its most dominant components or scores while maintaining the relevant variation between the data points [8], hierarchical cluster analysis (HCA) for identifying similarities between the spectra of microorganisms using the distances between spectra [11], and linear discriminant analysis (LDA) for the classification of objects into groups or clusters by determining the similarity of a set of values from an unknown sample to a set of values measured from a set of known samples [12]. In addition, variable selection methods, more specifically, successive projections algorithm (SPA) [13] and genetic algorithm (GA) [14] in conjunction with LDA have improved the model performance compared with the full spectrum model. On the other hand, a standardized experimental protocol in relation to media preparation, incubation time and temperature, cell harvesting conditions, sample preparations, and NIR measurement should be followed to obtain reproducible data.

The present paper investigates the use of NIR spectroscopy and chemometric techniques as a quick and non-destructive method for classification of *P. aeruginosa* strains into two different categories: species of standard multidrug-resistant phenotype (resistant to more than one class of antibacterial agent), and sensitive to all classes of antibacterial agents tested. This study is the first to apply NIR spectroscopy coupled with variable selection techniques to identify *P. aeruginosa* with multidrug-resistant isolated from clinical material profile. We employed SPA and GA to select an appropriate subset of wavelengths for LDA that reflects a specific biochemical feature of each category.

2. Material and Thethods

2.1. Bacteria strains

Sixty-three strains of *P. aeruginosa*, with varied susceptibility profile of the different classes of antibacterial agents (penicillins, β-lactam, β-lactamase inhibitor combinations, cephems, monobactams, carbapenems, lipopeptides, aminoglycosides, and fluoroquinolones) from different clinical specimens (various secretions, urine culture, blood culture, and catheter tip) were used in this study. From these samples, 54 (class 1) species were confirmed of being standard multidrug-resistant phenotype (resistant to more than one class of antibacterial agent), and 44 (class 2) were sensitive to all classes of antibacterial agents tested.

2.2. Maintenance of strains

After confirmation of the etiological and phenotypic resistant patterns of species, their identifications were stored for 18 months (January 2013 to July 2014) in tubes of nutrient agar with eppendorffs (HIMEDIA) and kept in the dark during the experiments. The cultures were renewed every 90 days. For the study, species were later removed from the culture collection of the Laboratory of Mycobacteria, Federal University of Rio Grande of Norte, Brazil.

2.3. Etiologic identification of strains

2.4. Phenotypic identification of bacterial resistant

The agar disk diffusion technique was used to assess susceptibility to different classes of antimicrobials for confirmation. Test-resistant patterns such as combined disk and enzyme block were used for strains that showed reduced susceptibility halo of the third-generation cephalosporins or carbapenems. *P. aeruginosa* ATCC27853 was used for control and validation tests.

2.5. NIR spectroscopy

Bacterial species were transplanted in studies of stock for nutrient Agar (HIMEDIA) plate and incubated for 24 h in bacteriological incubator (35 °C). Each NIR (spectral resolution of 8 cm^{-1}, in duplicate) spectra were directly acquired in the reflectance mode, using a miniature Fourier-transform scanning spectrometer from ARCspectro ANIR (Neuchâtel, Switzerland). The portable NIR device uses an InGaAs photodiode (900-2600 nm) and the reflected light was directed to the spectrometer via a bundle of optical fibers (model R600-7-VIS-125 F, Ocean Optics, USA) linked to the probe end and the data acquisition and analyses were carried out by ARCspectro ANIR 1.64 software. Absorbance spectra of samples of *P. aeruginosa* were obtained from the spectrum of the reflectance standard (Labsphere, 8750) which was used in the background. disposable syringes (1 mL) were used to place the sample on an aluminum plate (0.1 mm thickness) and then the reflectance probe was positioned under each sample, which was in a petri dish containing nutrient agar. The probe reflectance was washed with ethanol (70% v/v) and dried with paper after each sample. Spectral measurements were done in an acclimatized room under a controlled temperature of 22 °C, 60% relative air humidity, and samples were allowed to equilibrate to this temperature before analysis.

2.6. Multivariate analysis

The data import, pre-treatment, and construction of chemometric classification models (SPA-LDA and GA-LDA) were implemented in MATLAB R2014a software (Mathworks Inc, Natick, MA, USA). NIR spectra were pre-processed by Savitzky–Golay smoothing with different windows (3, 5, 7, and 15), first polynomial order, derivatisation of first/second derivatives, and multiplicative scatter correction (MSC). Mean centering was applied to all spectra before performing variable subset selection and calibration. For SPA-LDA and GA-LDA model, the samples were divided into training, validation, and prediction sets by applying the classic Kennard–Stone (KS) uniform sampling algorithm to the NIR spectra [16]. Sample numbers in each set are presented in Table 1. Training samples were used in the modeling procedure (including variable selection for LDA), whereas the prediction set was only used in the final evaluation of the classification. The optimum number of variables for SPA-LDA and GA-LDA was determined from the minimum cost function G calculated for a given validation dataset:

$$G = \frac{N_V}{N_n} \quad \text{ð1Þ}$$

$$N_V \quad n=1 \quad ;$$

Table 1

The samples were phenotypically identified using methods pub- lished by the American Society of Microbiology [15].

Number of training, validation, and test spectra in each category.

Category	Set training	Validation	Test
(1) *P. aeruginosa* resistant	34	10	10
(2) *P. aeruginosa* sensitive	24	10	10
Total	58	20	20

where g_n is defined as

$$g_n = \frac{\sum r^2(x_n, m_{I(n)})}{\min_{I(m)\neq I(n)} \sum r^2(x_n, m_{I(m)})} \quad (2)$$

and $I(n)$ is the index of the true class for the nth validation object x_n. g_n is defined as risk of misclassification of the nth validation object x_n, $n = 1, \ldots, N_V$). In this definition, the numerator is the squared Mahalanobis distance between object x_n (of class index I_n) and the sample mean $m_{I(n)}$ of its true class. The denominator in Eq. (2) corresponds to the squared Mahalanobis distance between object x_n and the center of the closest wrong class.

GA-LDA was used to select variables employing the G function as cost function. The mutation and reproduction probabilities were kept constant at 10% and 60%, respectively. The initial population was 100 individuals, with 50 generations. The best solution (in terms of the fitness value) resulting from the three realizations of the GA was employed. Receiver-operating characteristic (ROC) analysis is recommended standard practice for test evaluation studies for non-binary tests [17]. For this study, measures of test accuracy such as sensitivity (probability that a test result will be positive when the disease is present), specificity (probability that a test result will be negative when the disease is not present) were calculated as important quality standards in test evaluation. Both have a maximum value of 1 and a minimum of 0. The sensibility and specificity can be calculated following the equations:

$$\text{Sensibility}(\%) = \frac{TP}{TP + FN} \times 100 \quad (3)$$

$$\text{Specificity}(\%) = \frac{TN}{TN + FP} \times 100 \quad (4)$$

where FN is defined as false negative and FP as false positive. TP is defined true positive and TN is defined true negative.

3. Results and discussion

3.1. NIR spectra

Fig. 1 shows the collected diffuse reflectance spectra of *P. aeruginosa* strains (resistant and sensitive), which illustrates that the lowest molecular absorption is in the short wavelength region (900-1408 nm) with higher values in the first overtone region (1490-1852 nm), and still higher absorbance levels in the combination region (2083-2500 nm).

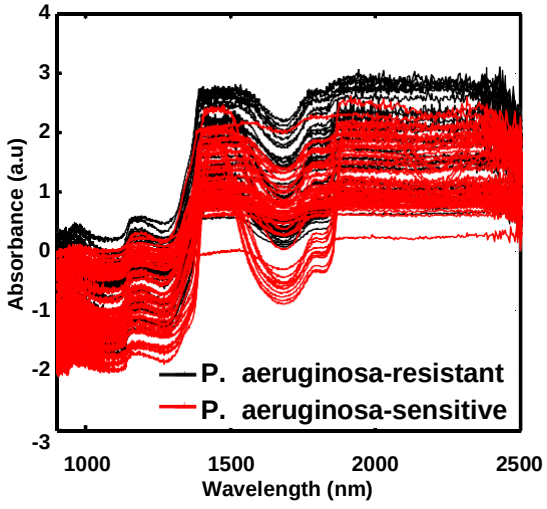


Fig. 1. Raw NIR spectra of *Pseudomonas aeruginosa* samples: (-) resistant, ( ) sensitive.

It can be seen in Fig. 1 that it features 900-1000 nm and 2100-2500 nm, which were removed from both spectral classes for construction of the classification models (SPA-LDA and GA-LDA). Baseline offsets and bias were present due to the light scattering. In an attempt to minimize the effects caused by the difficulty in obtaining an ideal spectrum without undesirable random variations, some pre-treatments were applied to the original spectra before construction of the classification models so that these variations did not influence the final results. Among the pre-processing techniques tested, the one showing better separation of the classes employing SPA-LDA and GA-LDA was the combination of Savitzky–Golay first derivative (15 points window, first degree polynomial), Savitzky–Golay smoothing (15 points window), and multiplicative scattering correction (MSC). These methods improve the signal–noise ratio (Savitzky–Golay smoothing), correct baseline displacement (Savitzky–Golay first derivative) and minimize the effects of light scattering (MSC). These effects are shown in Fig. 2.

As can be seen in Fig. 2, the spectra show some overtones and combination bands: 1143 nm (band marked with number 1), 1333 nm (band marked with number 2), 1384 nm (band marked with number 3), 1557 nm (band marked with number 4), 1753 nm (band marked with number 5), and 1871 nm (band marked with number 6). At 1143 nm, there is a second overtone of asymmetric stretching of $-CH_3$ methyl. At 1333 nm, there is the second overtone of C=C alkenes (vinyl group). At 1384 nm there is a first overtone of stretching and anti-symmetric O–H bond. At 1753 nm, there is the first overtone of anti-symmetric stretching of NH_2 and at 1871 nm, there is the second overtone of stretching O–H bend.

3.2. SPA-LDA and GA-LDA models

In this study, variable selection techniques (SPA and GA) in conjunction with LDA were applied to the spectral data set to investigate two different objectives simultaneously. First, to achieve a predictive method, with the goal of formulating a discrimination rule used to predict unknown *P. aeruginosa* strain (resistant and sensitivity) samples, measuring of test accuracy (sensitivity and specificity). Second, identification of the altered biochemical-bacteria fingerprint following SPA-LDA and GA-LDA models.

When SPA-LDA was used to segregate the two categories, resistant vs. sensitive, it resulted in the selection of 70 variables, namely, 1060, 1066, 1072, 1086, 1095, 1104, 1127, 1134, 1142, 1153, 1170, 1183, 1203, 1215, 1248, 1262, 1308, 1335, 1348, 1358, 1370, 1384, 1393, 1402, 1410, 1420, 1428, 1436, 1448, 1455, 1460, 1465, 1472, 1482, 1493, 1506, 1513, 1529, 1539, 1547, 1560, 1572, 1584, 1596, 1604,

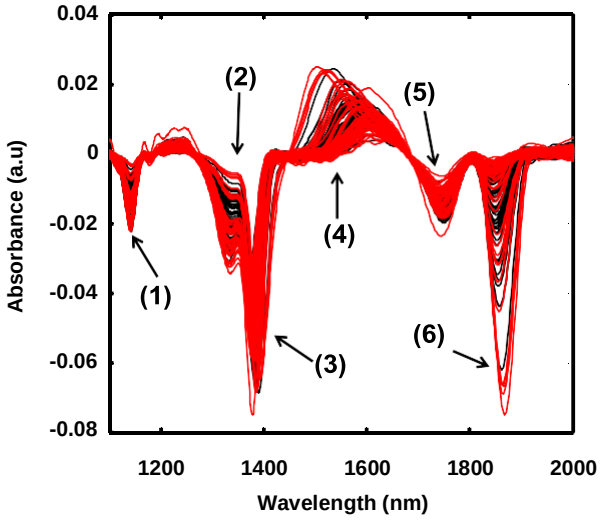


Fig. 2. NIR spectra with first derivative of the Savitzky–Golay using a window of fifteen points. (-) *P. aeruginosa* resistant, ( ) *P. aeruginosa* sensitive.

1614, 1638, 1653, 1673, 1698, 1721, 1738, 1753, 1770, 1790, 1806, 1819, 1843, 1857, 1868, 1885, 1902, 1917, 1929, 1940, 1956, 1971, 1987, 2006, and 2019 nm. Using the 70 selected wavelengths (Fig. 3A), the Fisher scores were obtained, and there was also a good segregation from each category, as shown in Fig. 3B. For *P. aeruginosa* strains resistant (category 1), the sensitivity and specificity obtained were 95%. For *P. aeruginosa* strains sensitive (category 2), the sensitivity and specificity obtained were also 95%. Examination of the selected wavelengths following SPA-LDA indicates that the main biochemical alterations may be associated with lipids, proteins, nucleic acids, and to a lesser extent, DNA vibrations. Several selected wavelengths for SPA-LDA appear to be of particular interest, namely, the variables at 1142 nm, 1335 nm, 1384 nm, representing the second overtone aromatic C–H bond, stretch overtone and deformation of the C–H bond and first overtone of stretching and anti-symmetric O–H bond, respectively.

Finally, Fig. 4A shows the plot scores with variable selection using GA-LDA, whose minimum point cost function was obtained with 32 wavelengths (Fig. 4B), namely, 1102, 1109, 1141, 1144, 1170, 1260, 1278, 1295, 1332, 1334, 1360, 1366, 1387, 1398, 1407, 1502, 1533, 1552, 1554, 1606, 1649, 1738, 1822, 1843, 1862, 1908, 1929, 1965, and 1999 nm. When GA-LDA was employed to predict resistant vs. sensitive, it was observed that using associated variables (32 selected) gives better segregation than SPA-LDA. As shown in Fig. 4A, there was again excellent separation from each category. The accuracy of GA-LDA for *P. aeruginosa* resistance and *P. aeruginosa* sensitivity achieved a sensibility of 93.0%. In the GA-LDA solution, some of the wavelengths selected appear to be of particular interest, namely, the variables at 1102 nm and 1108 nm, representing the C–H groups from lipids; 1387 nm and 1862 nm are related to the first overtone of stretching and anti-symmetric O–H bond and second overtone of stretching O–H bend, respectively.

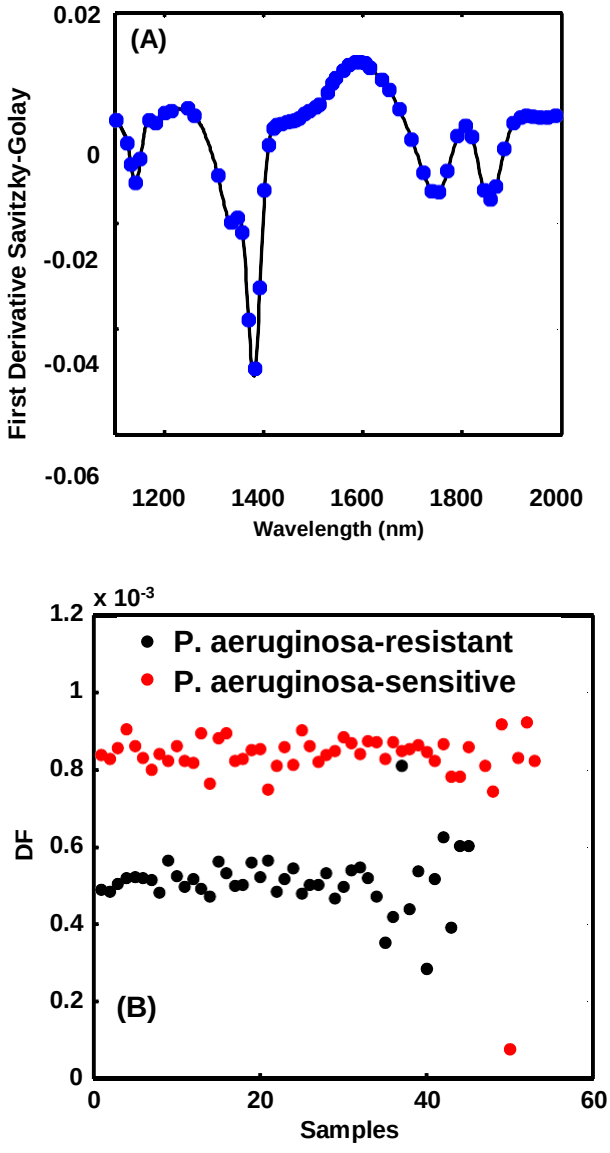


Fig. 3. The application of variable selection technique (SPA-LDA) to the segregation of *Pseudomonas aeruginosa* strains: (A) seventy wavenumber variables selected and (B) DF1 × DF2 discriminant function values calculated by using the variables selected by SPA-LDA from both categories: (●) *P. aeruginosa* resistant, (●) *P. aeruginosa* sensitive.

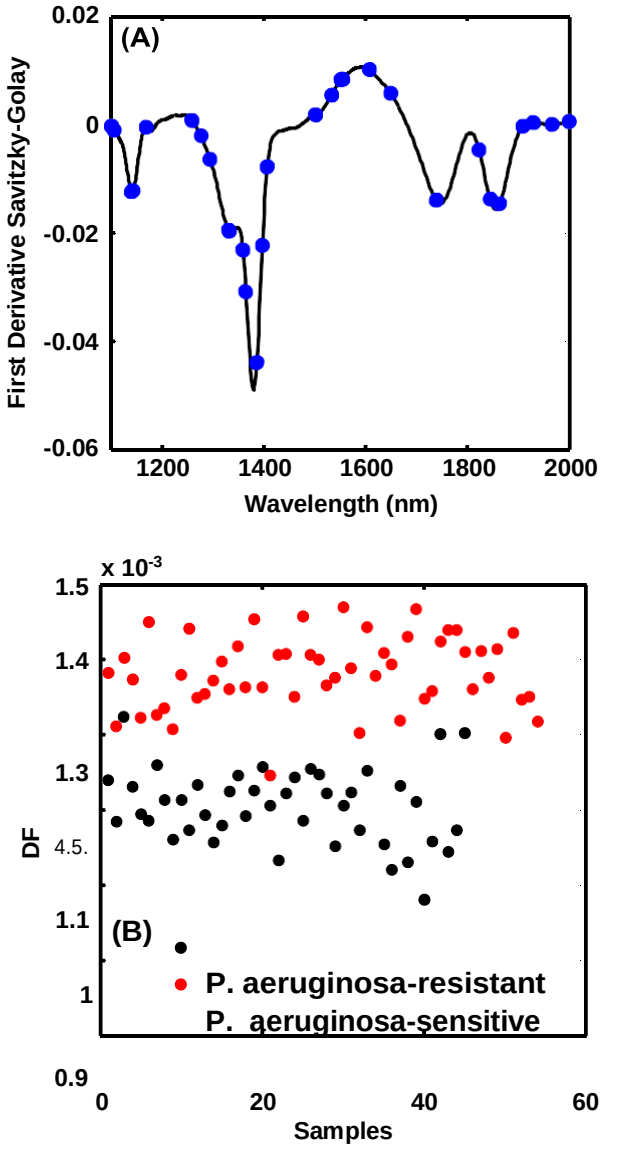


Fig. 4. The application of variable selection technique (GA-LDA) to the segregation of *Pseudomonas aeruginosa* strains: (A) thirty-two wavenumber variables selected, and (B) DF1 × DF2 discriminant function values calculated by using the variables selected by GA-LDA from both categories: (●) *P. aeruginosa* resistant, (●) *P. aeruginosa* sensitive.

4. Conclusion

This paper presented a non-destructive method to discriminate *P. aeruginosa* strains according to resistance vs. sensitivity, based on clinical samples from simple NIR spectroscopy measurements aided by recognition techniques with variable selection methods (SPA-LDA and GA-LDA). The metabolic fingerprint generated by NIR spectroscopy combining with variable selection methods is a powerful adjunct for implementation of infection control measures, emerging as an alternative for quick/rapid and cost-effective identification of microbiologic species.

The study clearly demonstrates that *P.aeruginosa* strains according to resistance vs. sensitivity can be clearly segregated using NIR spectroscopy with SPA-LDA and GA-LDA algorithms. The SPA-LDA presented better performance than GA-LDA, archiving a correct classification rate of 95%. The proposed method contributes to increasing the technological level of sensitive and specific tests which are faster, non-destructive and specific means of detecting and differentiating these microbial pathogens due to timely and accurate detection/identification being critical for patient treatment decisions and outcomes for millions of patients each year.

AcknowledgeThents

The authors would like to acknowledge the financial support from CAPES for a fellowship for Marques, A.S, the Graduate Program in Chemistry (PPGQ) of UFRN and the Laboratory of mycobacteria, malaria, and toxoplasmosis. The work was funded by grants from CNPq/Capes project (Grant 070/2012 and 442087/2014-4) and FAPERN (Grant 005/2012).

References

[1] M.P. Fink, D.R. Snydman, M.S. Niederman, K.V. Leeper, R.H. Johnson, S. Heard, et al., Treatment of severe pneumonia in hospitalized patients: results of a multicenter, randomized, double-blind trial comparing intravenous ciprofloxacin with imipenem-cilastatin, antimicrob, Agents Chemother. 38 (1994) 547–557.

[2] E.B. El Amari, E. Chamot, R. Auckenthaler, J.C. Peche, C. Van Delden, Influence of previous exposure to antibiotic therapy on the susceptibility pattern of *Pseudomonas aeruginosa* bacteremic isolates, Clin. Infect. Dis. 33 (2001) 1859—1864.

[3] D.J. Diekema, M.A. Pfaller, R.N. Jones, G.V. Doern, P.L. Winokur, A.C. Gales, et al., Survey of bloodstream infections due to Gram-negative bacilli: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, and Latin America for the SENTRY Antimicrobial Surveillance Program, 1997, Clin. Infect. Dis. 29 (1999) 595–607.

[4] F. Karpati, J. Jonasson, Polymerase chain reaction for the detection of *Pseudomonas aeruginosa*, *Stenotrophomonas maltophilia* and *Burkholderia cepacia* in sputum of patients with cystic fibrosis, Mol. Cell. Probes 10 (1996) 397–403.

[5] J.L. Burns, L. Saiman, S. Whittier, D. Larone, J.A.Y. Krzewinski, Z. Liu, et al., Comparison of agar diffusion methodologies for antimicrobial susceptibility testing of *Pseudomonas aeruginosa* isolates from cystic fibrosis patients, J. Clin. Microbiol. 38 (2000) 1818—1822.

[6] B.F. Woolfrey, J.M.K. Fox, R.T. Lally, C. Quall, Broth microdilution testing of *Pseudomonas aeruginosa* and aminoglycosides: need for employing dilutions differing by small arithmetic increments, J. Clin. Microbiol. 16 (1982) 663–667.

[7] F. Pasteran, O. Veliz, M. Rapoport, L. Guerriero, A. Corso, Sensitive and specific modified Hodge test for KPC and metallo-beta- lactamase detection in Pseudomonas aeruginosa by use of a novel indicator strain, *Klebsiella pneumoniae* ATCC 700603, J. Clin. Microbiol. 49 (2011) 4301—4303.

[8] A.S. Marques, J.T.N. Nicácio, T.A. Cidral, M.C.N. de Melo, K.M.G. de Lima, The use of near infrared spectroscopy and multivariate techniques to differentiate *Escherichia coli* and *Salmonella Enteritidis* inoculated into pulp juice, J. Microbiol. Methods 93 (2013) 90—94.

[9] A.S. Marques, E.P. Moraes, M.A.A. Júnior, A.D. Moura, V.F.A. Neto, R.M. Neto, et al., Rapid discrimination of klebsiella pneumoniae carbapenemase 2 — producing and non-producing klebsiella pneumoniae strains using near-infrared spectroscopy (NIRS) and multivariate analysis, Talanta 134 (2015) 126–131.

[10] L.E. Rodriguez-Saona, F.M. Khambaty, F.S. Fry, E.M. Calvey, Rapid detection and identification of bacterial strains by Fourier transform near-infrared spectroscopy, J. Agric. Food Chem. 49 (2001) 574—579.

[11] M. Harz, M. Kiehntopf, S. Stöckel, P. Rösch, E. Straube, T. Deufel, et al., Direct analysis of clinical relevant single bacterial cells from cerebrospinal fluid during bacterial meningitis by means of micro-Raman spectroscopy, J. Biophotonics 2 (2009) 70—80.

[12] A.M. Hamid, A.K. Jarmusch, V. Pirro, D.H. Pincus, B.G. Clay, G. Gervasi, et al., Rapid discrimination of bacteria by paper spray mass spectrometry, Anal. Chem. 86 (2014) 7500—7507.

[13] M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, P.N.T. Moreira, O.D.P. Neto, G.E. José, et al., The successive projections algorithm for spectral variable selection in classification problems, Chemom. Intell. Lab. Syst. 78 (2005) 11—18.

[14] H.S. Tapp, M. Defernez, E.K. Kemsley, FTIR spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils, J. Agric. Food Chem. 51 (2003) 6110—6115.

[15] G.L. Gilardi, Pseudomonas and Related Genera, in: E.W. Ballows, W.J. Hausler, K. Herrmann (Eds.), American Society for Microbiology, 5th ed., Man. Clin. Microbiol. 1991, pp. 429–441.

[16] R. Kennard, L. Stone, Computer aided design of experiments, Technometric 11 (1969) 137–148.

[17] K.T. Cheung, J. Trevisan, J.G. Kelly, K.M. Ashton, H.F. Stringfellow, S.E. Taylor, et al., Fourier-transform infrared spectroscopy discriminates a spectral signature of endometriosis independent of inter-individual variation, Analyst 136 (2011) 2047—2055.

Capítulo 5

Assessment of the effects of As(III) treatment on cyanobacteria lipidomic profiles by LC-MS and MCR-ALS

Aline S. Marques

Carmen Bedia

Romà Tauler

Kássio M. G. Lima

Analytica Bioanalytica Chemistry, 2016

Contribuição:

- Eu desenvolvi o método para cultivo das cianobactérias utilizadas.
- Eu cultivei, repliquei e mantive a cultura de cianobactérias.
- Eu fiz a extração dos lipídeos a partir de um procedimento padronizado.
- Eu coordenei a aquisição dos dados LC-MS.
- Eu fiz o tratamento quimiométrico.
- Eu escrevi a primeira versão do manuscrito,

Aline de Sousa Marques Prof. Dr. Kássio Michell Gomes de Lima

Anal Bioanal Chem
DOI 10.1007/s00216-016-9695-5

RESEARCH PAPER

Assessment of the effects of As(III) treatment on cyanobacteria lipidomic profiles by LC-MS and MCR-ALS

Aline S. Marques [1] & Carmen Bedia[2] & Kássio M. G. Lima[1] & Romà Tauler[2]

Received: 3 April 2016 / Revised: 25 May 2016 / Accepted: 6 June 2016

Abstract Cyanobacteria are a group of photosynthetic, nitrogen-fixing bacteria present in a wide variety of habitats such as freshwater, marine, and terrestrial ecosystems. In this work, the effects of As(III), a major toxic environmental pollutant, on the lipidomic profiles of two cyanobacteria species (*Anabaena* and *Planktothrix agardhii*) were assessed by means of a recently proposed method based on the concept of regions of interest (ROI) in liquid chromatography mass spectroscopy (LC-MS) together with multivariate curve resolution alternating least squares (MCR-ALS). Cyanobacteria were exposed to two concentrations of As(III) for a week, and lipid extracts were analyzed by ultrahigh-performance liquid chromatography/time-of-flight mass spectrometry in full scan mode. The data obtained were compressed by means of the ROI strategy, and the resulting LC-MS data sets were analyzed by the MCR-ALS method. Comparison of profile peak areas resolved by MCR-ALS in control and exposed samples allowed the discrimination of lipids whose concentrations were changed due to As(III) treatment. The tentative identification of these lipids revealed an important reduction of the levels of some galactolipids such as monogalactosyldiacylglycerol, the pigment chlorophyll a and its degradation product, pheophytin *a*, as well as carotene compounds such as 3-hydroxycarotene and carotene-3,3′-dione, all of these compounds being essential in the photosynthetic process. These results suggested that As(III) induced important changes in the composi- tion of lipids of cyanobacteria, which were able to com- promise their energy production processes.

Keywords LC-MS · MCR-ALS · Cyanobacteria · As(III)

Introduction

Arsenic (As) is a well-known toxic element naturally present in the Earth's crust [1]. Several studies have shown the different responses of microorganisms exposed to this metal, such as the demethylation process in fungi [2], contamination in food [3], and bioaccumulation, transformation, and release in *Microcystis* [4]. Arsenic is considered to be one of the most toxic environmental substances. Both natural and anthropogenic sources contribute to the presence of As in environmental media [5].

Cyanobacteria are the only known prokaryotes that perform oxygen-evolving photosynthesis, commonly found in diverse environments including marine and freshwater, as well as in terrestrial habitats [6]. They are responsible for about 50 % of earthly photosynthesis, and they are the main contributors of the global oxygen cycle. Metals, such as arsenic, are known to exercise a negative influence on cyanobacteria photosynthesis [7, 8]. Cyanobacteria photosynthetic processes are known to occur in cyanobacteria thylakoid membrane [9]. Lipids of cyanobacteria have an important function in the thylakoid membrane structural components, and consequently, they are vital for their life processes [10]. The lipid composition of cyanobacteria thylakoid membrane is mostly formed by galactolipid species of monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG). These lipids are crucial for the stability and activity of photosynthesis systems [11]. A decrease in their concentration has been related to a reduction

* Romà Tauler
roma.tauler@idaea.csic.es

[1] Biological Chemistry and Chemometrics, Universidade Federal do Rio Grande do Norte (UFRN-IQ), Natal 59071-970, Brazil

[2] Institute of Environmental Assessment and Water Diagnostic (IDAEA-CSIC), Jordi Girona 18-26, 08028 Barcelona, Spain

Published online: 16 June 2016

of the photosynthetic activity, changes in chloroplast ultrastructure, and growth arrest [12].

Liquid chromatography coupled to mass spectrometry (LC-MS) is currently being employed in lipid analysis due to its high resolution and sensitivity, enabling the simultaneous identification of multiple constituents in the same analyzed sample [13]. However, this analytical methodology, when applied in complex lipidomic studies, generates a huge amount of data in the analysis of a single biological sample [14]. A recently introduced strategy of selection of mass spectrometry regions of interest (ROIs) is proposed to handle the extremely large size of these LC-MS data sets, using affordable computer resources, without mass accuracy loss [15]. The ROI-based method significantly decreases the size of LC-MS data sets, allowing simultaneous analysis of multiple data sets in the study of the same system under different experimental conditions.

Multivariate curve resolution alternating least squares (MCR-ALS) is a chemometric tool applied to resolve multicomponent responses from unknown unresolved mixtures [16–19]. MCR-ALS has been applied to the resolution of elution/concentration and mass spectra profiles for the different components existing in complex cell lipid mixtures analyzed by chromatographic methods [14]. Recently, it has also been applied successfully to resolve complex omics profiling problems, such as in lipidomics [18, 19] and metabolomics [20] studies.

The aim of this work was to evaluate the changes operating on lipidomic profiles of cyanobacteria (*Anabaena* and *Planktothrix agardhii*) when they were treated with As(III). As_2O_3 was selected as a source of As(III) since it is the main precursor of most of As compounds [21]. To accomplish this objective, a recently introduced data analysis strategy which combines the ROI concept in LC-MS together with MCR-ALS was applied. Changing lipids were identified and preliminary biochemical interpretations of the results are discussed.

Materials and methods

Chemicals and solvents

As_2O_3 was purchased from Sigma. Analytical grade methanol and chloroform were purchased from Merck and Carlo Erba, respectively. HPLC Gradient Grade acetonitrile was purchased from Fischer Chemicals. Lipid standards were obtained from Avanti Polar Lipids.

Cyanobacteria culture and As(III) exposure

Anabaena and *P. agardhii* were both obtained from Culture Collection of Algae and Protozoa (CCAP1403/21 and CCAP1459/11, respectively; SAMS Limited) and were grown aseptically in 500-mL Erlenmeyer flasks containing BG-11 medium (with $NaNO_3$ as a source of nitrogen) [22] at 25 °C under continuous fluorescent white light with an intensity of 15 μmol photons m^{-2} s^{-1}, until an optical density of 0.1–0.2 at 750 nm (OD_{750}) was reached. Then, the cultures were transferred to 50-mL Falcon flasks with different concentrations (0, 1, and 2 mM) of As(III); cells were maintained in the mentioned conditions for 7 days. The choice of these As(III) concentrations is based on a previous toxicity test on cyanobacteria (data not shown) in which the concentration 1 mM started to reduce normal cell growth.

Estimation of protein content

Protein content was measured using Pierce™ BCA Protein Assay Kit (Thermo Scientific). Bovine serum albumin was used as a protein standard. In a 96-well plate, 20 μL of every sample was mixed with 100 μL of the working solution (BCA reagents A and B, 1:50) and then incubated at 37 °C for 1 h. Absorbance was measured at 562 nm (Synergy 2 Multi-Mode Reader, Biotek Instruments, Inc.).

Microscopy and chlorophyll spectra

Cyanobacteria were visualized through a Nikon eclipse 90i microscope using a ×60 oil objective, and images were captured with a Nikon DS-Ri1 digital camera. Chlorophyll spectra were measured in a Synergy 2 Multi-Mode Reader (Biotek Instruments, Inc.) UV–Visible spectrophotometer in the range of 300–800 nm.

Lipid extraction

For every sample, 250 mL of cyanobacteria cultured in 500-mL Erlenmeyer flasks was used. Cyanobacteria cultures were centrifuged (5000 rpm, 20 min, 4 °C) and the superna- tants were discarded to recover cell pellets. The extraction has been performed as previously described [19]. Briefly, 100 μL of Mili-Q water was added to the cyanobacteria pellets and the suspension was transferred to borosilicate glass tubes. This was followed by the addition of 500 μL of chloroform and 250 μL of methanol in each tube. This mixture was fortified with different internal standards of lipids (C18(plasmalogen)-18:1 phosphatidylcholine, C18(plasmalogen)-18:1 phosphatidylethanolamine, C16(plasmalogen) lysophosphatidylcholine, 1,2,3-17:0 triacylglycerol, 1,3-17:0 D5 diacylglycerol, 17:0 monoacylglycerol, 17:0 cholesteryl ester, 16:0 D31-18:1 phosphatidic acid, 16:0 D31-18:1 phosphatidylcholine, 16:0 D31-18:1 phosphatidylethanolamine, 16:0 D 31-18:1 phosphatidylserine, 16:0 D31-18:1 phosphatidylglycerol, 16:0 D31-18:1 phosphatidylinositol, 17:1 lysophosphatidylcholine, 17:1 lysophosphatidylglycerol, 17:1 lysophosphatidylserine, 17:1 lysophosphatidic acid, 17:1 lysophosphatidylcholine,

17:1 lysophosphatidylinositol), 200 pmol each (10 μL of 20 μM stock solutions in absolute ethanol). Next, samples were vortexed and sonicated until they appeared dispersed. In order to increase the efficiency of the extraction, samples were left under agitation overnight in a water bath at 48 °C, as described previously for other lipid extractions [23]. Samples were then cooled and evaporated under N_2 stream and subsequently resuspended in 500 μL of methanol. The resulting suspension was then transferred to 1.5-mL Eppendorf tubes and evaporated again. Then, samples were resuspended in 150 μL of methanol and were centrifuged (10,000 rpm, 3 min) to discard all precipitates from the samples. Finally, 130 μL of each sample supernatant was taken and transferred to UPLC vials for LC injection.

Liquid chromatography mass spectrometry

LC-MS analysis was performed using a Waters Acquity UPLC system connected to a Waters LCT Premier orthogonal accelerated time-of-flight mass spectrometer (Waters) operated in positive ion electrospray ionization mode. Full scan spectra from 50 to 1800 Da were acquired, and acquisition of every MS spectrum was taken approximately 0.2 s.

Mass accuracy and reproducibility were maintained by using an independent reference spray via the LockSpray interference. The analytical column was a 100 × 2.1-mm inner diameter, 1.7-μm C8 Acquity UPLC bridged ethylene hybrid (Waters). The two mobile phases were: phase A, H_2O 2 mM ammonium formate, and phase B, MeOH 1 mM ammonium formate. Both contained 0.2 % of formic acid. The flow rate was 0.3 mL min^{-1}. Tthe gradient of A/B solvents started at 20:80 and changed to 10:90 in 3 min, from 3 to 6 min remained at 10:90, changed to 1:99 in 6 min until minute 15, remained 1:99 until minute 18, and finally returned to the initial conditions until minute 20. The column was held at 30 °C.

Chemometric data analysis

Raw data files obtained from the LC-MS analysis of every sample were transformed to CDF format by Databridge function of MassLynx™ V. 4.1 software. LC-MS data were uploaded into the MATLAB workspace environment (version R2012b, The Mathworks Inc.) using the functions mzcdfread and mzcdf2peaks of the Bioinformatics Toolbox (The Mathworks Inc.). LC-MS data were arranged and aligned according to their *m*/*z* values in a square regular data matrix, with retention times in the rows and selected *m*/*z* values in the columns. Different strategies were adapted to build such a data matrix. We have recently proposed a new protocol based on the selection of the ROIs [15]. This strategy allows for the selection of *m*/*z* traces whose intensity signals are higher than preselected threshold values (not noise or spurious measurements), and it takes advantage of the sparsity of the measured MS data and of their mass accuracy. Using this approach, full-scan chromatographic MS matrices with the highest experimental mass accuracy are obtained with low storage requirements. A relatively low number (200–500) of high-resolution MS (intensity, *m*/*z*) signals were finally taken into account, with more than 100-fold computer storage reduction (Fig. 1).

Every MS-ROI data matrix was then normalized taking into account the areas of the internal standards added and the protein content measured for each sample. This normalization was done by multiplying each matrix by a factor obtained as follows: (200 × mg protein)/mean area of standards, where 200 refers to the picomoles of stan- dards added in the lipid extraction. After normalization, individual MS-ROI matrices obtained in the analysis of every *Anabaena* and *P. agardhii* samples were used to build four different augmented matrices. In Fig. 1, the exemplary construction of a MS-ROI augmented matrix was given. Since MS-ROI individual data matrices have different number of ROI *m*/*z* values, a preliminary step of ROI rearrangement to consider all those ROI *m*/*z* values (common and not common) with significant traces (retention times where MS intensity values are higher than the threshold values at the considered ROI *m*/*z* value) is performed. In case an individual matrix does not have a particular significant trace at this ROI *m*/*z* value, low random intensity values at the noise level are assigned. See Gorrochategui et al. [15] for more details about MS-ROI augmented matrix building. As shown in Fig. 1, the MS-ROI augmented data matrix D_{aug} is built with four control samples and four treated samples, and the final dimensions of this augmented matrix are the total number of retention times considered in the eight chromatographic runs times the number of finally considered *m*/*z* ROI values. The first MS-ROI augmented matrix was built with the four non-treated (control) and the four 1-mM As(III)-treated *Anabaena* samples. The second MS-ROI augmented matrix had the four non-treated (control) and the four 2-mM As(III)-treated *Anabaena* samples. The third MS-ROI augmented matrix had the four non-treated (control) and the four 1-mM As(III)-treated *P. agardhii* samples. Finally, the fourth MS-ROI augmented matrix was built with the four non-treated (control) and the four 2-mM As(III)-treated *P. agardhii* samples. In order to perform this matrix augmentation column-wise, the columns of every ROI data matrix should have the same *m*/*z* ROI values. See Gorrochategui et al. [15] for details on how to implement this in practice.

In this work, the MCR-ALS method [16, 17] has been used for the analysis of the augmented ROI data matrices described above. Since the MCR-ALS method has been described

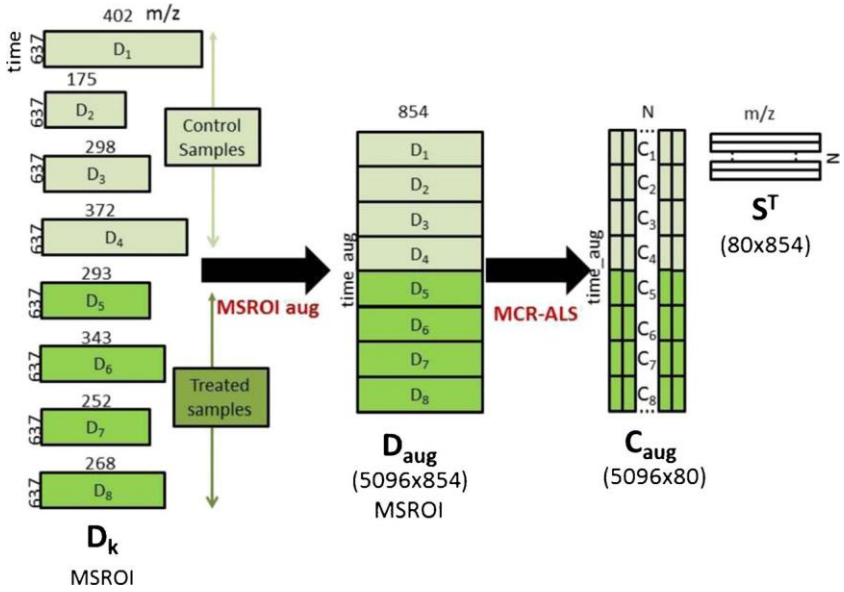


Fig. 1 Example of MS-ROI data matrix augmentation and application of MCR-ALS to the simultaneous analysis of eight individual MS-ROI data matrices: four control samples and four treated samples. See [15] and Eq. 2

elsewhere [15–17], only a brief explanation is given below. This method is based on the fulfillment of a bilinear measurement model, which assumes that the measured MS spectra along the chromatographic separation are the sum of the MS spectra of the lipids of the investigated system (plus some background, solvent, and instrumental noise contributions) weighted by their relative concentrations. This bilinear model can be described by the simple data matrix equation

$$D = CS^T + E \quad (1)$$

In the case of a single LC-MS-ROI data matrix, D contains the MS intensity signals at the selected *m*/*z* ROI values (columns) and retention times (rows). C has the elution or concentration profiles of resolved components, and S^T has the resolved pure MS-ROI spectra of these components. E has the non-explained variance and experimental error. MCR-ALS solves Eq. 1 by an Alternating Least Squares algorithm which calculates iteratively the concentration, C, and the pure MS-ROI spectra, S^T, matrices fitting optimally the experimental data matrix, D. MCR-ALS optimization requires the initial proposal of a number of components and of an initial estimate of either the C or S^T matrix. A sufficiently large number of components (for instance, 80 or more) can be initially selected to explain most of the data variance (e.g., more than 90 % of the total data variance) and to properly describe the experimental data features (chromatographic peaks). This number of components can also include some components that are not strictly related to the lipid signals but to the solvent, background, or other detector and instrumental signal contributions. Initial estimates of either the C or S^T matrix are obtained from the more distinct rows (MS spectra) or columns (chromatograms) of the data matrix. These are good initial estimates for MCR-ALS due to the high selectivity of the LC-MS data. Once the number of components and initial estimates of either C or S^T are provided, during the ALS optimization, various constraints are applied to model appropriately the shapes of the C and S^T profiles [24]. For ROI LC-MS data, non-negativity was applied to ensure the concentration and spectral profiles to be positive. Also, during the ALS optimization, the MS spectra were normalized to have all maximum intensity equal to 1, to avoid scale indeterminacies (intensity ambiguity) and to give relative quantitative information in the concentration profiles. Since the MS-ROI spectra are sparse and very selective (a lot of zero values), rotation ambiguities [25] are practically not present in this case. Fitting MCR-ALS results can be performed by using the explained data variance calculated as: $R^2(\%) = \sum d^2 / \sum d^2 \times 100$, where d_{ij} and d^*_{ij} are the corresponding values for $i = 1,\ldots,I$ (number of data matrix rows, retention values) and $j = 1,\ldots,J$ (number of data matrix columns, number of *m*/*z* ROI values), d_{ij} is the MS-ROI individual intensity value in the experimental matrix D, and d^*_{ij} is the corresponding value calculated by MCR-ALS using the optimal C and S^T matrices finally obtained, as has been shown in previous works [16, 17].

The MCR-ALS bilinear model can be extended to the simultaneous analysis of multiple individual data matrices using the column-wise augmentation strategy described in Fig. 1. In this particular case, the simultaneous analysis of multiple MS-ROI data matrices coming from the same type of cyanobacteria culture (either *Anabaena* or *P. agardhii*) under different conditions (control samples and As(III)-treated samples at two concentrations, 1 and 2 mM) was performed using the appropriate MS-ROI column-wise augmented data matrix (see Fig. 1). Equation 1 for MCR analysis of a single data matrix is now extended

to Eq. 2 for MCR analysis of the new column-wise MS-ROI augmented data matrices [16, 17].

$$\begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_K \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix}^T \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_k \end{bmatrix}^T \quad (2)$$

$$\begin{bmatrix} D_3 \\ \vdots \\ D_K \end{bmatrix} = \begin{bmatrix} C_3 \\ \vdots \\ C_k \end{bmatrix} S \quad \text{or} \quad \begin{bmatrix} E_3 \\ \vdots \\ E_k \end{bmatrix} = C_{aug} S + E_{aug}$$

In this equation, the column-wise augmented MS-ROI data matrix, D_{aug}, is decomposed by the bilinear model into the product of a column-wise augmented matrix, C_{aug}, having the elution profiles of the resolved components in every individual MS-ROI data matrix, by the MS spectra matrix, S^T, having the pure MS-ROI spectra of these components. E_{aug} is the corresponding matrix which will have the model unexplained variance (for the considered number of components) and experimental error. It is important to remark here that the proposed MCR-ALS data analysis strategy does not imply any shift nor shape assumption nor correction of the chromatographic elution profiles of the same component in the different chromatographic runs. It only needs that the *m*/*z* values match among the different MS spectra included in D or D_{aug} and that every component can be defined by a unique pure MS spectrum. This data analysis procedure has high flexibility and resolution power, it takes profit of the high resolution of the MS detector, and it allows for the lack of reproducibility of chromatographic separation conditions.

Once the MCR-ALS results are obtained, the C_{aug} and S^T matrices are examined in detail to extract qual- itative (identification) and quantitative information and to allow the interpretation of the investigated effects (by As(III) treatment). The elution/concentration profiles resolved by MCR-ALS for every component are in the columns of the C_{aug} matrix for all chromatographic runs simultaneously analyzed. Comparison of the areas of these elution profiles provides the relative quantitative information about a particular resolved component in the different analyzed samples. Comparison of the areas between the control and As(III)-treated samples allows for the statistical assessment of the produced effects at a particular significance level (*p* values below 0.05). Only those components showing statistically significant differ- ences between their concentrations (peak areas) in the control and As(III)-treated samples are further consid- ered for their chemical compound identification. This identification is performed looking for the MCR-ALS MS-ROI-resolved spectrum corresponding to the same component in the S^T matrix. The *m*/*z* ROI values of these spectra have full-scan MS mass accuracy and are used for compound identification using the Human Metabolome Database (http://www.hmdb.ca) and Lipid Maps (http://www.lipidmaps.org) online databases.

Lipids are assigned from *m*/*z* values associated to smaller delta values (difference between query mass and adduct mass), and they were compared with those found in previous works.

Results

Morphological changes and chlorophyll content of cyanobacteria exposed to arsenic

First, morphological changes of cyanobacteria (*Anabaena* and *P. agardhii*) were assessed after 7 days of arsenic exposure. As depicted in the optical microscope images of Fig. 2, cyanobacteria filaments showed a visible decrease of green coloration along with the increasing concentration of As(III). At 2 mM As(III), *P. agardhii* species presented considerable fading of their green pigmentation while *Anabaena* species showed complete discoloration.

Since green pigmentation in cyanobacteria is associated with the composition of photosynthetic pigments like chlorophyll [6], changes in this green pigmentation were first analyzed by UV–Visible spectrometry (Fig. 3).

The absorbance spectra of treated cyanobacteria were recorded. The results showed that absorption bands between 400 and 500 nm were practically eliminated after arsenic treatment, especially in the case of *Anabaena* (Fig. 3A). Considering that the main peak of chlorophyll is around 450 nm [26], the results suggested that the decrease of green coloration was related to chlorophyll degradation. An important decline of chlorophyll under arsenic exposure has already been described in previous works [4]. Rocchetta et al. [27] also showed that algal cells exposed to heavy metals such as copper and chromium suffered from multiple biochemical alterations affecting the photosynthesis. In this work, in order to study the changes in lipid composition due to As(III) treatment, an exhaustive untargeted lipidomic study was performed.

Chemometric analysis of LC-MS data of cyanobacteria exposed to As(III)

The effects of different concentrations (0, 1, and 2 mM) of arsenic after 7 days of exposure on the two cyanobacteria (*Anabaena* and *P. agardhii*) cultures were investigated by LC-MS in positive ion mode. LC-MS raw data were first normalized (using internal standards and total protein content) and compressed according to the ROI strategy (see ʙMaterials and methods^). As mentioned, the use of the ROI strategy allowed substantial reduction of computer storage and processing time without losing mass accuracy [15].

In Fig. 4, the positive ion mode full-scan LC-MS chromatograms (after selection of the ROIs) are compared for

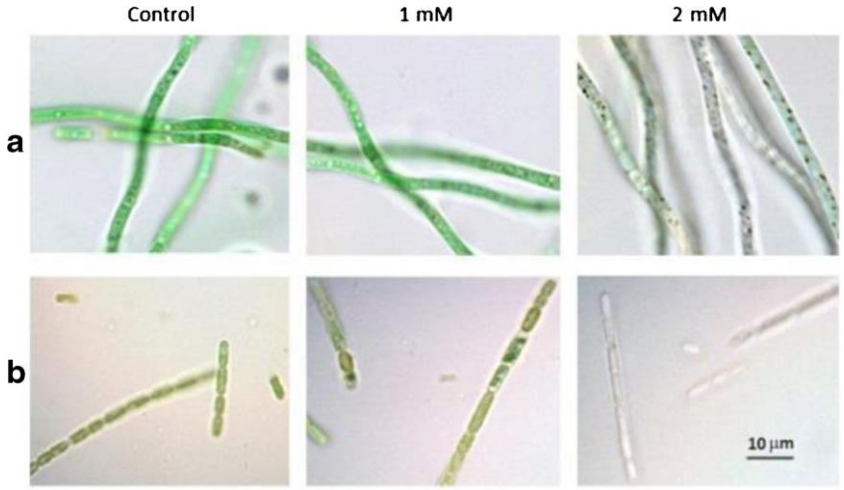


Fig. 2 Light microscope pictures of cyanobacteria filaments under As(III) treatment. *Anabaena* (a) and *P. agardhii* (b) cyanobacteria species incubated for 7 days with 0, 1, and 2 mM of As(III) in Blue-Green medium

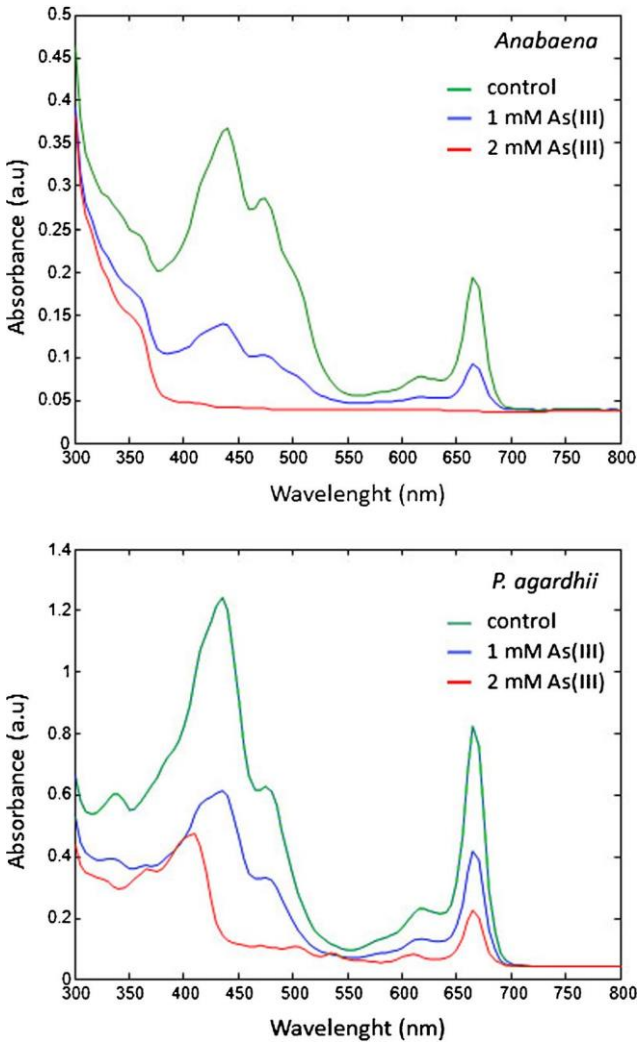


Fig. 3 Chlorophyll absorption spectra of As(III)-treated cyanobacteria. *Anabaena* (a) and *P. agardhii* (b) exposed for 7 days to 0, 1, and 2 mM As(III) (*brown*, *blue*, and *yellow lines*, respectively)

samples extracted from *Anabaena* cultures without any As(III) treatment (Fig. 4A) to those obtained after 7 days of exposure with 1 mM As(III) (Fig. 4B). Distinct features are clearly noticed in these chromatograms already showing the effects of metal treatment.

In order to investigate the effects on lipid concentration changes in both cyanobacteria cultures due to the As(III) treatment at 1 and 2 mM, four different MS-ROI column-wise augmented data matrices were built. Each new column-wise augmented data matrix contained eight MS-ROI individual data matrices, the first four being the control replicates and the last four the treatment replicates, as explained in ʙMaterials and methods.^ In Fig. 5, the first MS-ROI augmented matrix for *Anabaena* at 1-mM As(III) treatment is shown as an example.

The eight chromatographic sections distinguished correspond to the eight investigated samples. As shown in the picture, plots of the LC-MS-ROI data for the control samples (four regions on the left) had higher intensity signals than the plots of the four 1-mM As(III)-treated samples (four regions on the right), suggesting an important loss of lipid species concentrations due to the As(III) treatment.

The augmented LC-MS-ROI data matrices of four control and four treated cyanobacteria samples were then analyzed by MCR-ALS (see Fig. 1), using non-negativity constraints, to resolve elution and spectral profiles of the sample lipid constituents. MS-resolved pure spectra (S^T) were also normalized to have the same equal intensity, transferring all relative quantitative information to elution profiles (C). MCR-ALS was performed using a large enough number of components (80) to explain enough data variance and to assure the resolution of a sufficiently large number of elution and MS spectra profiles with reasonable chemical shapes. In the four cases, *f*, the explained variances were between 93 and 95 %. A small number of the MCR-ALS-resolved components could be related to solvent and background contributions, whereas all other had

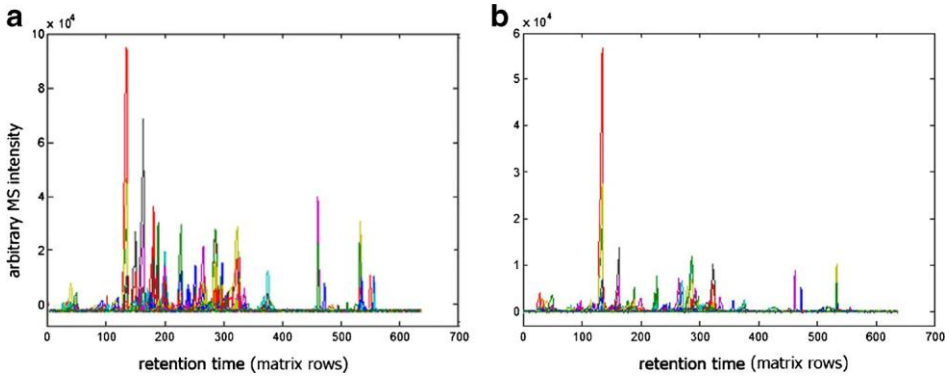


Fig. 4 LC-MS-ROI chromatogram of a single sample (from one single-data matrix). *Anabaena* control (a) and 1-mM As(III)-treated (b) samples. In the *X*-axis are the number of retention times in the analysis of one sample (giving one data matrix with 637 rows or retention times; see Fig. 1)

reasonable chromatographic and spectral shapes and features. Only these ones were investigated in more detail, and their concentration changes between the control and As(III)-treated samples were estimated.

Once the elution profiles corresponding to the same component in the control and treated samples of the same augmented MS-ROI matrix were resolved, they were further analyzed to assess the possible concentration changes due to As(III) treatment. As explained in ʙMaterials and methods,^ this was performed by comparison of the peak areas and/or heights of these two types of samples. A statistical Student's *t* test was used in every case to assess for the statistical difference between the peak areas/heights of the elution profiles of the same component in the control and treated samples. As an example, the MCR-ALS-resolved elution profiles and mass spectra for two components that presented significant changes between controls and treatments are represented in Fig. 6 for both *Anabaena* 1-mM and *P. agardhii* 1-mM augmented matrices. On one hand, in *Anabaena* 1-mM samples on the left of Fig. 6, component 79 exhibited an important decrease in the treated samples (four sections on the right), as shown in Fig. 6A, with an *m*/*z* ROI value of 800.6217 (Fig. 6B), which was further identified as MGDG(36:2). In the case of *P. agardhii* 1-mM samples on the right of Fig. 6, component 21 presented a similar decreasing profile in the treated samples (Fig. 6E). With an *m*/*z* ROI value of 551.4229 (Fig. 6F), this molecule was further identified as 3-*cis*-hydroxy-carotene. Bar plots of the areas and the heights of the resolved peaks in both examples represented in both cases (Fig. 6C, D, G, H) also confirmed the changes produced by As(III) treatment.

In Tables 1 and 2, lipids showing significant changes according to the *t* test (at the $p < 0.05$ significance level) are

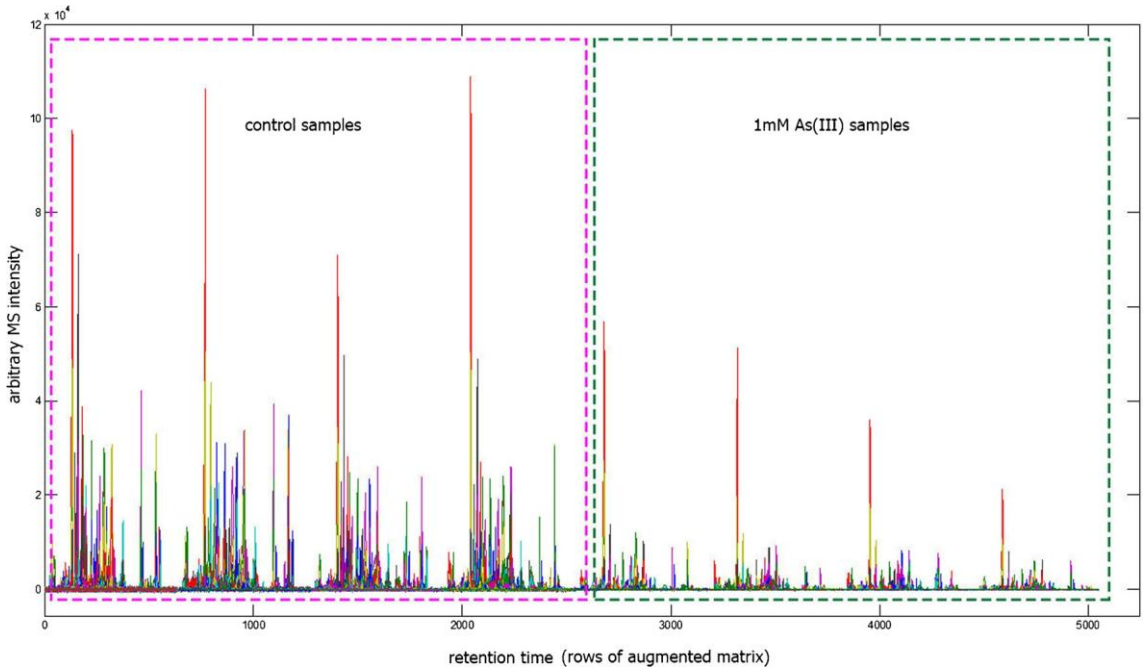


Fig. 5 LC-MS-ROI chromatograms of multiple samples (from one column-wise augmented data matrix). On the *left* are four *Anabaena* control samples (section marked on *pink slashed square*). On the *right* are four 1-mM As(III)-treated samples (section marked on *green slashed square*). In the *X*-axis are the total number of retention times in the analysis of the eight samples (giving one column-wise augmented data matrix with $8 \times 637 = 5096$ rows or retention times; see Fig. 1)

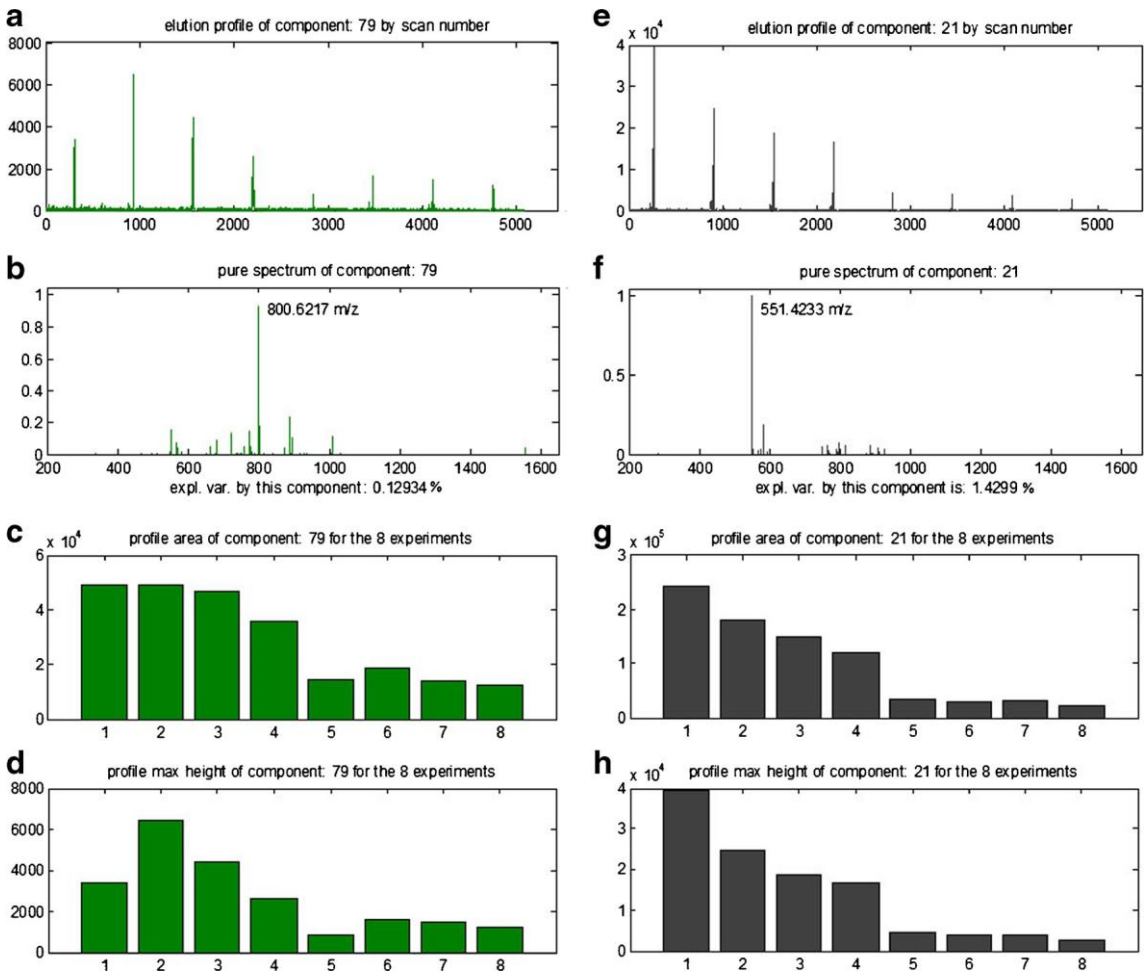


Fig. 6 Examples of results obtained in the MCR-ALS simultaneous analysis of multiple LC-MS-ROI chromatograms (from the MCR-ALS analysis of column-wise augmented data matrices having four control samples and four 1-mM As(III)-treated samples; see Eq. 2). Results are given for component 79 (*left*) in the analysis of *Anabaena* sample and for component 21 (*right*) in the analysis of *P. agardhii* sample. a, c MCR-ALS-resolved elution profiles (C_{aug} in Eq. 2). b, f MCR-ALS-resolved MS pure spectra (S^T in Eq. 2) with the *m/z* ROI values at their intensity maxima. c, g Bar plots of the integrated peaks areas of each replicate. d, h Bar plots of the peaks heights of each replicate

listed with their retention times, measured *m/z*, explained variance (R^2), identified compounds, adduct, calculated *m/z*, mass error (in parts per million), area and height fold changes, and *p* values. These lipids are given and summarized for *Anabaena* and *P. agardhii* cyanobacteria cultures treated with 1 and 2 mM As(III), respectively. Most of the listed lipids are associated with cyanobacteria membranes and with pigments related to the photosynthesis process. Interestingly, at 1-mM As(III) treatment (Table 1), a 4- to 5-fold decrease of several MGDG species has been detected for both cyanobacteria species. These galactolipids are associated with cyanobacteria membrane and are known to be involved in thylakoid membrane packing and to provide a mechanism for the separation of the components of the photosystem within the thylakoid membrane [12]. In addition, as mentioned before, we observed a significant loss in the concentrations of pigments related to the photosynthetic system in 1-mM *Anabaena* samples, such as chlorophyll a, and its sub-product, pheophytin a. Also, we found a large 4-fold decrease of some carotene compounds such as 3-*cis*-hydroxy-carotene and carotene-3,3′-dione. Carotenes have key roles in light absorption for use in the photosynthesis process, as well as a protective role of chlorophyll from photodamage [28]. In *P. agardhii*, we also found a significant 5-fold decrease of 3-*cis*-hydroxy-carotene, as has been shown in the example of Fig. 6. Another interesting feature observed is the decrease of triacylglycerol (51:0), only found in *Anabaena* samples.

Similar results were obtained for 2-mM As(III) treatment (Table 2). In this case, the decrease of MGDGs that showed a reduction in their levels at 1 mM was more intense in both cyanobacteria. Also, a new MGDG appeared reduced at this concentration. In *Anabaena* samples, the decline of pheophytin and carotene concentrations was more pronounced (more than 10-fold). For *P. agardhii* samples, pheophytin a appeared reduced in the treated samples, and the reduction of 3-*cis*-hydroxy-carotene levels observed at

Table 1 Summary of molecules that showed significant changes between untreated and 1-mM As(III)-treated samples for both *Anabaena* and *P. agardhii* using the proposed LC-MS-ROI-MCR-ALS methodology

RT	Measured *m/z*	R^2	Identified compound	Adduct	calculated *m/z*	Mass error (ppm)	Area fold change	Height fold change	Area *p* value	Height *p* value
Anabaena 1 mM										
9.12	551.4229	3.0424	3-*cis*-Hydroxy-b.e-carotene *b	M+H	551.4253	4.4	0.21	0.25	0.002	0.001
5.17	565.4057	0.1326	E.e-carotene-3.3′-dione *b	M+H	565.4040	3.0	0.28	0.15	0.041	0.029
12.93	567.4538	0.7971	Unknown	–	–	–	0.18	0.16	0.003	0.002
3.49	583.4128	0.1129	Diatoxanthin 3.6-epoxide *b	M+H	583.4146	3.0	0.21	0.15	0.001	0.007
11.55	619.6013	0.1602	Unknown	–	–	–	0.38	0.34	0.001	0.001
1.32	626.1850	0.2677	Unknown	–	–	–	0.19	0.15	0.015	0.035
4.58	675.6788	35.879	Unknown	–	–	–	0.40	0.43	0.026	0.019
16.29	686.6451	0.1356	Unknown	–			0.39	0.42	0.004	0.001
6.16	766.5462	0.815	MGDG(34:5) *a	$[M+NH_4]$	766.546	0.3	0.18	0.11	0.016	0.022
10.21	800.6217	0.1293	MGDG(36:2) *a	$[M+NH_4]$	800.6247	3.7	0.33	0.30	0.001	0.027
18.4	866.8210	1.1585	TG(51:0)	$[M+NH_4]$	866.818	3.5	0.11	0.09	0.017	0.032
11.11	871.5736	4.8752	Pheophytin a*b	[M+Na]	871.5737	0.2	0.14	0.21	0.039	0.001
9.84	893.5477	2.2746	Chlorophyll a*b	[M+H]	893.542589	5.7	0.21	0.12	0.018	0.007
P. agardhii 1 mM										
3.96	310.3083	0.0651	Unknown	–	–	–	0.3	0.4	0.0093	0.0116
9.15	551.4233	1.4298	3-*cis*-Hydroxy-b.e-carotene *b	M+H	551.4253	3.6	0.2	0.2	0.0101	0.0224
12.91	567.4542	0.5982	Unknown	–	–	–	0.2	0.2	0.0556	0.0086
1.66	594.1597	0.165	Unknown	–	–	–	0.2	0.3	0.0336	0.0067
11.55	619.6026	0.1348	Unknown	-	-	-	0.3	0.2	0.0078	0.0065
8.06	763.5351	0.3065	MGDG(35:5) *a	[M+H]	763.536	1.2	0.2	0.2	0.0172	0.0026
6.88	768.566	9.202	MGDG(34:4) *a	$[M+NH_4]$	768.5627	4.3	0.2	0.2	0.0071	0.0071
7.22	794.5805	0.24	MGDG(36:5) *a	$[M+NH_4]$	794.5776	3.6	0.3	0.1	0.0308	0.0115

Molecules are ordered by their *m/z* values. Lipids related to cyanobacteria membranes and pigments related to the photosynthesis process have been marked with *a or *b, respectively. Fold change values >1 mean that the concentrations of the compound increased after As(III) treatment. Fold change values between 0 and 1 mean a decrease of compound concentration

Table 2 Summary of molecules that showed significant changes between untreated and 2-mM As(III)-treated samples for both *Anabaena* and *P. agardhii* using the proposed UPLC-MS-ROI-MCR-ALS methodology

RT	Measured *m/z*	R^2	Identified compound	Adduct	Calculated *m/z*	Mass error (ppm)	Area fold change	Height fold change	Area *p* value	Height *p* value
Anabaena 2 mM										
9.21	550.4159	0.4266	Unknown	–	–	–	0.03	0.010	0.013	0.034
9.12	551.4229	3.0891	3-*cis*-Hydroxy-b.e-carotene *b	M+H	551.4253	4.4	0.02	0.007	0.002	0.001
5.17	565.4064	0.1376	E.e-carotene-3.3′-dione *b	M+H	565.404007	4.2	0.09	0.039	0.038	0.027
12.93	567.4544	0.7739	Unknown	–	–	–	0.01	0.001	0.003	0.002
1.6	594.1601	0.372	Unknown	–	–	–	0.22	0.220	0.001	0.005
11.55	619.6033	0.1602	Unknown	–	–	–	0.26	0.168	0.001	0.001
1.38	626.1857	0.2845	Unknown	–	–	–	0.21	0.251	0.035	0.024
4.58	675.6788	35.492	Unknown	–	–	–	0.27	0.261	0.017	0.008
16.29	686.6452	0.1242	Unknown	–	–	–	0.10	0.049	0.007	0.003
6.16	766.5473	0.8663	MGDG(34:5)*a	$[M+NH_4]$	766.5463	0.1	0.01	0.001	0.007	0.001
8.68	772.5933	0.5143	MGDG(34:2)*a	$[M+NH_4]$	772.5933	0.0	0.02	0.003	0.010	0.041
10.21	800.6139	0.1552	MGDG(36:2)*a	$[M+NH_4]$	800.6247	13.5	0.13	0.082	0.001	0.014
11.11	871.5744	4.5519	Pheophytin a*b	[M+H]	871.5737	−0.8	0.01	0.001	0.033	0.001
P. agardhii 2 mM										
3.96	310.3089	0.0631	Unknown	–	–	–	0.10	0.14	0.0098	0.0097
1.07	495.1252	0.089	Unknown	–	–	–	0.25	0.26	0.0246	0.04
12.95	536.4350	0.3259	Unknown	–	–	–	0.03	0.03	0.0105	0.0126
9.13	551.4242	3.0649	3-*cis*-Hydroxy-b.e-carotene *b	M+H	551.4253	2.0	0.03	0.01	0.0032	0.0087
12.92	567.4532	0.573	Unknown	–	–	–	0.01	0.00	0.0366	0.0076
1.66	594.1595	0.172	Unknown	–	–	–	0.17	0.17	0.0247	0.013
11.55	619.6035	0.1387	Unknown	–	–	–	0.13	0.14	0.0064	0.0057
10.09	680.4791	0.2172	Unknown	–	–	–	0.08	0.04	0.0488	0.0457
8.07	763.5356	0.3116	MGDG(35:5)*a	[M+H]	763.5360	0.6	0.19	0.16	0.031	0.0132
6.89	768.5644	9.7156	MGDG(34:4)*a	$[M+NH_4]$	768.562	3.1	0.04	0.07	0.0033	0.0044
7.75	770.5815	6.4446	MGDG(34:3)*a	$[M+NH_4]$	770.5776	5.1	0.11	0.12	0.0601	0.0471
6.48	792.5657	4.4156	MGDG(36:6)*a	$[M+NH_4]$	792.5620	4.6	0.09	0.09	0.0656	0.0476
7.23	794.5806	0.2492	MGDG(36:5)*a	$[M+NH_4]$	794.5785	2.5	0.07	0.06	0.032	0.0182
11.15	871.5709	8.1386	Pheophytin a*b	[M+Na]	871.5737	3.3	0.06	0.06	0.0427	0.0383

Molecules are ordered by their *m/z* values. Lipids related to cyanobacteria membranes and with pigments related to the photosynthesis process have been marked with *a or *b, respectively. Fold change values >1 mean that the concentrations of the compound increased after As(III) treatment. Fold change values between 0 and 1 mean a decrease of compound concentration

1 mM was more severe (30-fold). The area fold changes of the more significant features at 1- and 2-mM As(III) treatments for both cyanobacteria species are compared and summarized in Fig. 7.

Discussion

A LC-MS-ROI-MCR-ALS methodology has been proposed to perform the untargeted lipidomic analysis of two cyanobacteria species (*Anabaena* and *P. agardhii*) af- ter 7 days of As(III) exposure at 1 and 2 mM. The ap- plication of the ROI strategy allowed the compression of the large amount of data from full-scan LC-MS data with- out any loss of information. Comparison of the elution profiles resolved by MCR-ALS provided information about which lipids more significantly changed their concen- tration after cyanobacteria exposure to different doses of As(III). The results of these analyses revealed strong affecta- tions on the concentrations of some chemical constituents in- volved in photosynthesis, which is the essential survival pro- cess of cyanobacteria. Although the changes were not identi- cal for both species of cyanobacteria, the families of com- pounds affected were similar and followed the same concen- tration reduction pattern. Among the compounds involved in As(II) toxicity, the most significant changes were on the concentration levels of some MGDG galactolipid species. These compounds are the most abundant lipids in all photosynthetic tissues, including those of higher plants, algae, and cyanobacteria. They are directly linked to cyanobacteria membrane, with two main functions: to allow efficient packing of the thylakoid membrane and to separate the components of the photosystem within the thylakoid membrane [10]. In a previous study, it has been shown that a low ratio of MGDG to DGDG is crucial to maintain the stability and proper behavior of protein membranes [11]. In addition to galactolipids, a significant drop in the concentration of photosynthetic pigments was observed. On one hand, in *Anabaena*, chlorophyll was almost completely suppressed, as well as its sub-product, pheophytin a. The concentration loss of pheophytin a was also observed in the case of *P. agardhii* at 2 mM of As(III) treatment. On the other hand, the concentrations of some carotene compounds such as 3-*cis*-hydroxy-carotene and caro- tene-3,3′-dione were importantly reduced. Whereas chlo- rophyll a has a key role in absorbing energy from light and acting as a primary electron donor in the electron transport chain of photosynthesis, carotenes behave as allied compounds to chlorophyll in the photosynthetic process as they are able to absorb light energy and transfer it to chlorophylls and also to protect chlorophyll from photodamage. This loss of pigmentation is in agreement with the observed discoloration of cyanobacteria and with the alterations of the UV spectra of the chlorophyll of both species (the results suggest that chlorophyll a is degraded to pheophytin a, which in turn is also reduced). The strong reduction of pigment concentration levels together with the reduction of ga- lactolipid concentration levels indicated that As(III) exposure induced a collapse of the photosynthetic process.

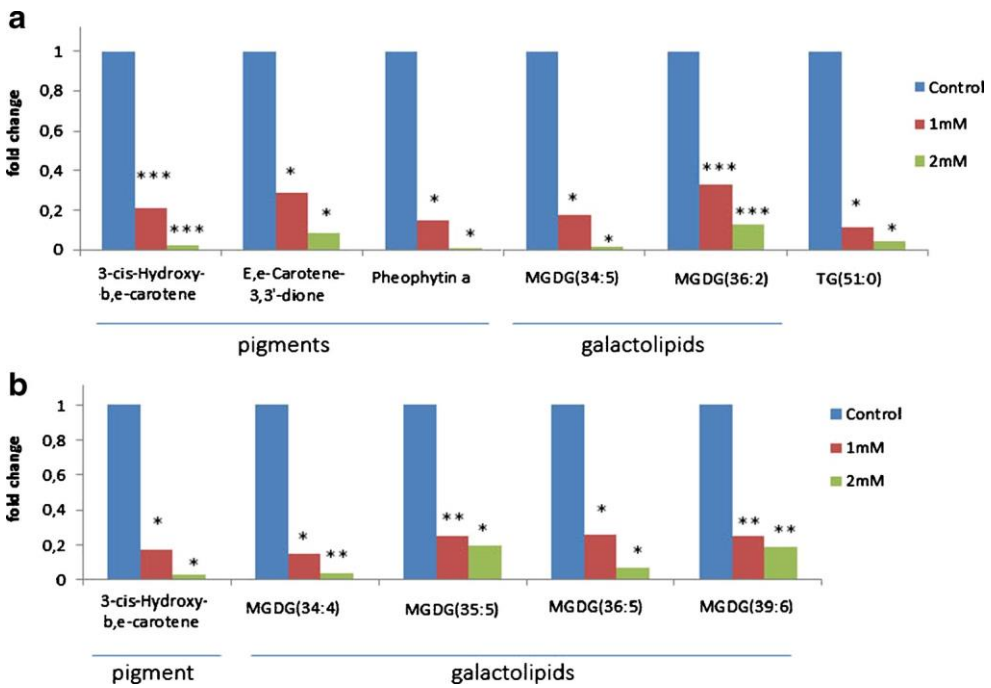


Fig. 7 Areas fold change representation of the common changing lipids observed at 1 and 2 mM of As(III) treatment for the two cyanobacteria species. Common changes between the two As(III) concentrations for *Anabaena* (a) and P. *agardhii* (b). The statistical significance of each bar is represented with *asterisks*: *$p < 0.05$, **$p < 0.01$, ***$p < 0.005$

Altogether, a general disruption of the photosynthetic process has been detected after As(III) treatment in both cyanobacteria species, which reveals the strong toxicity induced by this metal, present in a variety of contamination sources.

Conclusions

The combination of a non-target methodology based on the use of the regions of interest (ROI) and multivariate curve resolution analysis on LC-MS full-scan data has been shown to be a powerful approach for the investiga- tion of major lipidomic changes in cyanobacteria cultures when exposed to the highly toxic metal As(III). Lipid composition changes on *Anabaena* and *P. agardhii* cyanobacteria cultures when exposed for 7 days to As(III) showed that these changes were mainly associated with changes in the thylakoid cyanobacteria membrane composition (MGDG) and with photosynthetic compounds such as chlorophyll a and some carotenoids. Overall, the results of this study suggested that the presence of increasing concentrations of As(III) induced the collapse of the whole cyanobacteria photosynthetic process.

Acknowledgments The authors would like to acknowledge the financial support from the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES) and Brazilian National Council for Scientific and Technological Development (CNPq) for a 1-year fellowship to Aline Marques in the Chemometrics Research Group at IDAEA-CSIC, Barcelona, Spain. K.M.G. Lima acknowledges the CNPq Grant (305962/2014-0) for financial support. This work was funded by grants from CNPq/Capes project (grant 070/2012) and by the CHEMAGEB project (FP/2007-2013)/ERC Grant Agreement no. 320737). The authors also thank Dr. Benjamín Piña and Claudia Rivetti (IDEAE/CSIC) for providing the cyanobacteria species control.

Compliance with ethical standards All authors have accepted the content of this paper and the principles of ethical and professional conduct. Sources of funding are given in the Acknowledgements section. This research has no conflict of interest and has not involved humans or animals

References

1. Ferrari SG, Silva PG, González DM, Navoni JA, Silva HJ. Arsenic tolerance of cyanobacterial strains with potential use in biotechnology. Rev Argent Microbiol. 2013;45(03):174–9.
2. Su S, Zeng X, Feng Q, Bai L, Zhang L, Jiang S, et al. Demethylation of arsenic limits its volatilization in fungi. Environ Pollut. 2015;204:141–4.
3. Phan K, Sthiannopkao S, Heng S, Phan S, Huoy L, Wong MH, et al. Arsenic contamination in the food chain and its risk assessment of populations residing in the Mekong River basin of Cambodia. J Hazard Mater. 2013;262:1064–71.
4. Wang Z, Luo Z, Yan C. Accumulation, transformation, and release of inorganic arsenic by the freshwater cyanobacterium *Microcystis aeruginosa*. Environ Sci Pollut Res. 2013;20(10): 7286–95.
5. Joseph T, Dubey B, McBean EA. A critical review of arsenic exposures for Bangladeshi adults. Sci Total Environ. 2015;527–528: 540–51.
6. de Oliveira VE, Neves Miranda MAC, Soares MCS, Edwards HGM, de Oliveira LFC. Study of carotenoids in cyanobacteria by Raman spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc. 2015;150:373–80.
7. Yin X-X, Chen J, Qin J, Sun G-X, Rosen BP, Zhu Y-G. Biotransformation and volatilization of arsenic by three photosynthetic cyanobacteria. Plant Physiol. 2011;156(3): 1631–8.
8. Cassier-Chauvat C, Chauvat F. Responses to oxidative and heavy metal stresses in cyanobacteria: recent advances. Int J Mol Sci. 2015;16(1):871.
9. Vermaas WFJ. Photosynthesis and respiration in cyanobacteria. In: Encyclopedia of life sciences. New York: Wiley; 2001. doi:10.1038/npg.els.0001670.
10. Chen D, Yan X, Xu J, Su X, Li L. Lipidomic profiling and discovery of lipid biomarkers in *Stephanodiscus* sp. under cold stress. Metabolomics. 2013;9(5):949–59.
11. Hölzl G, Dörmann P. Structure and function of glycoglycerolipids in plants and bacteria. Prog Lipid Res. 2007;46(5):225–43.
12. Dörmann P, Benning C. Galactolipids rule in seed plants. Trends Plant Sci. 2002;7(3):112–8.
13. Li S, Xu J, Jiang Y, Zhou C, Yu X, Zhong Y, et al. Lipidomic analysis can distinguish between two morpholog- ically similar strains of *Nannochloropsis oceanica*. J Phycol. 2015;51(2):264–76.
14. Lima KMG, Bedia C, Tauler R. A non-target chemometric strategy applied to UPLC-MS sphingolipid analysis of a cell line exposed to chlorpyrifos pesticide: a feasibility study. Microchem J. 2014;117: 255–61.
15. Gorrochategui E, Jaumot J, Tauler R. A protocol for LC-MS metabolomic data processing using chemometric tools. Protocol Exchange. 2015. doi:10.1038/protex.2015.102.
16. Jaumot J, Gargallo R, de Juan A, Tauler R. A graphical user-friendly interface for MCR-ALS: a new tool for multi- variate curve resolution in MATLAB. Chemom Intell Lab Syst. 2005;76(1):101–10.
17. Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: new features and applications. Chemom Intell Lab Syst. 2015;140:1–12.
18. Gorrochategui E, Casas J, Porte C, Lacorte S, Tauler R. Chemometric strategy for untargeted lipidomics: biomarker detection and identification in stressed human placental cells. Anal Chim Acta. 2015;854:20–33.
19. Bedia C, Dalmau N, Jaumot J, Tauler R. Phenotypic malignant changes and untargeted lipidomic analysis of long-term exposed prostate cancer cells to endocrine disruptors. Environ Res. 2015;140:18–31.
20. Farrés M, Piña B, Tauler R. Chemometric evaluation of *Saccharomyces cerevisiae* metabolic profiles using LC–MS. Metabolomics. 2015;11(1):210.–24
21. Grund SS, Hanusch K, Wolf HU. Arsenic and arsenic compounds. Ullmann's encyclopedia of industrial chemistry. Weinheim: Wiley-VCH; 2005.
22. Koek MM, Jellema RH, van der Greef J, Tas AC, Hankemeier T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics. 2011;7(3): 307–28.
23. Merrill Jr AH, Sullards MC, Allegood JC, Kelly S, Wang E. Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatogra- phy tandem mass spectrometry. Methods. 2005;36(2):207–24.

24. Tauler R. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. J Chemom. 2001;15(8):627–46.
25. de Juan A, Jaumot J, Tauler R. Multivariate curve resolution (MCR). Solving the mixture analysis problem. Anal Methods. 2014;6(14):4964–76.
26. Olson JM, Romano CA. A new chlorophyll from green bacteria. Biochim Biophys Acta. 1962;59(3):726–8.
27. Rocchetta I, Küpper H. Chromium- and copper-induced inhibition of photosynthesis in *Euglena gracilis* analysed on the single-cell level by fluorescence kinetic microscopy. New Phytol. 2009;182(2):405–20.
28. Domonkos I, Kis M, Gombos Z, Ughy B. Carotenoids, versatile components of oxygenic photosynthesis. Prog Lipid Res. 2013;52(4):539–61.

Capítulo 6 – Conclusões e perspectivas

1. NIR E ATR-FTIR..... 77

2. LC-MS..... 78

3. PERSPECTIVAS FUTURAS E RECOMENDAÇÕES..... 78

1. NIR e ATR-FTIR

- As infecção bacterianas possuem um representação significativa quanto se trata de causas de morbidade e mortalidade humana, com alta taxas de mortalidade relacionadas ao tempo de diagnóstico por métodos tradicionais e tratamentos tardios. Devido a isso a proposta de classificação multivariada combinada com a espectroscopia na região do infrvermelho (NIR e ATR-FTIR) pode ser justificada, pois a mesma possui precisão, velocidade e baixo custo efetivo quando comparados aos métodos convencionais de diagnóstico.
- Os modelos de classificação propostos foram aplicados nos espectros NIR e ATR-FTIR obtidos a partir de amostras de Staphylococcus aureus em cinco concentrações diferentes em sangue humano (Capítulo 2), DNA de Klebsiella pneumoneae produtora e não produtora de Carbapenemase (Capítulo 3), colônias de Pseudomonas aeruginosa sensível e resistente (Caoítulo 4). Com taxas de classificação e valores de sensibilidade significativos.
- O GA-LDA apresentou melhor desempenho em relação a sensibilidade, em todos os casos, com maior eficiência nos grupos alterados, possibilitando o diagnóstico correto destes e tratamento adequado. Em geral, os métodos de seleção de variáveis (GA e SPA, particularmente o primeiro) apresentaram melhores resultados de desempenho em comparação com a redução de variáveis pelo método (PCA). Foi notável a melhora no desempenho alcançado com o uso dos métodos de seleção de variáveis.

- As diferenças espectrais encontradas tiveram suas bandas atribuídas a (Capítulo 2) lipídeos, proteínas [Amida II (~1550 cm-1) e C=o de aminoácidos (~1400 cm-1)] e vibrações de DNA (~1080 cm-1), (Capítulo 3) assinaturas de DNA/RNA (-1756 nm), amidas secundárias (~1632 nm) e (Capítulo 4) C-H de lipídeos (1102 e 1108 nm).
- O desempenho na classificação das bactérias já mencionas nos Capítulos 2, 3 e 4 proporcionou uma perspectiva de detecção precoce da presença de bactérias e até mesmo de resistência bacteriana, implicando em um método mais rápido e barato de diagnóstico, podendo ser um método de suporte aos métodos convencionais.

2. LC-MS

- As cianobactérias, quando submetidas a fontes de contaminação por metais pesados, apresentam alterações em seu processo fotossintético. Avaliando que é atribuído as cianobactérias 50% da fotossíntese terreste, sendo uma das principais contribuintes para o ciclo de oxigênio global, pode ser justificado a proposta de desenvolver um método de detecção das mudanças que ocorrem na estrutura células dessas bactérias quando expostas a diferentes concentrações de Arsênio na forma de A2O3.
- Devido aos processos fotossintéticos ocorrerem na membrana tilacóide esta ser basicamente composta por lipídios, esse estudo avaliou o perfil lipidômico dessas bactérias por cromatografia líquida acoplada com espectrometria de massa.
- A combinação do MCR-ALS utilizando as regiões de interesse (ROI) em dados de LC-MS mostraram-se potenciais na determinação de mudanças no perfil lipidômicos da *Anabaena sp*. e *Planktothrix agardhii* quando expostas a Arsênio(III) durante 7 dias.
- As massas majoritárias encontradas estão relacionadas com galactolipideos (MGDG), clorofila e alguns carotenoides.

3. Perspectivas futuras e recomendações

- Avaliar outras espécies de bactérias e verificar a eficiencia das mesmas com os métodos espectroscópico NIR e ATR-FTIR em conjunto com métodos de seleção de variáveis SPA-LDA e GA-LDA ampliando o espectro amostral, possibilitando uma maior aplicabilidade em rotinas laboratoriais.
- Miniaturização de instrumentos para medições de campo que empregam LEDs que emitem radiação em comprimentos de onda previamente selecionados (por exemplo, UV, Vis, NIR e IR), para uso portátil desta técnica, podendo ser aplicadas ‘in locu’.
- Validar os método e padronizar a operacionalização, para posterior aplicação como método padrão de diagnóstico.

APÊNDICE A

Hyperspectral Raman spectroscopy image analysis of Cyanobacteria exposed to Arsenic (III) using multivariate curve resolution-alternating least squares (MCR-ALS)

Aline S. Marques

Carmen Bedia

Romà Tauler

Kássio M. G. Lima

Manuscrito em fase de correção

Contribuição:

- Eu desenvolvi o método para cultivo das cianobactérias utilizadas.
- Eu cultivei, repliquei e mantive a cultura de cianobactérias.
- Eu fiz a extração dos lipídeos a partir de um procedimento padronizado.
- Eu fiz o tratamento quimiométrico.
- Eu escrevi a primeira versão do manuscrito.

Aline de Sousa Marques Prof. Dr. Kássio Michell Gomes de Lima

Hyperspectral Raman spectroscopy image analysis of Cyanobacteria exposed to Arsenic (III) using multivariate curve resolution-alternating least squares (MCR-ALS)

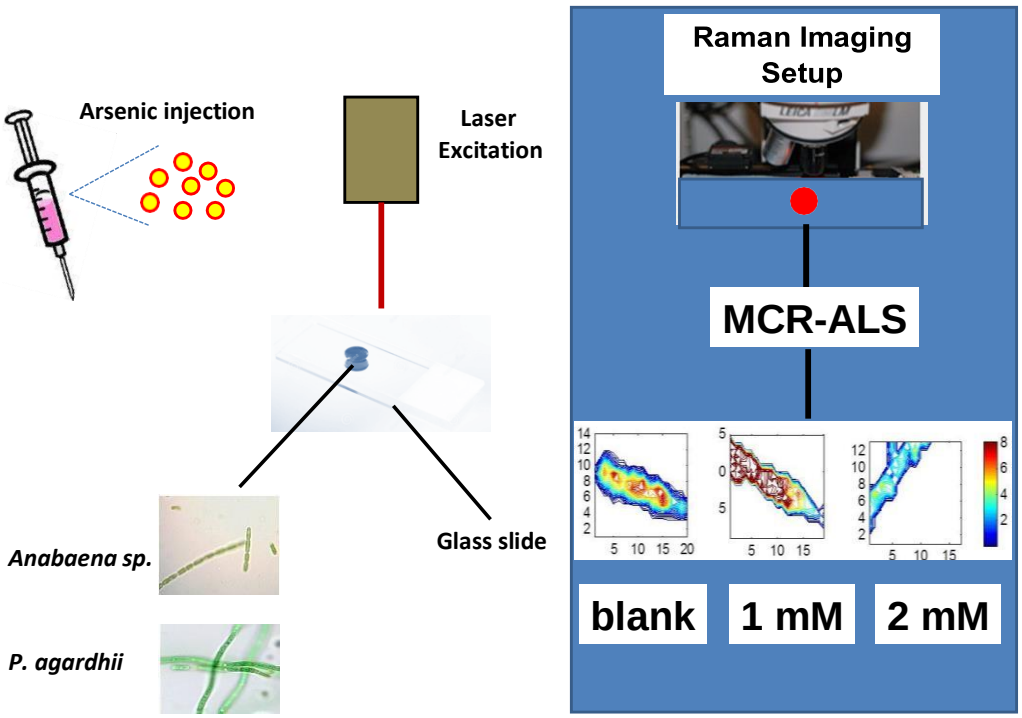
Aline S. Marques,[1] Carmen Bedia,[2] Roma Tauler,[2] Kássio M. G. Lima,[1*]

[1]UFRN–IQ, Biological Chemistry and Chemometrics, Natal 59071–970, Brazil

[2]IDAEA–CSIC, Jordi Girona 18–26, Barcelona 08028, Spain

*Author to whom correspondence should be addressed: e-mail kassiolima@gmail.com

ToC



Abstract: Cyanobacteria, the only known prokaryotes that perform oxygen-evolving photosynthesis, are receiving strong attention in applied research **.** This paper shows the application of hyperspectral Raman imaging combined with multivariate curve resolution-alternating least squares (MCR-ALS) to study changes on the concentrations of biological substances when two cyanobacteria strains (*Anabaena* sp*.* and *Planktothrix agardhii*) were exposed to different concentrations of Arsenic.. Our strategy allowed the estimation of the relative concentration and spectra profiles of the main components of the samples exposed to different concentrations of arsenic, as well as of blank unexposed samples, showing how they change after one-week time of treatment with the metal. Comparison of concentration distribution maps resolved by MCR–ALS in control and exposed samples, makes possible to observe the distributions of chlorophyll, sugars (glucose, galactose), and uronic acids in these cyanobacteria samples. As a conclusion of this work, hyperspectral Raman imaging coupled to the MCR-ALS method has been confirmed to be an appropriate tool to investigate how the presence of arsenic causes changes in the behavior of certain endogenous cyanobacteria metabolites after As(III) treatment over time.

Keywords: cyanobacteria; hyperspectral images; Raman; MCR-ALS; As(III).

Introduction

In the last years, hyperspectral imaging associated with vibrational spectroscopies (Near Infrared, Fluorescence and Raman) has gained increasing interest in different research and applied areas such as in food,[1] environmental,[2] research,[3] biomedical[4] and pharmaceutical[5,6] fields. This technique provides spatial information about the distribution of chemical components on the scanned area, generating data sets which can be arranged in three dimensions, two spatial dimensions which describes the mapped area and one spectral dimension associated with spectral variables.[7] Hyperspectral images result from multiwavelength spectroscopy readings of hundreds of contiguous spatial positions (pixels) of the target sample.[8]

Raman spectroscopy presents significant interest due its high specificity, low sensitivity to water and requirement for minimal sample preparation.[9] Raman is a scattering phenomenon as opposed to absorption/emission in fluorescence and its active molecules are more photostable compared with fluorophores, which are rapidly photo bleached for this narrow spectral features can be separated from the broad band autofluorescence[10]. Raman imaging is a powerful technique for the analysis of various kinds of biological targets such as cells, tissues, and living organisms[11]. It gives large amounts of spatially resolved biochemical (proteins, nucleic acids, lipids and CH_2 groups)[12] data. Raman spectroscopy can differentiate the spectral fingerprint of many molecules, resulting in very high multiplexing capabilities.

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) is a chemometric tool applied to resolve multiple component responses from

unknown unresolved mixtures as described in[13–16]. MCR-ALS has been already applied to extract relevant information from Raman images [17,18] for different type of biological samples.[5,19] This method has some features which can be adapted to different type of three-way data such as: i) it can be used with different model complexities (trilinear and not trilinear); ii) it has a simple algorithmic implementation based on matrix inversion; iii) eigenvalue–eigenvector decomposition of the experimental data matrix is used to determine the number of independent contributions; iv) it means that several constraints can be simply applied during the ALS optimization with increasing reliability of the solutions obtained.[20]

The aim of this work was to test the feasibility of the developed method using Raman images of two cyanobacteria (*Anabaena sp.* and *Planktothrix agardhii*) combined with MCR-ALS to evaluate possible changes on biological substances when these cyanobacteria are exposed to different concentrations (1 and 2 mM) of Arsenic. Cyanobacteria are the only known prokaryotes that perform oxygen-evolving photosynthesis, commonly found in diverse environments including marine and freshwater, as well as in terrestrial habitats.[21] When these cyanobacteria are exposed to a metal, such as arsenic, the photosynthesis is influenced negatively .[22,23] Arsenic (As) is considered to be one of the most toxic metal substances for the environments.[24] Several studies have shown the responses of different microorganisms exposed to this metal, such as the demethylation process in fungi,[25] the contamination of food and its bioaccumulation[26] and transformation and release in Microcystis.[27]

This work is divided in three sections: first, experimental details on how the cyanobacteria cells were exposed to As (III) and on their analysis by Raman

spectroscopic imaging are described. Second, the application of MCR-ALS for the simultaneous analysis of these Raman images on two different species of cyanobacteria at different concentrations of Arsenic is given. Finally, from the results obtained, preliminary identification of the cyanobacteria chemical constituents that changed with As(III) exposure and the possible biochemical implications are discussed.

Materials and methods

Cyanobacterial culture and arsenic exposure

Anabaena sp. and *Planktothrix agardhii* from Culture Collection of Algae and Protozoa (CCAP1403/21 and CCAP1459/11 respectively, SAMS Limited) were cultivated aseptically in 500 mL Erlenmeyer flasks containing Blue-Green medium (http://www.ccap.ac.uk/media/documents/BG11.pdf) at 25 ºC under continuous fluorescent white light with an intensity of 15 µmol m-2 s-1, until an optical density of 0.1-0.2 at 750 nm (OD750) was reached. Then, the cultures were transferred to 50 mL Falcon flasks with different concentrations (0, 1 and 2 mM) of As(III) (As_2O_3, Sigma) and cells were maintained in the mentioned conditions for 7 days.

Raman imaging

The hyperspectral Raman image were obtained with a DXRXi Raman microscope (Thermo Scientific Inc., USA) and the OMNICxi software (Thermo Scientific Inc., USA), coupled to a 455 nm wavelength laser. The 0.90/100x objective was used. The maps were recorded with spatial resolution of 1µm in both directions. The integration time was equal 0.16s and fixed for each scan. One drop of each sample (*Anabena sp.* and *Planktothrix agardhii* with 0, 1 and

2 mM of As(III))was dispensed on a silica-glass slide and leave to dry and a single spectrum in each point was recorded in the range 3400-150 cm^{-1} with 8 scans. The laser power was 1.5 mW. Six Raman maps were captured with 14x20 points for Ana00 (control), 15x19 points for Ana01 (1mM), 15x14 points for Ana02 (2mM), 24x14 points for Pltx00 (control), 16x22 points for Pltx01 (1mM) and 23x15 points Pltx02 (2mM).

Data analysis and pre-processing

In this work, the Raman image data files in HDR format were uploaded into the MATLAB (version R2012b, The Mathworks Inc,) using the functions read_envihdr and envidataread2 of the Bioinformatics Toolbox (The Mathworks Inc.). Asymmetric least squares (AsLS) and Savitzky–Golay (SG) smoothing methods were applied to eliminate undesired light scattering, instrumental, and background effects,[28] as shown in Figure 1 for an image data (*Anabaena sp.* without As(III) exposure) example.

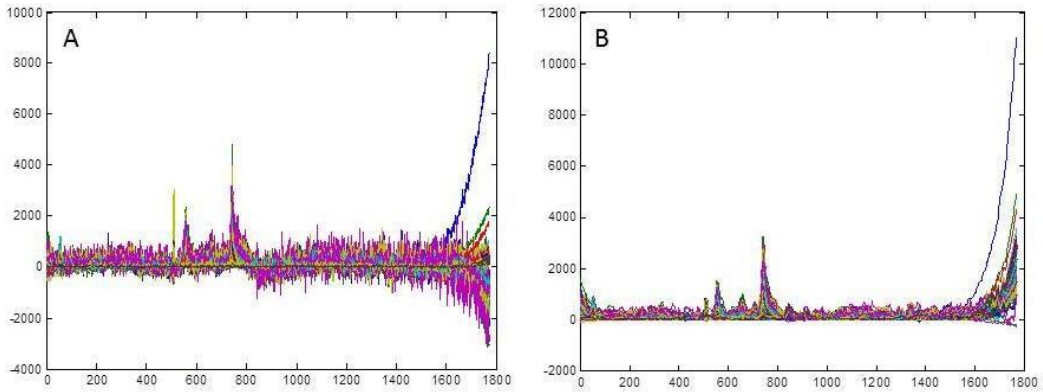


Fig. 1. A. Raw data spectra of *Anabaena sp.* without As(III) exposure, B. Anabaena sp. spectra after AsLs + SG pretreatment, C. Raw data spectra of

Planktothrix agardhii without As(III) exposure, D. Planktothrix agardhii spectra after AsLs + SG pretreatment.

Asymmetric least squares (AsLs) is a pretreatment method proposed by Eilers[29] that remove the possible baseline and background contributions. In the Raman case, AsLs reduced well the fluorescence contribution in the measured spectra. Savitzky–Golay[30] is a smoothing technique used in this paper to remove the noise contributions of Raman signals with a window of five points width and first polynomial degree.

Chemometric data analysis

The Multivariate curve resolution–alternating least squares method has been already used for the analysis of the hyperspectral images[31–33] due to the easy of the introduction of the external spectral and spatial information about the image during its resolution and to the ability to work with single and multiset image arrangements. MCR-ALS is used to decompose the hyperspectral images into their signatures or pure spectra of the image constituents and into their concentration (relative amounts) on the image (distribution map).[31] A brief description of MCR-ALS in given here, for more information about it see the references.[14,20,34] This method suppose that the mixture spectra in every pixel can be described as a linear combination of the spectra of the constituents weighted by their relative amounts. MCR-ALS performs a bilinear decomposition of a data matrix that can be described as:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \qquad \text{Equation 1}$$

where **D** is the raw data matrix (original raw pixel mixture spectra). To apply the MCR-ALS, the hyperspectral data cube is unfolded into a **D** matrix, as shown in

Figure 2. **C** is the matrix of the relative concentrations of the chemical constituents present in the pixels of the scanned image (from which their spatial distribution maps can be derived), $\mathbf{S}^{\mathbf{T}}$ is the matrix of their pure spectra (spectral signatures of the chemical constituents) and **E** has the experimental error in the raw measurement.s

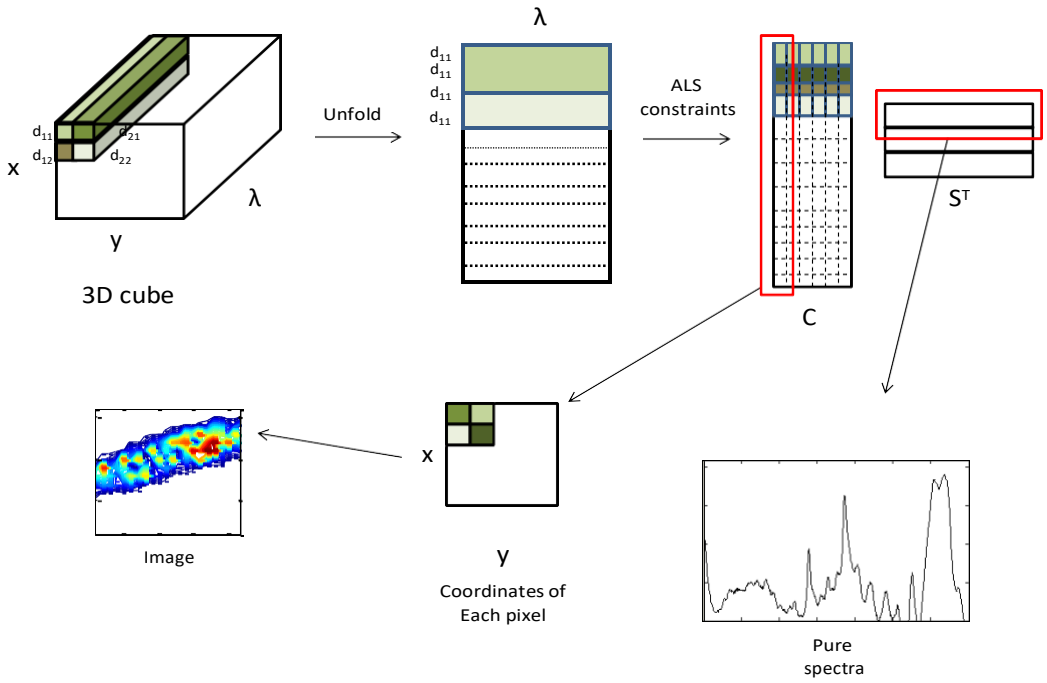


Fig. 2. Schematic representation of the bilinear matrix decomposition of a hyperspectral image using the MCR-ALS method.

A flowchart of the main steps of the MCR-ALS algorithm [17] when applied to hyperspectral images is illustrated in t Figure **3.**

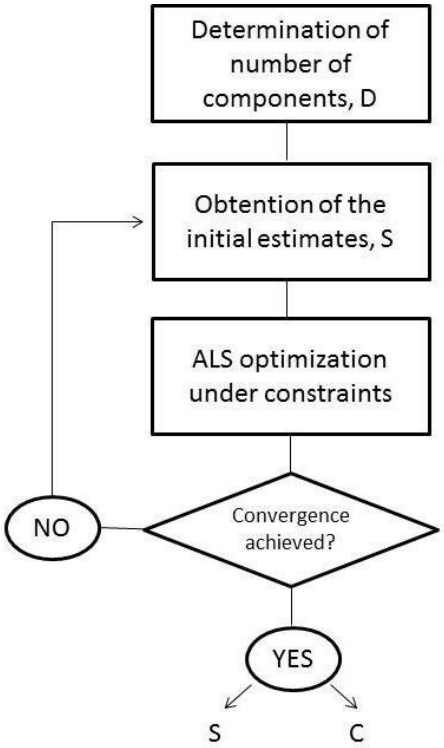


Fig. 3. Flowchart for general steps of MCR-ALS algorithm

MCR-ALS can be easily extended to the analysis of multiple data matrices. When multiple images are simultaneously analyzed, every sample image are folded in a data matrix and arranged in a new column-wise augmented data matrix, with the spectral, λ, direction in the common columns dimension. See previous works where this approach has been explained in detail [38,39]. In this work, the simultaneous MCR-ALS analysis of the hyperspectral images coming from the different samples of cyanobacteria culture (three of *Anabaena sp.* and three of *P. agardhii*) under different concentrations of As(III) (0, 1 and 2 mM of As(III) from As_2O_3) was performed using the column-wise augmented matrix, previously described, and the bilinear model that can be written in matrix form as:

$$\mathbf{D_{aug}} = [\mathbf{D_1};\mathbf{D_2};\mathbf{D_3}] = [\mathbf{C_1};\mathbf{C_2};\mathbf{C_3}]\,\mathbf{S^T} + [\mathbf{E_1};\mathbf{E_2};\mathbf{E_3}] = \mathbf{C_{aug}S^T} + \mathbf{E_{aug}} \quad \textbf{Equation 2}$$

where $\mathbf{D_{aug}}$ is the column-wise augmented matrix that has the data images from the three cyanobacteria samples at the different As(II) concentrations. According to equation 2, $\mathbf{D_{aug}}$ augmented data matrix, is decomposed by the extension of the same bilinear model as in Equation 1, but now for the simultaneous analysis of multiple data matrices (hyperspectral images). $\mathbf{C_{aug}}$ in Equation 2**,** is the column-wise augmented matrix with the relative concentrations of the constituents (distribution maps) at every pixel coordinates of every individual hyperspectral image corresponding to a individual cyanobacteria sample. $\mathbf{S^T}$ in Equation 2 has the pure spectra of these different constituents resolved by MCR-ALS. Note that these spectra will be the same (common) for all the simultaneously analyzed hyperspectral images (cyanobacteria samples in $\mathbf{D_1}$, $\mathbf{D_2}$ and $\mathbf{D_3}$ data matrices).

In the MCR-ALS procedure, Equations 1 and 2 are solved by an alternating least squares (ALS) optimization under a set of constraints. In this case, non-negativity constraints were applied to both, to the image concentrations, **C** or $\mathbf{C_{aug}}$, and to the pure spectra, $\mathbf{S^T}$ **(**Equations 1 and 2). Furthermore, to avoid scale ambiguities during ALS, spectra normalization was also applied**.**

After the application of MCR-ALS, the identification of every one of these three other components resolved by MCR-ALS was performed by comparison of the Raman bands with those previously reported in the literature.[35, 36] Finally to evaluate the quality of the fitting achieved with MCR-ALS, the percentage of

lack-of-fit (lof, give a measure of fit quality in relative terms with the same units as the measured data and comparable with experimental relative error estimations) (Eq. 2) and the percentage of explained variance (R^2) (Eq. 3) are calculated with the equations above:

$$\mathbf{lof}(\%) = \mathbf{100} \times \sqrt{\frac{\sum_{ij} \mathbf{d_{ij} - d'_{ij}}}{\sum_{ij} \mathbf{d_{ij}^2}}}$$

(2)

$$\mathbf{R^2} = \mathbf{100} \times (\mathbf{1} - \frac{\sum_{ij} \mathbf{e^2}}{\sum_{ij} \mathbf{d_{ij}^2}})$$

(3)

where dij is the element of the hyperspectral image data matrix **D**, and d´ is the corresponding element of this data matrix recalculated by the ALS model. Here, eij are the elements of the **E** matrix, and dij are the elements of the raw data set **D**.

Results and discussion

Images obtained with Nikon eclipse 90i microscope using a 60x oil objective equipped with Nikon DS-Ri1 digital camera of cyanobacteria (*Anabaena sp.* and *Planktothrix agardhii*) samples showing morphological changes when As(III) was added at different concentrations (0, 1 and 2 mM) are given in Figure 4.

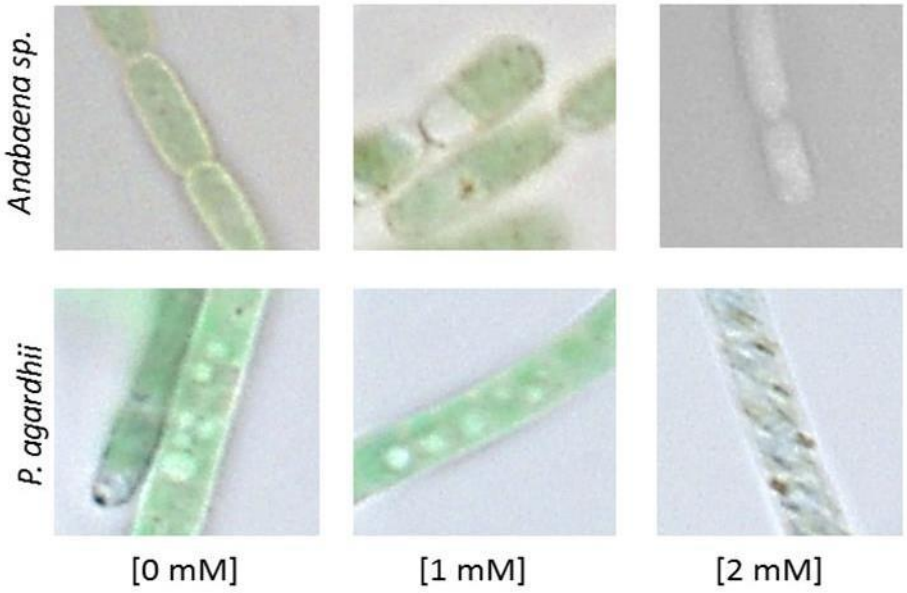


Fig. 4. Microscope images of *Anabaena sp.* and P*lanktothrix agardhii* in presence of 0, 1 e 2 mM of arsenic.

As shown in Figure 4, the increase in the concentration of arsenic in the culture medium, caused a fading of green pigmentation, being more intensively observed in *Anabaena sp.* cells. This coloration can be associated to changes in the composition of photosynthetic pigments, such as chlorophyll (see reference 36).

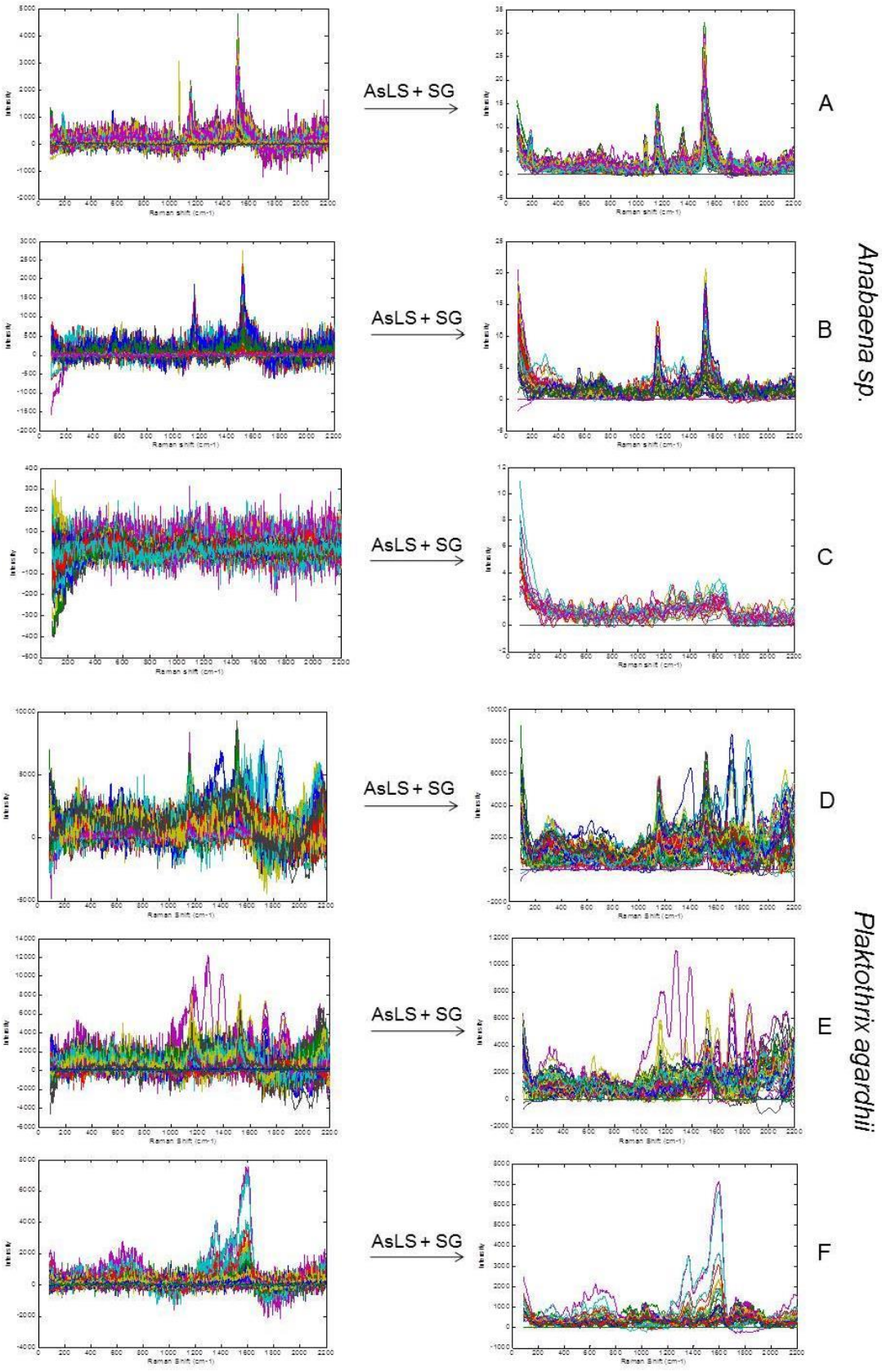

AsLS + SG
A
AsLS + SG
B
AsLS + SG
C
AsLS + SG
D
AsLS + SG
E
AsLS + SG
F
Anabaena sp.
Plaktothrix agardhii

Fig. 5. Raman spectra pure and pre-treated with AsLs and Savitzky-Golay after 7 days of exposure to different arsenic concentrations of A. *Anabaena sp.* [0 mM], B. *Anabaena sp.* [1 mM], C. *Anabaena sp.* [2 mM], D. *Plaktothrix agardhii* [0 mM], E. *Plaktothrix agardhii* [1 mM] and F. *Plaktothrix agardhii* [2 mM].

Figure 5 shows the raw Raman spectra of the different analyzed *Anabaena sp.* (A, B and C) and *Plaktothrix agardhii* (D, E and F) samples (images), and the same spectra once they have been pretreated with Asymmetric Least Squares (AsLs) and Savitzky-Golay (SG). Unwanted contributions like light scattering, instrumental noise and baseline effects were significantly reduced using the combination of these two pretreatment methods. Fluorescence major contributions were removed satisfactorily cutting their contributions in the spectral range 2400-2200 cm^{-1}.

Table 1 and 2 summarize the MCR-ALS results obtained for raw and pretreated (AsLs + SG) *Anabaena sp* data. Good data fitting results were obtained with pretreated data, with a significant improvement when compared to results with raw data. The same results can be seen in Table 2 to *Planktothrix agardhii* data. Best MCR-ALS results are given in italic in both Tables.

Table 1: MCR-ALS results of Raman hyperspectral imaging *Anabaena sp.* data.

Constraints	**Anabaena_0MM**		**Anabaena_01MM**		**Anabaena_02MM**	
	Lack of fit (%)	R^2(%)	Lack of fit (%)	R^2(%)	Lack of fit (%)	R^2(%)
Raw data						
Non-negativity	12.6	78.7	21.7	65.6	41.2	49.2

Pretreated data (AsLs + SG)						
Non-negativity	2.4	95.8	2.40	95.8	1.2	95.6

Table 2. MCR-ALS results of Raman hyperspectral imaging *Plaktothrix agardhii* data

Constraints	**Planktothrix_0M**		**Planktothrix _01MM**		**Planktothrix _02MM**	
	Lack of fit (%)	R^2(%)	Lack of fit (%)	R^2(%)	Lack of fit (%)	R^2(%)
Raw data						
Non-negativity	22.5	76.5	24.8	70.7	22.5	76.3
Pretreated data (AsLs + SG)						
Non-negativity	6.7	92.8	10.2	91.1	4.9	97.1

The two cyanobacteria augmented data matrices, , one with the 3 samples *Anabaena sp.* and another one with the 3 samples *Planktothrix agardhii* at the different As(III) concentrations were then built up (and analyzed by MCR-ALS (see method section) using non-negativity constraints to concentration and spectra profiles (as mentioned in Materials and methods section). MCR-ALS was applied to these two augmented data matrices using 3 components. When more components were used no additional information was

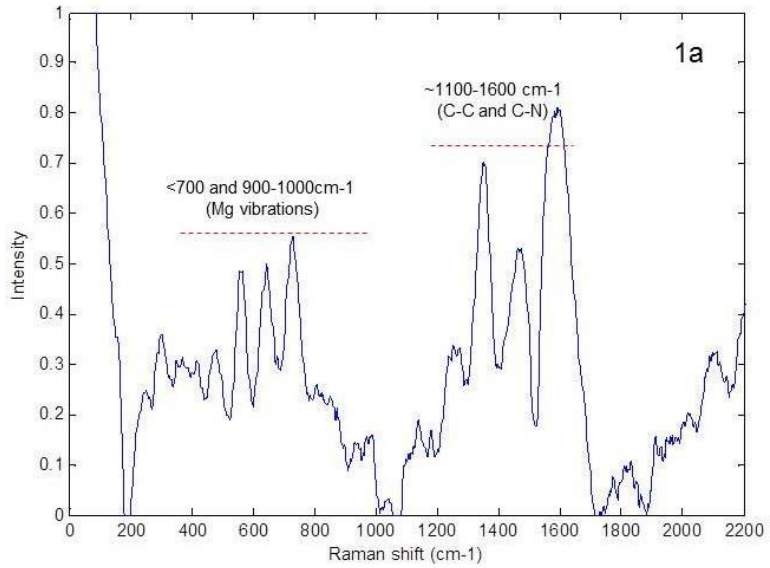
obtained related to the changes in the composition of the analyzed cyanobacteria samples Table 3 shows final MCR-ALS data fitting results for the two augmented matrices of *Anabaena sp*. and *Plaktothrix agardhii* with three components. In this case only the MCR-ALS results for the pretreated data are shown.

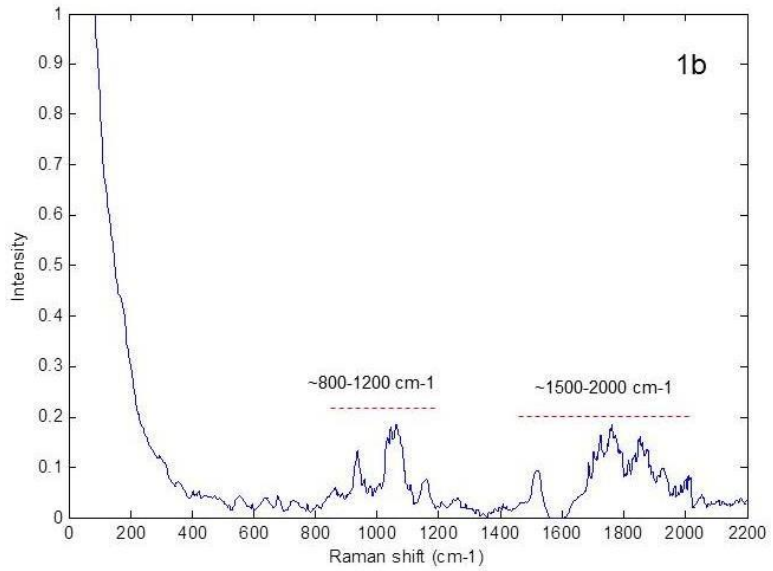
Table 3. MCR-ALS results of Raman hyperspectral imaging *Anabaena sp.* and *Plaktothrix agardhii* augmented data matrices,

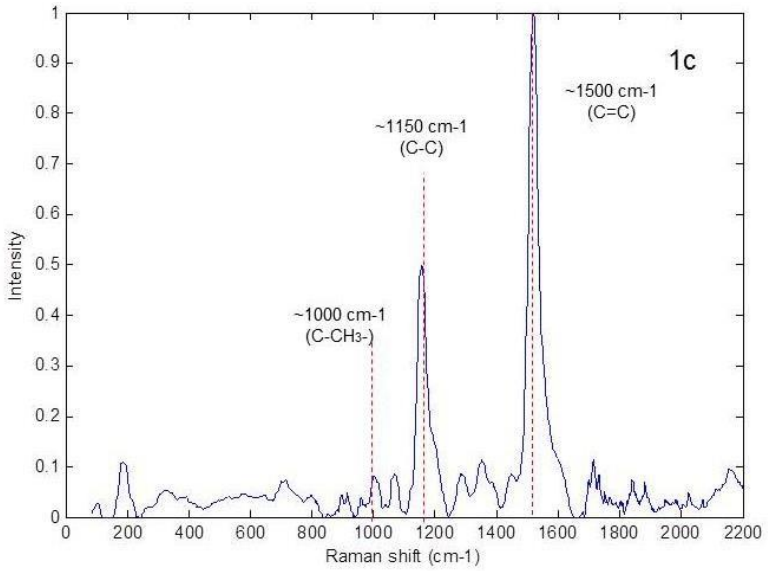
Constraints	**Anabaena_aug**		**Planktothrix _aug**	
Pretreated data (AsLs + SG)	Lack of fit (%)	R^2(%)	Lack of fit (%)	R^2(%)
Non-negativity	2.61	94.68	7.47	90.41

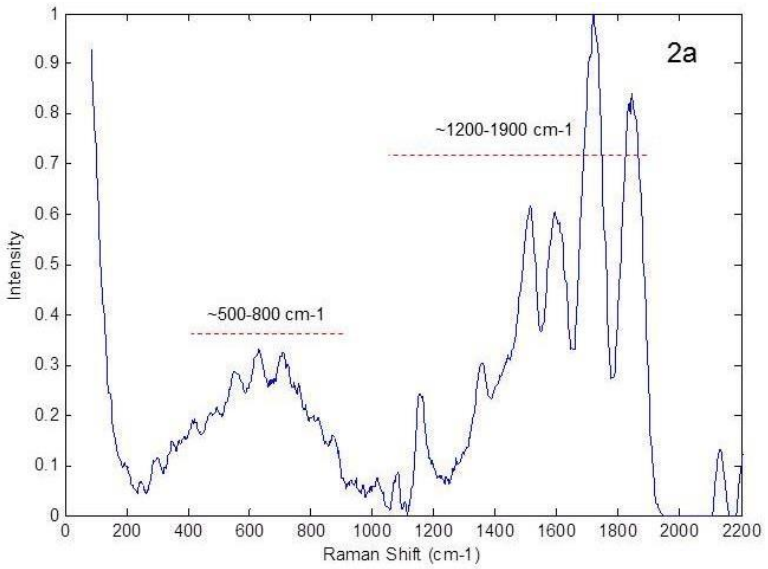
Figure 6. Shows the Raman spectral signatures resolved by MCR-ALS for *Anabaena sp* (components A, B and C) and *P. agardhii samples* (components D, E and F) when they were exposed to As(III) . *Anabaena sp* component A, showed intense Raman bands below 700 cm^{-1} and at 900-1000 cm^{-1} (magnesium vibrations) and ~ 1100-1600 cm^{-1} (CC and CN bonds).) Raman spectral signatures of component C (Figure 7.1c *Anabaena sp.*) and of component F (Figure 7.2c , *P. agardhii)* show bands at ~ 1500 cm^{-1} (C = C), ~ 1150 cm^{-1} (CC) and ~ 1000 cm^{-1} (C-CH 3)). Raman spectral signatures of components B (Figure 7.1b), D (Figure 7.2) and E (Figure 7.2b) have bands at 1200- 1900 cm^{-1} and 500-800 cm^{-1}, 800-1200 cm^{-1} and 300 cm^{-1}, 1150 cm^{-1} (CC) , 1361 cm^{-1} and 1500 cm^{-1} (C = C) respectively. When compared to library

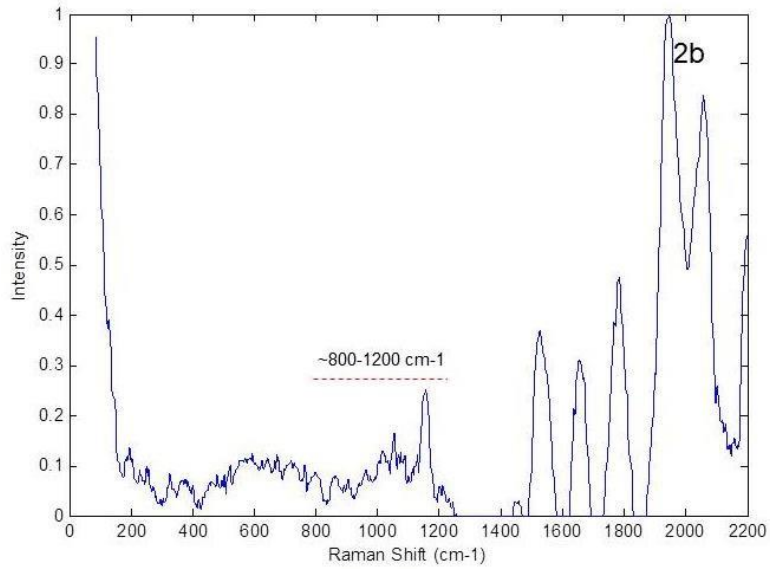
spectrum this resulted appear similar to the Raman spectra of Algae with carotenoids and chlorophyll composition as defined in a previous work.[36, 37]






1c
~1500 cm-1
(C=C)
~1150 cm-1
(C-C)
~1000 cm-1
(C-CH3-)
Intensity
Raman shift (cm-1)


2a
~1200-1900 cm-1
~500-800 cm-1
Intensity
Raman Shift (cm-1)



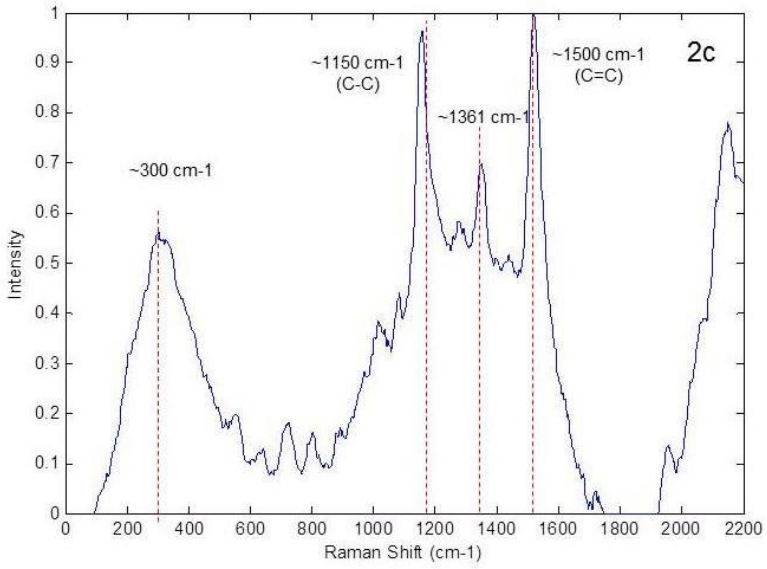
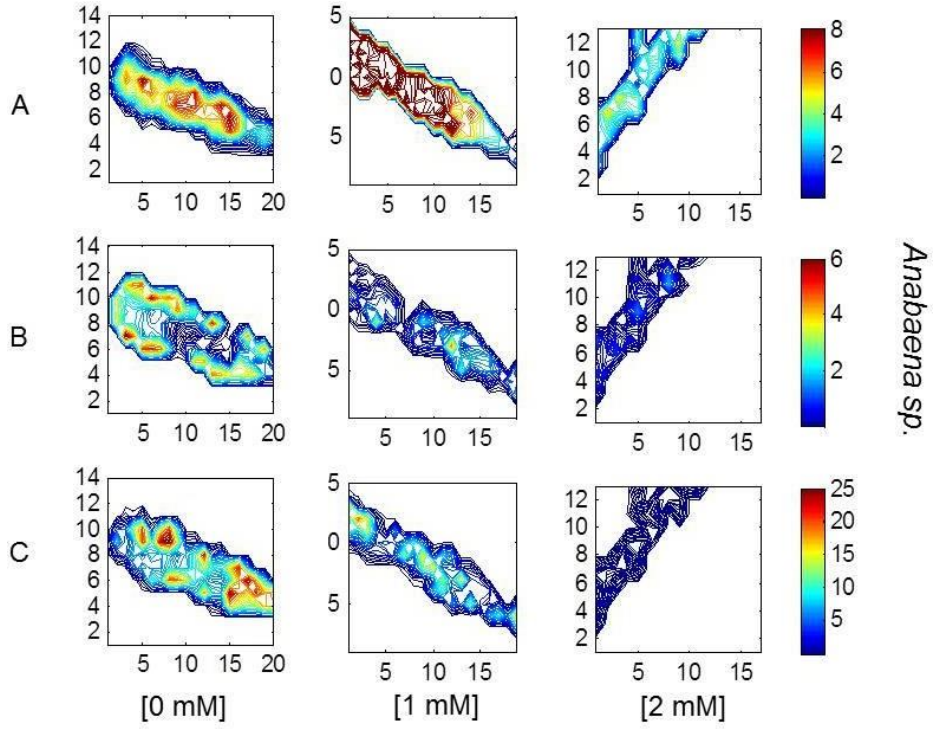


Fig. 6. Resolved MCR-ALS spectra for each component. 6a). *Anabaena sp*.- component A, 6b). *Anabaena sp.*- component B, 6c). *Anabaena sp.*- component C, 6d). *P. agardhii*- component D, 6e). *P. agardhii*- component E and 6f) *P. agardhii*- component F.

Figure 7 gives the spatial distribution (maps) of the three different components resolved by MCR-ALS for *Anabaena sp*. and *Planktothrix agardhii* samples at the different investigated concentrations of As(III) (A, B, C for *Anabaena sp*., and D, E and F for *Planktothrix agardhii*). A decrease in the

presence of the first component (A) can be observed. This component decreased drastically when twice the concentration of As(III) was added. In the maps B and C As(III) causes also changes in the contributions of MCR-ALS components 2 and 3 . In the same way MCR-ALS components related to Planktothrix agardhii (D, E and F) show the changes produced by As(III), with increasing and reduction on the contribution of these components and on the concentration of the chemical compounds associated to them.



!

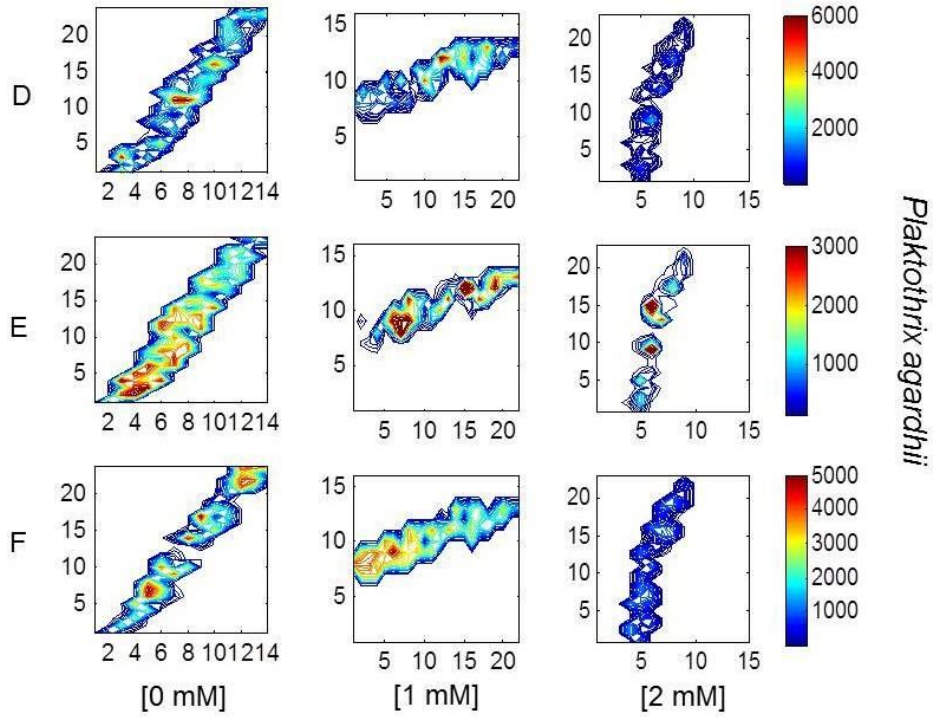


Fig. 7. Raman image of A. *Anabaena sp.* without As(III), B. *Anabaena sp.* with 1 mM of As(III), C. *Anabaena sp.* with 2 mM of As(III), D. *P. agardhii* without As(III), E. *P. agardhii* with 1 mM of As(III) and F. *P. agardhii* with 2 mM of As(III). The letters are the components, and the columns the concentrations!

Conclusions

In this paper, Raman hyperspectral images of two cyanobacteria species (*Anabaena sp.* and *Planktothrix agardhii* at different concentrations of arsenic (As(III)) were analyzed with the Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) method.). As results of this analysis showed that the presence of arsenic (As(III)) during the growth of these two cyanobacteria (*Anabaena sp.* and *Planktothrix agardhii*) affected strongly the molecular

structures associated to the photosynthetic processes such as chlorophyll a and b and carotenoids.

Acknowledgements

The authors would like to acknowledge the financial support from Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES) and *Brazilian* National Council for Scientific and Technological Development (CNPq) for a one year fellowship to Aline Marques in the chemometrics research group at IDAEA-CSIC, Barcelona, Spain. K.M.G. Lima acknowledges the CNPq Grant (305962/2014–0) for financial support. This work was funded by grants from CNPq/Capes project (Grant 070/2012) and by the CHEMAGEB project (FP/2007-2013)/ERC Grant Agreement n. 320737). The authors also thank Dr. Benjamín Piña and Claudia Rivetti (IDEAE/CSIC) for providing the cyanobacteria strains control.

References

1 A. López-Maestresalas, J. C. Keresztes, M. Goodarzi, S. Arazuri, C. Jarén and W. Saeys, *Food Control*, 2016, **70**, 229–241.

2 O. Sytar, M. Brestic, M. Zivcak, K. Olsovska, M. Kovar, H. Shao and X. He, *Sci. Total Environ.*, 2016.

3 J. Cheng, A. Volkmer, L. D. Book and X. S. Xie, *J. Phys. Chem. B*, 2002, **106**, 8493–8498.

4 V. Zheludev, I. Polonen, N. Neittaanmaki-Perttu, A. Averbuch, P. Neittaanmaki, M. Gronroos and H. Saari, *Biomed. Signal Process. Control*, 2015, **16**, 48–60.

5 G. P. S. Smith, C. M. McGoverin, S. J. Fraser and K. C. Gordon, *Adv. Drug Deliv. Rev.*, 2015, **89**, 21–41.

6 M. Boiret, A. de Juan, N. Gorretta, Y. M. Ginot and J. M. Roger, *J. Pharm.*

Biomed. Anal., 2015, **103**, 35–43.

7 C. J. G. Colares, T. C. M. Pastore, V. T. R. Coradin, L. F. Marques, A. C. O. Moreira, G. L. Alexandrino, R. J. Poppi and J. W. B. Braga, *Microchem. J.*, 2016, **124**, 356–363.

8 X. Zhang, A. de Juan and R. Tauler, *Appl. Spectrosc.*, 2015, **69**, 993–1003.

9 S. Stewart, R. J. Priore, M. P. Nelson and P. J. Treado, *Annu. Rev. Anal. Chem.*, 2012, **5**, 337–360.

10 P. Crow, B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad and N. Stone, *Br. J. Cancer*, 2005, **92**, 2166–2170.

11 H. Kano, H. Segawa, M. Okuno, P. Leproux and V. Couderc, *J. Raman Spectrosc.*, 2016, **47**, 116–123.

12 L. Ashton, K. A. Hollywood and R. Goodacre, *Analyst*, 2015, **140**, 1852–1858.

13 J. Jaumot, R. Gargallo, A. de Juan and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2005, **76**, 101–110.

14 J. Jaumot, A. de Juan and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2015, **140**, 1–12.

15 E. Gorrochategui, J. Casas, C. Porte, S. Lacorte and R. Tauler, *Anal. Chim. Acta*, 2015, **854**, 20–33.

16 C. Bedia, N. Dalmau, J. Jaumot and R. Tauler, *Environ. Res.*, 2015, **140**, 18–31.

17 F. B. Lavoie, N. Braidy and R. Gosselin, *Chemom. Intell. Lab. Syst.*, 2016, **153**, 40–50.

18 M. B. Mamián-López and R. J. Poppi, *Microchem. J.*, 2015, **123**, 243–251.

19 M. Ando and H. Hamaguchi, *J. Biomed. Opt.*, 2013, **19**, 11016.

20 K. M. G. Lima, C. Bedia and R. Tauler, *Microchem. J.*, 2014, **117**, 255–261.

21 V. E. de Oliveira, M. A. C. Neves Miranda, M. C. S. Soares, H. G. M. Edwards and L. F. C. de Oliveira, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, 2015, **150**, 373–80.

22 X.-X. Yin, J. Chen, J. Qin, G.-X. Sun, B. P. Rosen and Y.-G. Zhu, *Plant Physiol.*, 2011, **156**, 1631–1638.

23 C. Cassier-Chauvat and F. Chauvat, *Int. J. Mol. Sci.*, 2015, **16**, 871–886.

24 T. Joseph, B. Dubey and E. A. McBean, *Sci. Total Environ.*, 2015, **527**–**528**, 540–551.

25 S. Su, X. Zeng, Q. Feng, L. Bai, L. Zhang, S. Jiang, A. Li, R. Duan, X. Wang, C. Wu and Y. Wang, *Environ. Pollut.*, 2015, **204**, 141–144.

26 K. Phan, S. Sthiannopkao, S. Heng, S. Phan, L. Huoy, M. H. Wong and K. W. Kim, *J. Hazard. Mater.*, 2013, **262**, 1064–1071.

27 Z. Wang, Z. Luo and C. Yan, *Environ. Sci. Pollut. Res.*, 2013, **20**, 7286–7295.

28 J. Peng, S. Peng, A. Jiang, J. Wei, C. Li and J. Tan, *Anal. Chim. Acta*, 2010, **683**, 63–68.

29 P. H. C. Eilers, *Anal. Chem.*, 2003, **75**, 3631–3636.

30 J. Luo, K. Ying and J. Bai, *Signal Processing*, 2005, **85**, 1429–1434.

31 X. Zhang and R. Tauler, *Anal. Chim. Acta*, 2013, **762**, 25–38.

32 G. L. Alexandrino, M. R. Khorasani, J. M. Amigo, J. Rantanen and R. J. Poppi, *Eur. J. Pharm. Biopharm.*, 2015, **93**, 224–230.

33 L. de M. França, M. F. Pimentel, S. da S. Simões, S. G. Junior, J. M. Prats-Montalbán and A. Ferrer, *Eur. J. Pharm. Biopharm.*, 2016, **104**, 180–188.

34 J. Jaumot and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2010, **103**, 96–107.

35 C. Matthaus, B. Bird, M. Miljkovic, T. Chernenko, M. Romeo, M. Diem, *Methods in Cell Biology*, 2008, **89**, 275-308.

36 Y. Y. Huang, C. M. Beal, W. W Cai, R. S. Ruoff, E. M. Terentjev, *Biotech. and Bioeng*., 2010, **105**, 889-897.

37 H. Abramczyk, B. Brozek-Pluska, *Anal. Chim. Acta*, 2016, **909**, 91-100.

38 A. De Juan, F. S. L. Borba, T. Jawhari, R. S. Honorato, *The Analyst*, 2017, **142**, 1177-1178.

39 X. Zhang, A. De Juan, R. Tauler, *Talanta*, 2016, **146**, 1-9.

Captions for Figure

Fig. 1. (A) Raw data spectra of *Anabaena sp.* without As(III) exposure; (B) Anabaena sp. spectra after AsLs + SG pretreatment; (C) Raw data spectra of *Planktothrix agardhii* without As(III) exposure; (D) Planktothrix agardhii spectra after AsLs + SG pretreatment.

Fig. 2. Schematic representation of the bilinear matrix decomposition of a hyperspectral image using the MCR-ALS method.

Fig. 3. Flowchart for general steps of MCR-ALS

Fig. 4. Images of *Anabaena sp.* and P*lanktothrix agardhii* in presence of 0, 1 e 2 mM of arsenic.

Fig. 5. Raman spectra pure and pre-treated with AsLs and Savitzky-Golay after 7 days of exposure to different arsenic concentrations of A. *Anabaena sp.* [0 mM], B. *Anabaena sp.* [1 mM], C. *Anabaena sp.* [2 mM], D. *Plaktothrix agardhii* [0 mM], E. *Plaktothrix agardhii* [1 mM] and F. *Plaktothrix agardhii* [2 mM].

Fig. 6. Raman image of A. *Anabaena sp.* without As(III), B. *Anabaena sp.* with 1 mM of As(III), C. *Anabaena sp.* with 2 mM of As(III), D. *P. agardhii* without As(III), E. *P. agardhii* with 1 mM of As(III) and F. *P. agardhii* with 2 mM of As(III).

Fig. 7. Spectra recovered to each component 1a. *Anabaena sp*.- component A, 1b. *Anabaena sp.*- component B, 1c. *Anabaena sp.*- component C, 2a. *P. agardhii*- component D, 2b. *P. agardhii*- component E and 2c. *P. agardhii*- component F.