



---

Universidade Estadual de Campinas – UNICAMP

Instituto de Química – IQ

Departamento de Físico-Química

TESE DE DOUTORADO

Aplicação de Métodos de Resolução de Curvas Multivariada a Dados  
Experimentais Gerados por Diferentes Técnicas Analíticas: Uma Visão  
Exploratória da Biossíntese da Violaceína.

Raimundo Clecio Dantas Muniz Filho

Orientadora: Profa. Dra. Márcia Miguel Castro Ferreira

Campinas, julho 2011

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE  
QUÍMICA DA UNICAMP**

M925a	<p>Muniz Filho, Raimundo Clecio Dantas. Aplicação de métodos de resolução de curvas multivariada a dados experimentais gerados por diferentes técnicas analíticas: uma visão exploratória da biossíntese da violaceína / Raimundo Clecio Dantas Muniz Filho. -- Campinas, SP: [s.n], 2011.</p> <p style="text-align: center;">Orientador: Profa. Dra. Márcia Miguel Castro Ferreira.</p> <p style="text-align: center;">Doutorado - Universidade Estadual de Campinas, Instituto de Química.</p> <p style="text-align: center;">1. MCR-ALS. 2. PARAFAC. 3. Violaceína. I. Ferreira, Márcia Miguel Castro. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.</p>
-------	---

**Título em inglês:** Application of multivariate curve resolution to experimental data generated by several analytical techniques: an exploratory overview of violacein biosynthesis

**Palavras-chaves em inglês:** MCR-ALS, PARAFAC, Violacein

**Área de concentração:** Físico-Química

**Titulação:** Doutor em Ciências

**Banca examinadora:** Profa. Dra. Márcia Miguel Castro Ferreira (orientadora), Dr. André Messias Krell Pedro (UNILEVER), Prof. Dr. Francisco Benedito Teixeira Pessine (IQ-UNICAMP), Prof. Dr. Fábio César Gozzo (IQ-UNICAMP), Prof. Dr. Marcelo Martins Sena (DQ-UFMG)

**Data de defesa:** 11/07/2011



Dedico esta Tese aos meus pais, irmãos e familiares, em especial, à vó Creusa (*in memoriam*). À minha querida esposa, Márcia Patrícia, e a nosso pequenino(a) que está por vir, por todo amor, dedicação e compreensão em todos os momentos dessa dura jornada.



## Agradecimentos

Lembro-me como se fosse hoje, era um 25 de fevereiro de 2007, quando parti de Teresina, capital do Estado do Piauí, rumo a desconhecida, para mim, Campinas. A possibilidade de fazer doutorado em uma instituição de tamanho renome, como a UNICAMP, era minha força propulsora para vencer a barreira dos quatro longos anos longe das pessoas que sempre vivemos e amamos. Hoje, percebo que todo esse tempo já ficou para a história, e que sem a ajuda de diversas pessoas, o mesmo teria sido muito mais pesado e lento. Por isso, gostaria de agradecer a todos que de alguma forma contribuíram para que essa jornada chegasse ao fim, em especial:

- À Deus, por permitir que mais um objetivo que almejei durante minha vida fosse alcançado.
- À Profa. Márcia, pela orientação, ensinamentos, confiança e amizade.
- Aos Professores Volpe e Duran por permitirem que eu utilizasse seus laboratórios para realização dos experimentos.
- Aos Mestres que contribuíram com a minha formação (Pedro Vazquez, Yuji Takahata, Márcia, Cheti, Bruns, Munir, Marcelo Ganzarolli e Pessine).
- Ao Prof. Romà Tauler e a seu grupo pelos ensinamentos e hospitalidade.
- Aos Técnicos (Cláudia Martelli, José, Rita de Cássia, Divino, Chico, Míriam e Ricardo) e ao pessoal da CPG (Bel, Miguel e Gabriela).
- Ao Dorival Martins, por todos os ensinamentos de como trabalhar com a *Chromobacterium violaceum*.
- Aos Amigos do LQTA, em especial o Samuel, pela amizade e companheirismo até nas madrugadas de experimento, e a todos os outros Amigos.
- Ao CNPq pelo suporte financeiro tanto no doutorado pleno quanto no doutorado sanduíche



# Curriculum Vitae

Raimundo Clecio Dantas Muniz Filho

---

## 1.0 Formação Acadêmica/Titulação

- 2007-2011
- Doutorado em Química.  
Universidade Estadual de Campinas, UNICAMP, Campinas, SP.  
Título: Aplicação de métodos de resolução de curvas multivariada a dados experimentais gerados por diferentes técnicas analíticas: uma visão exploratória da biossíntese da violaceína.  
Orientador: Profa. Dra. Márcia Miguel Castro Ferreira.  
Bolsista: Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq.
- 2005-2007
- Mestrado em Química.  
Universidade Federal do Piauí, UFPI, Teresina, PI.  
Título: Desidrogenação do metanol promovida por complexos heterobimetálicos de Ru(II)-Sn(II) como catalisador: um estudo teórico.  
Orientador: Prof. Dr. Alexandre Araújo de Souza.  
Bolsista: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES.
- 2001-2005
- Graduação em Licenciatura Plena em Química.  
Universidade Estadual do Piauí, UESPI, Teresina, PI.  
Título: Análise dos aspectos físico-químicos e bacteriológicos de águas de lagoas da cidade de Teresina-PI.  
Orientador: Prof. MSc. Francílio de Carvalho Oliveira.  
Bolsista: Programa Institucional de Bolsas de Iniciação Científica, PIBIC.

## 2.0 Produção científica

### 2.1 Artigos Publicados

Muniz Filho, R. C. D.; Sousa, S. A. A.; Pereira, F. S.; Ferreira, M. M. C. Theoretical study of acid-catalyzed hydrolysis of epoxide. *The Journal of Physical Chemistry A*, 114, 5187-5194, 2010.

Muniz Filho, R. C. D.; Moura, E. M.; Rocha, W. R.; Souza, A. A. Methanol dehydrogenation promoted by a heterobimetallic Ru(II)-Sn(II) complex as catalyst: a density functional study. *Journal of Molecular Structure, TEOCHEM*, 816, 77-84, 2007.

## 2.2 Artigos Submetidos

Muniz Filho, R. C. D.; Volpe, P. L. O.; Duran, N.; Ferreira, M. M. C. The violaceín biosynthesis monitored by multi-wavelength fluorescence spectroscopy and PARAFAC. *Analytica Chimica Acta*.

Muniz Filho, R. C. D.; Tauler, R.; Ferreira, M. M. C. Exploring the violacein biosynthesis by application of multivariate curve resolution method on fused UV-VIS absorption, fluorescence and liquid chromatography mass spectrometry data. *Analytica Chimica Acta*.

Cruz, A. G.; Cadena, R. S.; Faria, J. A. F.; Bolini, H. M. A.; Muniz Filho, R. C. D.; Ferreira, M. M. C. PARAFAC: adjustment for modeling consumer studies. *Food Research International*

## 3.0 Trabalhos em congressos

Muniz Filho, R. C. D.; Volpe, P. L. O.; Duran, N.; Tauler, R. Ferreira, M. M. C. A kinetic model for violacein production by *Chromobacterium violaceum* obtained by multivariate curve resolution and UV-vis and fluorescence data, 2010, Antuérpia, Bélgica.

Muniz Filho, R. C. D.; Volpe, P. L. O.; Duran, N.; Ferreira, M. M. C. Obtaining information from a biosynthetic pathway using multi-wavelength fluorescence spectroscopy and Chemometric Tools. VII Colloquium Chemometricum Mediterraneum, 2010, Granada, Espanha.

Melo, M. P. R.; Muniz Filho, R. C. D. Bueno, M. I.M.; Ferreira, M. M. C. Potential of energy dispersive X-ray and Near infrared techniques coupled with chemometrics in the calibration of water and alcohol contents in samples of cachaça. VII Colloquium Chemometricum Mediterraneum, 2010, Granada, Espanha.

Sousa, S. A. A.; Muniz Filho, R. C. D.; Ferreira, M. M. C. Energética envolvida no rearranjo intramolecular indólico da biossíntese da violaceína: um estudo teórico. XV Simpósio Brasileiro de Química Teórica, 2009, Poços de caldas, Minas Gerais.

Martins, J. P. A.; Muniz Filho, R. C. D.; Silva, F. P.; Ferreira, M. M. C. Estudo teórico do mecanismo de abertura de anéis epoxídicos. XIV Simpósio Brasileiro de Química Teórica, 2007, Poços de caldas, Minas Gerais.

Muniz Filho, R. C. D.; Moura, E. M.; Rocha, W. R.; Souza, A. A. Methanol dehydrogenation using  $[\text{Ru}(\text{Cp})(\text{PH}_3)_2(\text{SnF}_3)]$ : a theoretical study. Brazillian Meeting on Inorganic Chemistry, 2006, Fortaleza, Ceará.

Muniz Filho, R. C. D.; Souza, A. A. Estudo teórico das frequências vibracionais do siligermano por diferentes níveis de teoria. Congresso Brasileiro de Química, 2005, Belém, Pará.

Muniz Filho, R. C. D.; Oliveira, F. C.; Marques, L. G. A. Análise dos aspectos físico-químicos existentes nas águas de lagoas da cidade de Teresina-PI. Congresso Brasileiro de Química, 2004, Fortaleza, Ceará.

#### 4.0 Disciplinas Cursadas no Doutorado

Disciplinas	Conceito
Introdução à quântica e espectroscopia	A
Tópicos especiais em físico-química XII - quimiometria	A
Métodos computacionais em álgebra linear	B
Planejamento e otimização de experimentos	A

#### 5.0 Outras Informações relevantes

Estágio de doutorado sanduíche durante 6 meses no grupo de quimiometria ambiental do professor Dr. Romà Tauler em Barcelona, Espanha.

Trabalho intitulado "Obtaining information from a biosynthetic pathway using multi-wavelength fluorescence spectroscopy and chemometric tools" agraciado como o melhor pôster apresentado no VII Colloquium Chemometricum Mediterraneum, 2010, Granada-Espanha.

Participação no programa de estágio à docência (PED) nas disciplinas físico-química II e físico-química experimental II.



## Resumo

A aplicação de métodos de resolução de curvas multivariada (MCR-ALS e PARAFAC) a dados de espectroscopia de fluorescência multicomprimento de onda, absorção molecular e cromatografia líquida de ultra eficiência com detecção por espectrometria de massas foi avaliada como uma nova e complementar metodologia para se obter informação inicial sobre a biossíntese de produtos naturais. Para demonstrar o poder dessa proposta, a biossíntese da violaceína pela *Chromobacterium violaceum* foi usada como exemplo. A violaceína é um metabólito secundário que exibe amplo espectro de atividade biológica e o entendimento de sua rota biossintética pode ser de grande interesse para o planejamento de fármacos. Para tanto, uma cultura da *C. violaceum* foi cultivada em biorreator, onde, a cada duas horas, alíquotas do meio de cultura eram coletadas para subseqüentes análises. Na primeira parte deste trabalho, as matrizes de emissão-excitação das amostras revelaram o comportamento dinâmico de fluoróforos envolvidos na biossíntese. O método de análise de fatores paralelos (PARAFAC) foi aplicado para resolver os perfis de fluorescência, onde seis espectros puros de emissão e excitação foram obtidos e suas relações com a biossíntese da violaceína estabelecidas. Na segunda parte, o método de resolução de curvas multivariada por quadrados mínimos alternados (MCR-ALS) foi utilizado para extrair informações simultâneas dos dados gerados por UV-VIS, fluorescência e UPLC-MS das amostras coletadas durante a biossíntese. Cinco compostos foram resolvidos e informações espectrais e temporais dessas espécies foram obtidas. Compostos com massas moleculares de 453, 465 e 479 u, até então não reportados na literatura como intermediários dessa biossíntese, foram identificados. Por fim, a metodologia baseada na aplicação de métodos de resolução de curvas MCR-ALS e PARAFAC provou ser uma poderosa ferramenta para investigar a biossíntese de produtos naturais.



## Abstract

The application of Multivariate Curve Resolution methods (MCR-ALS and PARAFAC) into multi-wavelength fluorescence spectroscopy, molecular absorption and ultra-performance liquid chromatography coupled to mass spectrometry data was evaluated as a new and complementary approach for the initial analysis of the biosynthesis of fluorescent natural products. To demonstrate the usefulness of this methodology, violacein biosynthesis by *Chromobacterium violaceum* was used as an example. Violacein is a secondary metabolite that exhibits a broad spectrum of biological activities and the understanding of its biosynthetic pathway can be of great interest for drug design. Therefore, a culture of *C. violaceum* was grown in a bioreactor from which aliquots were collected every two hours for subsequent analysis. In the first part of this work, the excitation–emission matrices of the samples have revealed the dynamic behavior of the fluorophores involved in this biosynthesis. The parallel factor analysis (PARAFAC) method was applied to resolve the fluorescence profiles where six pure emission and excitation spectra were obtained and their relationship with violacein biosynthesis established. In the second part, the multivariate curve resolution procedure based on alternating least squares optimization (MCR-ALS) was used for extracting information from fused data (UV-VIS absorption, fluorescence and ultra-performance liquid chromatography with mass spectrometry) gathered during the violacein biosynthesis. Five chemical components were resolved from which spectral and temporal information were obtained. Compounds with molecular masses 453, 465 and 479 u, until now unreported as intermediate of this biosynthesis, were identified. Finally, the methodology based on the application of curve resolution methods PARAFAC and MCR-ALS has proven to be a unique and powerful tool to achieve a deeper insight of the biosynthesis of natural products.



## Sumário

<b>LISTA DE ABREVIATURAS .....</b>	<b>XIX</b>
<b>LISTA DE FIGURAS .....</b>	<b>XXI</b>
<b>LISTA DE TABELAS .....</b>	<b>XXIII</b>
<b>INTRODUÇÃO .....</b>	<b>1</b>
Definições, notações e convenções .....	3
Resolução de curvas multivariada .....	7
MCR-ALS.....	19
PARAFAC.....	28
A violaceína e sua biossíntese.....	33
<b>OBJETIVOS.....</b>	<b>39</b>
<b>PARTE EXPERIMENTAL.....</b>	<b>41</b>
Organismo, meio de cultura e condições de cultivo.....	41
Coleta das amostras .....	41
Espectroscopia de fluorescência .....	41
Espectroscopia de absorção molecular na região UV-VIS .....	42
Cromatografia líquida de ultra eficiência-detector de massas (UPLC-MS).....	42
Modelagem dos dados .....	43
<b>RESULTADOS E DISCUSSÃO.....</b>	<b>47</b>
Fases de crescimento da bactéria.....	47
Dados de Fluorescência.....	48
Modelagem dos dados de fluorescência utilizando o método PARAFAC .....	52
Modelagem dos dados de fluorescência utilizando o método MCR-ALS .....	57

<b>Modelagem dos dados fundidos utilizando o método MCR-ALS .....</b>	<b>60</b>
<b>CONCLUSÕES .....</b>	<b>79</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>81</b>
<b>ANEXOS .....</b>	<b>89</b>

## LISTA DE ABREVIATURAS

**ACP** – Ácido Cromopirrólico

**AIP** – Ácido Indol Pirúvico

**ALS** – “Alternating Least Squares” - Quadrados Mínimos Alternados

**AsLS** – “Asymmetric Least Squares” – Quadrados Mínimos Assimétricos

**CORCONDIA** – “Core Consistency Diagnostic” – Diagnóstico de Consistência de Núcleo

**EFA** – “Evolving Factor Analysis” – Análise de Fatores Evolucionária

**FA** – “Factor Analysis” – Análise de Fatores

**HELP** – “Heuristic Evolving Latent Projection”

**ITFFA** – “Iterative Target Transformation Factor Analysis”

**MCR** – “Multivariate Curve Resolution” – Resolução de Curvas Multivariada

**MCR-ALS** – “Multivariate Curve Resolution – Alternating Least Squares”

**MEE** – Matriz de Emissão Excitação

**NADPH** – Nicotinamida Adenina Dinucleotídeo Fosfato Reduzida.

**NaN** – “Not a Number”

**PARAFAC** – “Parallel Factor Analysis” – Análise de Fatores Paralelos

**PARAFAC 2** – Análise de Fatores Paralelos 2

**PCA** – “Principal Component Analysis” – Análise de Componentes Principais

**PMT** – “Photomultiplier Tube” – Tubo fotomultiplicador

**RFA** – “Resolved Factor Analysis”

**SFA** – “Subwindow Factor Analysis”

**SIMPLISMA** – “Simple to Use Interactive Self Modeling Mixture Analysis”

**SVD** – “Singular Value Decomposition” – Decomposição de Valores Singulares

**TIC** – “Total Ion Chromatogram” – Cromatograma de Íons Totais

**UPLC-MS** – Cromatografia Líquida de Ultra eficiência-Acoplada a um Detector de Massas

**UV-VIS** – Ultravioleta-Visível

## **WFA – “Window Factor Analysis”**

## LISTA DE FIGURAS

<b>Figura 1.</b> Pesquisa realizada na base de dados “web of Science” com o radical “chemometr*” no campo tópicos, realizada em 06/04/2011.....	1
<b>Figura 2.</b> Representação gráfica de um “unfolding”. Um arranjo $\mathbf{X}$ ( $K \times I \times J$ ) é rearranjado em uma matriz desdobrada, $\mathbf{X}$ ( $I \times JK$ ).....	4
<b>Figura 3.</b> Representação do modelo de resolução de curvas multivariadas para uma matriz $\mathbf{D}$ constituída de $I$ amostras contendo $F$ componentes medidos em $J$ comprimentos de onda. ....	8
<b>Figura 4.</b> a) Representação gráfica dos dados cinéticos (simulados); b) logaritmo da magnitude dos valores singulares da matriz de dados. ....	10
<b>Figura 5.</b> Representação do mapa do posto químico em um cromatograma simulado. Os números na parte superior do gráfico indicam regiões com 0, 1, 2, 3 espécies químicas. ....	13
<b>Figura 6.</b> Possíveis configurações para matrizes com ordens em comum. a) matriz com uma ordem em comum e aumentada na direção das colunas; b) matriz com uma ordem em comum e aumentada na direção das linhas; c) matriz com duas ordens em comum, dados com estrutura “multi-way”. ....	15
<b>Figura 7.</b> Representação esquemática da análise de fatores evolucionária (EFA) em ambas as direções da matriz $\mathbf{D}$ . ....	20
<b>Figura 8.</b> a) Evolução dos valores singulares significativos da matriz $\mathbf{EFA\_F}$ ; b) involução dos valores singulares significativos da matriz $\mathbf{EFA\_T}$ ; c) evolução do posto da matriz $\mathbf{EFA\_F}$ ; d) rebaixamento do posto da matriz $\mathbf{EFA\_T}$ ; e) janela de concentração; f) perfis de concentração aproximados. ....	20
<b>Figura 9.</b> Efeitos de algumas restrições aplicadas na correção dos perfis de concentração. Adaptado da referência 60.....	23
<b>Figura 10.</b> Decomposição de um conjunto de dados “three-way” de acordo com um modelo trilinear expresso através de matrizes $\mathbf{C}$ , $\mathbf{Z}$ e $\mathbf{S}$ . $\mathbf{H}$ é um arranjo central binário de dimensões ( $F \times F \times F$ ).....	25
<b>Figura 11.</b> Implementação da restrição de trilinearidade no método MCR-ALS. Nesta representação $k=3$ e $f = 1,2...F$ .....	25
<b>Figura 12.</b> Remoção de picos de espalhamento de matrizes de emissão-excitação.....	32
<b>Figura 13.</b> Estrutura química da violaceína. ....	33
<b>Figura 14.</b> Mecanismo de reação para a biossíntese da violaceína. ....	36
<b>Figura 15.</b> Representação gráfica da decomposição de $\mathbf{D}$ por PARAFAC e de $\mathbf{D}_{\text{augment}}$ por MCR-ALS.....	43
<b>Figura 16.</b> (a) Espectros de absorção molecular na região UV-VIS dos extratos etanólicos; (b) Espectros de emissão dos extratos etanólicos quando excitados em 280 nm; (c) Cromatogramas de íons totais (TIC’s) dos extratos etanólicos. ....	44
<b>Figura 17.</b> Matriz de dados obtida para monitorar a biossíntese da violaceína. ....	45
<b>Figura 18.</b> Alterações visuais ocorridas no meio reacional durante a biossíntese da violaceína. a) amostras não-tratadas; b) amostras tratadas.....	47
<b>Figura 19.</b> Típica MEE do extrato etanólico da <i>C. violaceum</i> . ....	48

<b>Figura 20.</b> Evolução temporal dos fluoróforos envolvidos na biossíntese da violaceína. .	49
<b>Figura 21.</b> Superfície 3D e gráfico de contorno de uma MEE na região F. ....	50
<b>Figura 22.</b> Evolução temporal dos fluoróforos envolvidos na biossíntese da violaceína detectados na região espectral de menor energia.....	51
<b>Figura 23.</b> Estimativas dos perfis de concentração, emissão e excitação obtida por PARAFAC. i) toda a região espectral; ii) região B; iii) região C; iv) região D; v) região E; vi) região F.....	55
<b>Figura 24.</b> Visualização gráfica da matriz de dados formada pela fusão de diferentes técnicas analíticas.....	60
<b>Figura 25.</b> Gráfico dos valores singulares .....	61
<b>Figura 26.</b> Perfis de concentração dos componentes obtida por EFA. ....	61
<b>Figura 27.</b> Perfis de concentração e espectrais obtidos com MCR-ALS sob restrição de não-negatividade.....	62
<b>Figura 28.</b> Perfis espectrais obtidos por MCR-ALS com restrições de não-negatividade e trilinearidade nas matrizes de fluorescência. ....	63
<b>Figura 29.</b> Perfis de concentração e espectrais resolvidos por MCR-ALS sob restrições de não-negatividade e seletividade. ....	65
<b>Figura 30.</b> Perfis espectrais obtidos com MCR-ALS.....	67
<b>Figura 31.</b> Gráficos superiores, cromatogramas de íons totais; gráficos medianos, cromatograma de íons totais resolvidos por MCR-ALS; gráficos inferiores, correlogramas e produto cruzado dos cinco componentes modelados. ....	73
<b>Figura 32.</b> Espectros de massas obtidos no modo positivo via ionização por “electrospray” a) extrato etanólico da <i>C. Violaceum</i> em 1,1 minuto. B) solução padrão de L-triptofano. .	74
<b>Figura 33.</b> Espectros de massas de uma amostra de extrato etanólico da <i>C. Violaceum</i> nos tempos de retenções a) 4,60 min; b) 4,73 min; c) 4,47 min; d) 3,01 min; e) 2,61 min; f) 6,33 min. ....	75

## LISTA DE TABELAS

<b>Tabela 1.</b> Condições de medidas para cada região espectral. ....	51
<b>Tabela 2.</b> Variância explicada e CORCONDIA calculados para os modelos trilineares de 1 a 6 fatores com o método PARAFAC. ....	53
<b>Tabela 3.</b> Figuras de mérito obtidas para os modelos obtidos pelos métodos MCR-ALS e PARAFAC. ....	58
<b>Tabela 4.</b> Grau de similaridade entre os perfis recuperados por PARAFAC e MCR-ALS obtidos através do cálculo do arccosseno entre os dois vetores. Unidades em graus ( $^{\circ}$ ). 59	59

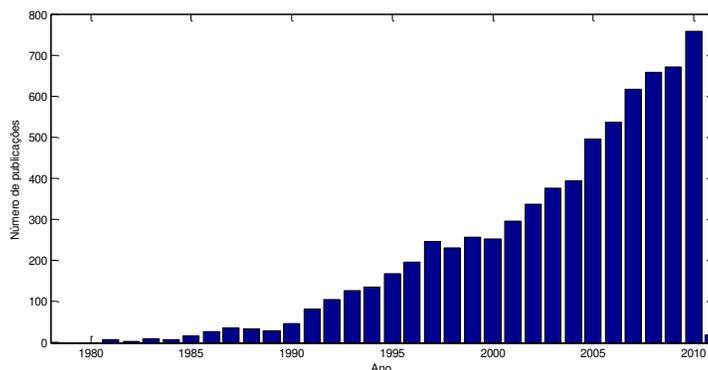


## INTRODUÇÃO

Há quase 40 anos, o professor Svante Wold definiu a quimiometria como a arte de extrair informação química relevante de dados produzidos experimentalmente<sup>1</sup>. Esta área da ciência, altamente interdisciplinar, utiliza conhecimentos da estatística multivariada, matemática aplicada e ciência da computação para tratar problemas da química, bioquímica, medicina, biologia e engenharia química.

Durante os últimos 30 anos, a quimiometria tem experimentado um grande crescimento, devido, principalmente, aos avanços no campo da informática, desenvolvimento de novas técnicas analíticas, bem como poderosos métodos matemáticos e estatísticos de análise<sup>2</sup>.

Para se ter uma ideia deste crescimento, realizou-se uma busca na base de dados “*web of Science*” utilizando o radical “*chemometr\**” no campo de busca tópicos, e o resultado obtido foi ordenado de acordo com o ano de publicação. O gráfico dessa pesquisa é apresentado na Figura 1, onde é possível observar um comportamento quase exponencial para o número de publicações nesta área durante as últimas décadas.



**Figura 1.** Pesquisa realizada na base de dados “*web of Science*” com o radical “*chemometr\**” no campo tópicos, realizada em 06/04/2011.

Ainda nessa pesquisa, pôde-se observar que o Brasil ocupa a oitava posição no cenário quimiométrico mundial, estando à frente de países como

## INTRODUÇÃO

---

Suécia, Austrália, Polônia, Dinamarca, Bélgica, Japão, Holanda, Noruega, dentre outros, com o total de 361 publicações.

As atividades de pesquisa em quimiometria podem ser agrupadas em oito áreas principais, a saber: calibração multivariada, análise exploratória e de classificação, planejamento e otimização de experimentos, resolução de curvas multivariada, processamento de sinal, controle estatístico multivariado de processos, figuras de mérito, métodos “multi-way” e de inteligência artificial.

No Brasil, cerca de 85% dos trabalhos em quimiometria referem-se às áreas de análise exploratória e classificação, calibração multivariada e planejamento e otimização de experimentos. A resolução de curvas multivariada ainda é muito pouco explorada em nosso país, apresentando, de acordo com a pesquisa realizada, apenas 8 artigos. Para se ter uma ideia, a Espanha e os Estados Unidos, países líderes nesta área, apresentam mais de 300 artigos cada um.

No sentido de divulgar a potencialidade desta área, o presente trabalho de tese versará sobre a aplicação de dois métodos de resolução de curvas multivariada a dados experimentais gerados por diferentes técnicas analíticas utilizadas para monitorar a biossíntese da violaceína.

A seguir, é apresentada uma revisão geral sobre métodos de resolução de curvas multivariada. Antes, a definição de alguns termos comumente utilizados, será abordada.

## Definições, notações e convenções

Booksh e Kowalski<sup>3</sup> propuseram a classificação das respostas produzidas por instrumentos, ou seja, os dados, de acordo com a álgebra tensorial. Assim, se uma medida de um sistema químico gera um único número como resposta, isto é, um escalar, como por exemplo, uma medida do pH de uma solução, diz-se que este dado é de ordem zero. Escalares serão simbolizados por letras minúsculas e em itálico,  $x$ .

Ao se coletar os valores de pH para  $I$  amostras tem-se um conjunto de números que podem ser arranjados na forma de um vetor. Letras minúsculas e em negrito serão utilizadas para representar vetores,  $\mathbf{x}$ . Por convenção, um vetor é um vetor coluna. Esta sequência de números deve ser analisada por métodos “one-way”, tais como, cálculos de média e desvio padrão.

Dados produzidos por instrumentos que geram uma sequência de números são classificados como de primeira ordem. Por exemplo, um espectro de UV-VIS com absorbâncias medidas em  $J$  comprimentos de onda. Ao se coletar estes espectros para  $I$  amostras, tem-se um conjunto de vetores que podem ser arranjados em uma matriz com dimensões  $I \times J$ . Matrizes são denotadas por letras maiúsculas e em negrito,  $\mathbf{X}$ . Este é um arranjo “two-way”, e pode ser analisado por métodos “two-way”, tais como, análise de componentes principais (PCA-“Principal Component Analysis”) e métodos de resolução de curvas multivariada (MCR-“Multivariate Curve Resolution”).

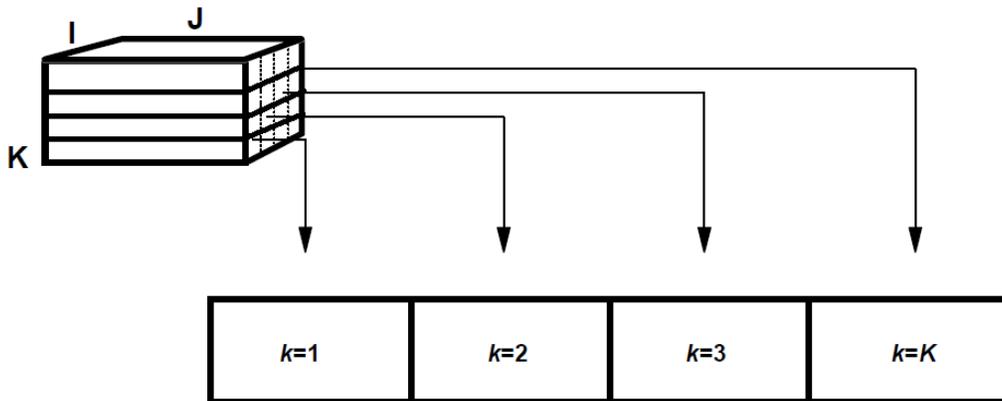
A transposição de uma matriz ou vetor é simbolizada por T sobrescrito,  $\mathbf{X}^T$  ou  $\mathbf{x}^T$ , respectivamente. A inversa da matriz  $\mathbf{X}$  é denotada por  $\mathbf{X}^{-1}$ . A pseudo-inversa de Moore-Penrose da matriz  $\mathbf{X}$  é representada por  $\mathbf{X}^+$  e corresponde a  $\mathbf{X}^+ = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ . A norma de Frobenius de uma matriz  $\mathbf{X}$  é simbolizada por  $\|\mathbf{X}\|$  e é calculada como a raiz quadrada da soma do quadrado dos seus elementos.

Se um instrumento hifenado produz uma matriz ( $I \times J$ ) de dados por amostra, diz-se que estes dados são de segunda ordem. Tomando  $k$  medidas em,

## INTRODUÇÃO

por exemplo, diferentes tempos, obtém-se um conjunto de matrizes que podem ser arrumadas em um arranjo “three-way” de dimensões  $k \times I \times J$ . Letras maiúsculas, em negrito, e sublinhadas, serão utilizadas para representar arranjos de N-ordens, tal que  $N > 2$ ,  $\underline{\mathbf{X}}$ . Arranjos com dimensões iguais ou superiores a três são classificados como dados de ordem superior ou “multi-way” e devem ser analisados por métodos adequados, tais como, análise de fatores paralelos (PARAFAC) e TUCKER3.

Os dados “three-way” podem ser rearranjados na forma de uma matriz desdobrada através de um procedimento ilustrado na Figura 2, conhecido como “unfolding”. Esse procedimento pode ser realizado nas direções  $I$ ,  $J$  e  $K$ .



**Figura 2.** Representação gráfica de um “unfolding”. Um arranjo  $\underline{\mathbf{X}}$  ( $K \times I \times J$ ) é rearranjado em uma matriz desdobrada,  $\mathbf{X}$  ( $I \times JK$ ).

Dois operações matriciais que são essenciais para o estudo de algoritmos que ajustam dados de ordem superior são os produtos de Kronecker e Khatri-Rao<sup>4</sup>.

O produto de Kronecker de duas matrizes  $\mathbf{A}$  e  $\mathbf{B}$  de tamanho  $M \times N$  e  $P \times Q$  é denotado pelo símbolo  $\otimes$  e é definido como na equação 1. O resultado é uma matriz de dimensões  $PM \times QN$ .

## INTRODUÇÃO

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \dots & a_{1N}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & \dots & a_{MN}\mathbf{B} \end{bmatrix} \quad \text{Eq.1}$$

Por exemplo, sejam  $\mathbf{A}$  e  $\mathbf{B}$  duas matrizes de dimensões  $3 \times 3$  e  $2 \times 3$ , respectivamente. O produto de Kronecker entre essas duas matrizes é:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} 1 & 3 & 5 & 2 & 6 & 10 & 3 & 9 & 15 \\ 2 & 4 & 6 & 4 & 8 & 12 & 6 & 12 & 18 \\ 4 & 12 & 20 & 5 & 15 & 25 & 6 & 18 & 30 \\ 8 & 16 & 24 & 10 & 20 & 30 & 12 & 24 & 36 \\ 7 & 21 & 35 & 8 & 24 & 40 & 9 & 27 & 45 \\ 14 & 28 & 42 & 16 & 32 & 48 & 18 & 36 & 54 \end{bmatrix}_{(6 \times 9)}$$

O produto de Khatri-Rao,  $\langle \odot \rangle$ , é um produto de Kronecker em blocos, onde as matrizes  $\mathbf{A}$  e  $\mathbf{B}$  são particionadas em blocos de tamanho  $M_I \times N_J$  e  $P_I \times Q_J$  para  $\sum_I M_I = M$ ,  $\sum_I P_I = P$ ,  $\sum_J N_J = N$  e  $\sum_J Q_J = Q$ .

$$\begin{aligned} \mathbf{A} \langle \odot \rangle \mathbf{B} &= \begin{bmatrix} [\mathbf{A}_{11}] & \dots & [\mathbf{A}_{1J}] \\ \vdots & \ddots & \vdots \\ [\mathbf{A}_{1I}] & \dots & [\mathbf{A}_{IJ}] \end{bmatrix} \langle \odot \rangle \begin{bmatrix} [\mathbf{B}_{11}] & \dots & [\mathbf{B}_{1J}] \\ \vdots & \ddots & \vdots \\ [\mathbf{B}_{1I}] & \dots & [\mathbf{B}_{IJ}] \end{bmatrix} \\ &= \begin{bmatrix} [\mathbf{A}_{11} \otimes \mathbf{B}_{11}] & \dots & [\mathbf{A}_{1J} \otimes \mathbf{B}_{1J}] \\ \vdots & \ddots & \vdots \\ [\mathbf{A}_{1I} \otimes \mathbf{B}_{1I}] & \dots & [\mathbf{A}_{IJ} \otimes \mathbf{B}_{IJ}] \end{bmatrix} \end{aligned} \quad \text{Eq.2}$$

Por exemplo, sejam  $\mathbf{A}$  e  $\mathbf{B}$  duas matrizes de dimensões  $3 \times 3$  e  $3 \times 3$ . O produto de Khatri-Rao entre essas duas matrizes é:

$$\mathbf{A} = \begin{bmatrix} [1 & 2] & [3] \\ [4 & 5] & [6] \\ [7 & 8] & [9] \end{bmatrix}, \mathbf{B} = \begin{bmatrix} [1] & [4 & 7] \\ [2] & [5 & 8] \\ [3] & [6 & 9] \end{bmatrix}$$

$$\mathbf{A} \langle \odot \rangle \mathbf{B} = \begin{bmatrix} 1 & 2 & 12 & 21 \\ 4 & 5 & 24 & 42 \\ 14 & 16 & 45 & 72 \\ 21 & 24 & 54 & 81 \end{bmatrix}$$

Outra definição importante a ser destacada é o conceito de linearidade. Suponha os seguintes modelos:

## INTRODUÇÃO

---

$$y_i = b_0 + b_1x_i + e_i; \quad i = 1, \dots, I \quad \text{Eq.3}$$

$$y_i = b_0 + b_1x_i + b_2x_i^2 + e_i; \quad i = 1, \dots, I \quad \text{Eq.4}$$

O modelo da equação 3 é linear tanto nos parâmetros quanto nas variáveis. O modelo da equação 4 é linear nos parâmetros, mas, não linear nas variáveis. O conceito de linearidade está associado com a forma em que os parâmetros se apresentam, e não as variáveis. Assim, na equação 4, mesmo o modelo sendo quadrático, o ajuste continua sendo linear, pois para um dado  $x$ ,  $y$  é uma função linear dos parâmetros  $b_0$ ,  $b_1$  e  $b_2$ .

Esta noção de linearidade pode ser estendida para a bilinearidade. Suponha um modelo PCA<sup>5</sup> com duas componentes principais para a matriz  $\mathbf{X}$ , com elementos:

$$x_{ij} = t_{i1}l_{1j} + t_{i2}l_{2j} + e_{ij}; \quad i = 1, \dots, I; j = 1, \dots, J \quad \text{Eq.5}$$

A parte modelada de  $\mathbf{X}$  é linear tanto em  $t_{i1}, t_{i2}$  para  $l_{1j}, l_{2j}$  fixos quanto em  $l_{1j}, l_{2j}$  para  $t_{i1}, t_{i2}$  fixos. Portanto, a equação 5 é chamada modelo bilinear de  $\mathbf{X}$ .

A noção de bilinearidade pode ser estendida para trilinearidade. Considere um modelo, com duas componentes, para o arranjo “three-way”  $\underline{\mathbf{X}}$ , com elementos:

$$x_{kij} = a_{k1}b_{1i}c_{1j} + a_{k2}b_{2i}c_{2j} + e_{kij}; \quad k = 1, \dots, K; i = 1, \dots, I; j = 1, \dots, J \quad \text{Eq.6}$$

A parte modelada de  $\underline{\mathbf{X}}$  é linear em  $a_{k1}, a_{k2}$  para  $b_{1i}, b_{2i}$  e  $c_{1j}, c_{2j}$  fixos,  $b_{1i}, b_{2i}$  para  $a_{k1}, a_{k2}$  e  $c_{1j}, c_{2j}$  fixos e em  $c_{1j}, c_{2j}$  para  $a_{k1}, a_{k2}$  e  $b_{1i}, b_{2i}$  fixos. Portanto, a equação 6 é chamada de modelo trilinear de  $\underline{\mathbf{X}}$ .

Por fim, outra definição importante é o conceito de posto de uma matriz. Matematicamente, o posto de uma matriz é definido como o menor valor entre o número de linhas ou colunas linearmente independentes. O posto matemático difere do posto químico (ou seja, o número de espécies químicas descritas de maneira linearmente independente pelo conjunto de dados), uma vez que os dados experimentais não estão livres de ruído. À frente, uma abordagem mais

adequada de como estimar o posto químico de uma matriz de dados será apresentada.

### Resolução de curvas multivariada

A Resolução de curvas multivariada (MCR-“Multivariate Curve Resolution”) descreve uma família de técnicas que ajudam a resolver misturas através da determinação de contribuições individuais das espécies químicas, mesmo quando nenhuma informação está disponível sobre sua natureza e composição<sup>6-10</sup>.

O principal pressuposto dessa família de técnicas é que os dados experimentais possam ser descritos por um modelo linear, similar à lei de Beer para misturas, onde a absorvância ( $A$ ) para um sistema multicomponente é a soma das absorvâncias individuais em um determinado comprimento de onda. Em outras palavras,

$$A = A_1 + A_2 + \dots + A_F = \varepsilon_1 b c_1 + \varepsilon_2 b c_2 + \dots + \varepsilon_F b c_F \quad \text{Eq.7}$$

onde  $\varepsilon_F$  é a absorvância molar da  $F$ -ésima espécie no comprimento de onda considerado;  $b$  é o caminho óptico e  $c_F$  é a concentração da  $F$ -ésima espécie. Caso o caminho óptico seja 1 cm, pode-se concluir que a absorvância total consiste no somatório do produto da absorvância molar pela concentração de cada espécie.

Para um conjunto de  $I$  amostras contendo  $F$  espécies e coletando as absorvâncias totais em  $J$  comprimentos de onda ( $\lambda_j$ ), tem-se:

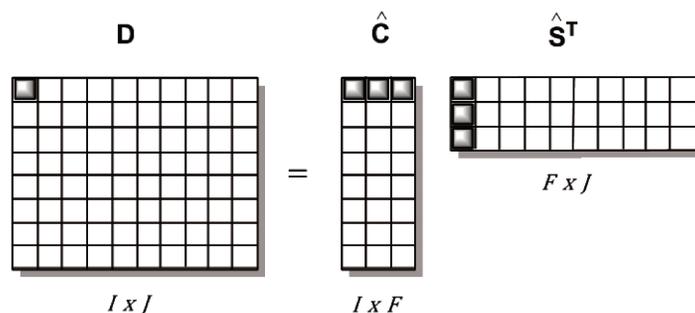
	$\lambda_1$	$\lambda_2$	$\dots$	$\lambda_J$
Amostra 1	$A_{(1,1)}$	$A_{(1,2)}$	$\dots$	$A_{(1,J)}$
Amostra 2	$A_{(2,1)}$	$A_{(2,2)}$	$\dots$	$A_{(2,J)}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Amostra $I$	$A_{(I,1)}$	$A_{(I,2)}$	$\dots$	$A_{(I,J)}$

## INTRODUÇÃO

Observe que a reunião das absorvâncias possibilita a obtenção de uma matriz de dados,  $\mathbf{D}$ . Os métodos baseados na resolução de curvas multivariada buscam decompor esta matriz  $\mathbf{D}$  em um produto de duas outras matrizes,  $\mathbf{C}$  e  $\mathbf{S}^T$ , que contêm informações sobre a concentração e absorvidade molar de cada espécie, respectivamente. Essa decomposição pode ser representada matematicamente pela equação

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{E} \quad \text{Eq.8}$$

e, visualmente, pela Figura 3, onde cada elemento da matriz  $\mathbf{D}$  corresponde ao produto interno entre um vetor linha da matriz  $\mathbf{C}$ , que contém a concentração de cada espécie na referida amostra, e um vetor coluna da matriz  $\mathbf{S}^T$ , que contém as absorvidades molares de cada espécie no comprimento de onda correspondente.  $\mathbf{E}$  é a matriz de resíduos que contém a parte não modelada.



**Figura 3.** Representação do modelo de resolução de curvas multivariadas para uma matriz  $\mathbf{D}$  constituída de  $I$  amostras contendo  $F$  componentes medidos em  $J$  comprimentos de onda.

Vale ressaltar que qualquer método baseado em uma técnica analítica que produza respostas linearmente aditivas pode ser tratado dessa maneira, por exemplo, espectroscopia de fluorescência, cromatografia e voltametria<sup>11</sup>.

Algumas questões surgem ao se analisar esta decomposição. A primeira, talvez seja como determinar o número de componentes, se nenhuma informação da natureza ou composição do sistema é conhecida? Como proceder para obter as matrizes  $\mathbf{C}$  e  $\mathbf{S}$ ?

## INTRODUÇÃO

---

Para responder estes questionamentos é necessário conhecer um pouco sobre a análise de fatores (FA – “Factor Analysis”)<sup>12</sup>, que é um método estatístico utilizado para descrever a variância dentre as variáveis observadas, em termos de um menor número de variáveis não observadas, chamadas fatores ou variáveis latentes. Essa análise baseia-se na decomposição de uma matriz no produto de duas, três ou mais matrizes utilizando algoritmos específicos para tais decomposições.

Um dos algoritmos mais utilizados é a decomposição de valores singulares (SVD-“Singular Value Decomposition”)<sup>13</sup>, na qual uma matriz **D** pode ser decomposta no produto de três matrizes, seguindo a equação 9.

$$\mathbf{D} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad \text{Eq.9}$$

As matrizes **U** ( $I \times I$ ), **S** ( $I \times J$ ) e **V**<sup>T</sup> ( $J \times J$ ) apresentam propriedades importantes. **S** é uma matriz diagonal que contém os valores singulares em ordem decrescentes. **U** e **V** são matrizes ortonormais, ou seja, são compostas de vetores ortogonais e normalizados. Essa última propriedade é representada em notação matricial como:

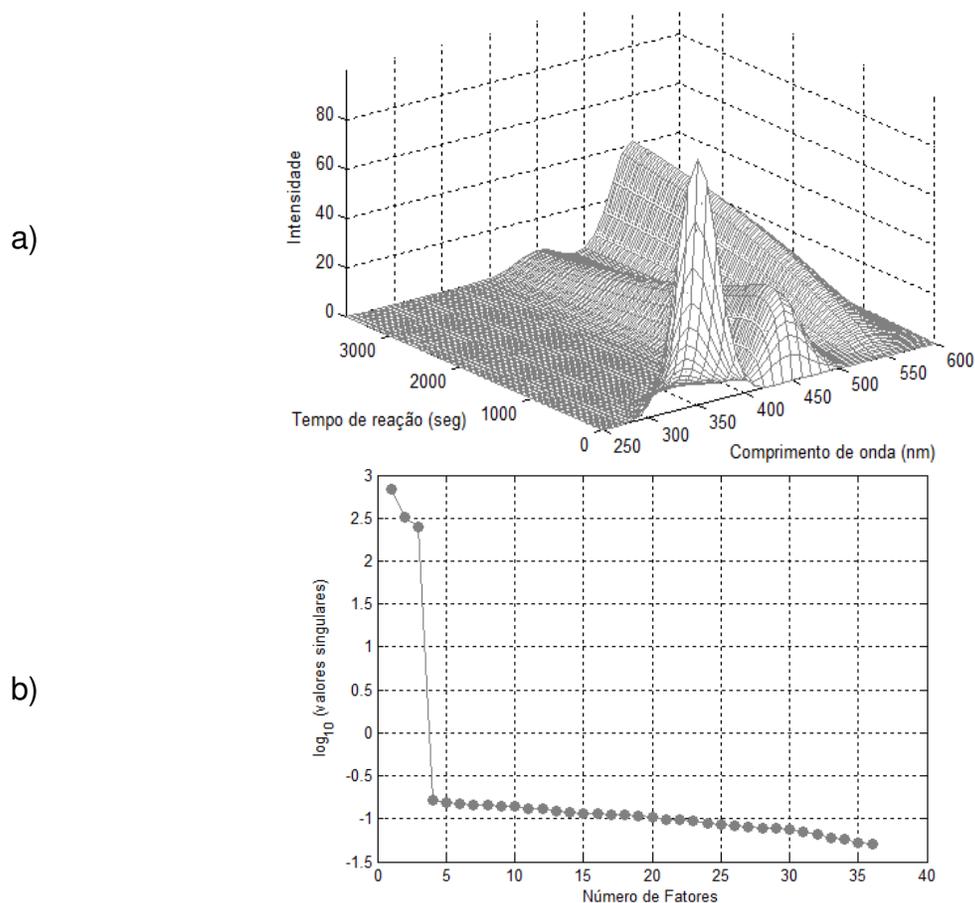
$$\mathbf{U}^T \mathbf{U} = \mathbf{I}; \mathbf{V} \mathbf{V}^T = \mathbf{I}, \quad \text{Eq.10}$$

onde **I** é uma matriz identidade ( $I \times I$ ) e ( $J \times J$ ), respectivamente.

Assumindo que os componentes químicos têm uma maior contribuição para a variância dos dados que outras fontes de variação, tais como: ruído, sinal de fundo e linha de base, é possível determinar o número de espécies (*F*) ou o posto químico de uma matriz de dados a partir da comparação dos respectivos valores singulares<sup>9</sup>. Veja o seguinte exemplo para esclarecer esta ideia.

Uma reação do tipo  $A \rightarrow B \rightarrow C$ , formada por duas reações consecutivas de primeira ordem, foi monitorada por espectroscopia de absorção molecular (Figura 4a). A matriz de dados obtida foi submetida a uma decomposição SVD, e os logaritmos dos valores singulares comparados graficamente (Figura 4b).

## INTRODUÇÃO



**Figura 4.** a) Representação gráfica dos dados cinéticos (simulados); b) logaritmo da magnitude dos valores singulares da matriz de dados.

Observe na Figura 4b que a distribuição logarítmica dos valores singulares diminui acentuadamente no quarto valor e então se estabiliza lentamente para os outros restantes. Como o quarto valor singular já se encontra ao nível de “ruído”, conclui-se que três é o posto químico ou pseudo-posto dessa matriz. De fato, a reação foi simulada seguindo uma cinética consecutiva de três espécies químicas ( $A$ ,  $B$  e  $C$ ). Vale destacar que nenhum conhecimento químico prévio foi necessário para a decomposição SVD, que é completamente matemática.

As matrizes obtidas na decomposição SVD podem ter suas dimensões reduzidas com base na determinação do posto químico. Assim,  $\mathbf{U}$ ,  $\mathbf{S}$  e  $\mathbf{V}^T$  passam a ter as dimensões  $\bar{\mathbf{U}}$  ( $I \times F$ ),  $\bar{\mathbf{S}}$  ( $F \times F$ ) e  $\bar{\mathbf{V}}^T$  ( $F \times J$ ) sem perder informação útil com relação à composição do sistema<sup>14</sup>.

## INTRODUÇÃO

---

A equação 9 pode, portanto, ser descrita como

$$\mathbf{D} = \mathbf{T} \mathbf{L}^T + \mathbf{E} \quad \text{Eq.11}$$

onde,  $\mathbf{T} = \bar{\mathbf{U}} \bar{\mathbf{S}}$  ( $I \times F$ ) é chamada de matriz de escores,  $\mathbf{L}^T = \bar{\mathbf{V}}^T$  ( $F \times J$ ) matriz de pesos e  $\mathbf{E}$  é a matriz de resíduos que contém a parte não modelada<sup>5</sup>.

A análise por SVD dá resultados únicos para escores e pesos, mas estes não têm, necessariamente, significado químico, ou seja, são fatores abstratos definidos de modo a maximizar a variância em  $\mathbf{D}$ . Nesse ponto é possível notar a semelhança entre um modelo puramente matemático, baseado na análise de fatores, representado na equação 11, e um modelo com significado químico apresentado na equação 8, onde as matrizes  $\mathbf{C}$  e  $\mathbf{S}$  contêm as informações relativas aos perfis de concentração e espectral de cada espécie, respectivamente.

Surge mais um questionamento. É possível relacionar as equações 8 e 11? A resposta é sim. Entretanto, é necessário encontrar uma única matriz de transformação,  $\mathbf{R}$ , onde  $\mathbf{R}^T = \mathbf{R}^{-1}$  que rotacione os fatores abstratos, escores e pesos, transformando-os em fatores com significado químico, perfis de concentração e espectrais. Esta transformação pode ser representada por:

$$\mathbf{D} = \mathbf{T} (\mathbf{R} \mathbf{R}^{-1}) \mathbf{L}^T + \mathbf{E} = (\mathbf{T} \mathbf{R})(\mathbf{L} \mathbf{R})^T + \mathbf{E} = \mathbf{C} \mathbf{S}^T + \mathbf{E} \quad \text{Eq.12}$$

O problema encontrado nessa operação é que qualquer matriz ortonormal de dimensões ( $F \times F$ ) pode ser tomada como uma matriz de transformação  $\mathbf{R}$ . Lembrando que  $\mathbf{R} \mathbf{R}^{-1} = \mathbf{I}$ , pode-se perceber que a rotação dos fatores abstratos não modifica o ajuste do modelo, ou seja,  $\mathbf{E} = \mathbf{D} - \mathbf{T} \mathbf{L}^T = \mathbf{D} - (\mathbf{T} \mathbf{R})(\mathbf{R}^{-1} \mathbf{L}^T) = \mathbf{D} - \mathbf{C} \mathbf{S}^T$ . Portanto, existe um número infinito de soluções possíveis e matematicamente equivalentes de  $\mathbf{C}$  e  $\mathbf{S}$ , e, devido a este fato, diz-se que esta decomposição sofre ambiguidade rotacional. Isto significa que os espectros ou perfis de concentração estimados para qualquer um dos componentes serão combinações lineares desconhecidas de seus espectros puros e dos verdadeiros perfis de concentração.

## INTRODUÇÃO

---

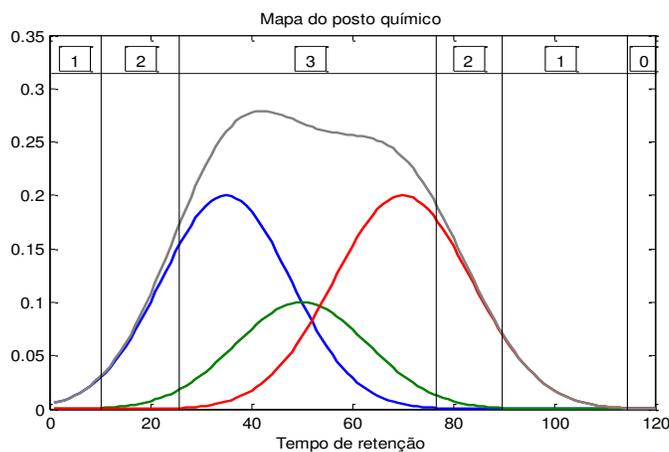
É possível reduzir consideravelmente este número infinito de possíveis soluções por meio de restrições derivadas de propriedades químicas ou matemáticas<sup>15</sup>. Quando uma restrição é implementada, ela deve ser expressa na forma de uma condição matemática (igualdade ou desigualdade). Por exemplo, as restrições de igualdade forçam os perfis ou alguns elementos dos perfis a serem iguais a determinados valores. Já as restrições de desigualdade fazem com que estes sejam maiores ou menores que dados valores. Em geral, as restrições de desigualdade são preferidas, pois modificam os perfis mais suavemente, minimizando perturbações na convergência do processo de resolução<sup>15</sup>.

Uma restrição natural que usualmente é aplicada às matrizes **C** e **S**, é a restrição de não-negatividade<sup>6,7,10,16,17</sup>. Esta advém do fato de que somente valores positivos para as concentrações e intensidades espectrais (registradas em muitos métodos espectroscópicos) têm significado físico. Assim, se um dado perfil deve ser não-negativo, os elementos negativos deste podem ser forçados a serem iguais a zero (restrição de igualdade) ou serem maiores que um valor negativo muito pequeno, geralmente na ordem do ruído (restrição de desigualdade). Outras restrições tais como unimodalidade (que fazem com que os perfis tenham um único máximo, tais como nos picos cromatográficos e nos sinais voltamétricos)<sup>10,16-20</sup>, balanço de massa ou “closure” (que força a soma dos perfis de concentração ser constante durante todo o processo)<sup>10,16,21,22</sup> e “hard modeling” (que fazem com que os perfis sigam modelos pré-definidos segundo as leis da físico-química)<sup>23-27</sup> também são restrições associadas às propriedades químicas.

As restrições de cunho matemático, geralmente referem-se às propriedades dos subespaços do conjunto de dados<sup>15</sup>. Neste caso, a matriz é dividida em pequenas partes e os postos químicos destas partes são obtidos<sup>28</sup>. Para facilitar o entendimento deste tipo de restrição, um exemplo de mapeamento do posto químico de um cromatograma é ilustrado na Figura 5. Aqui, é possível observar, através das curvas deconvoluídas, diferentes regiões que trazem informações sobre os componentes da amostra, a saber: i) região onde não há compostos (região zero). Essa informação é útil na resolução, pois indica onde não existem

## INTRODUÇÃO

espécies químicas; ii) região com informação seletiva (região 1). Nesse caso, o sinal registrado refere-se a apenas um composto, e essa informação é ideal para se resolver os perfis sem ambiguidade; iii) região parcialmente seletiva (região 2). O sinal obtido não traz informações sobre todos os compostos. Esta também é útil na resolução, pois indica onde determinadas espécies não existem; iv) região não seletiva (região 3). Todos os compostos aparecem nesse intervalo. Essa região não é útil para eliminar a ambiguidade rotacional.



**Figura 5.** Representação do mapa do posto químico em um cromatograma simulado. Os números na parte superior do gráfico indicam regiões com 0, 1, 2, 3 espécies químicas.

Estas restrições matemáticas estão relacionadas com a informação do posto local<sup>28</sup>, e a principal restrição dentro desse grupo é a de seletividade, que define janelas de concentração ou canais de resposta onde somente um componente contribui para o sinal. Nestas linhas da matriz  $\mathbf{C}$  ou colunas da matriz  $\mathbf{S}^T$ , os outros componentes são forçados a serem ausentes<sup>10</sup>.

A relevância desta restrição é facilmente compreendida, pois, o espectro puro de um composto pode ser obtido diretamente pelo espectro medido dentro de uma janela de concentração seletiva. O mesmo é válido para canais espectrais seletivos que permitem estimar perfis de concentrações puros. Como consequência, a ambiguidade rotacional na resolução final diminui drasticamente, já que boas definições dos perfis puros estão disponíveis a partir da informação seletiva. É importante notar que mesmo que não exista seletividade, a informação

## INTRODUÇÃO

---

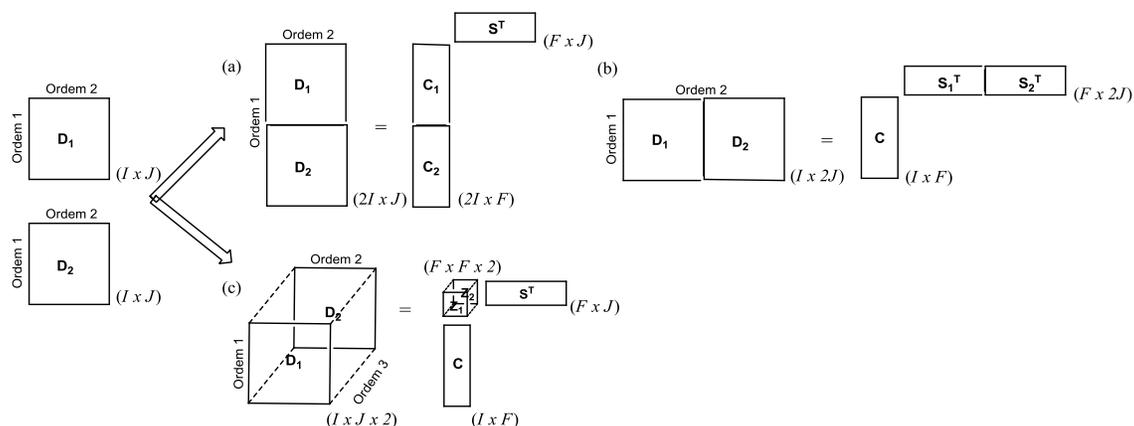
de posto local (regiões 2 na Figura 5) pode ser incorporada para reduzir o número de possíveis soluções inerentes à ambiguidade rotacional.

Se uma única matriz de dados é analisada, havendo seletividade para todas as espécies em uma das ordens (concentração ou espectral), a ambiguidade rotacional pode ser completamente superada e a verdadeira solução obtida. Quando não existe seletividade em nenhuma das duas ordens para qualquer componente, não existe garantia que as soluções corretas serão recuperadas<sup>10</sup>.

Outra forma de diminuir significativamente o número de possíveis soluções e resolver problemas de deficiência de posto pode ser alcançada quando mais de uma matriz, com pelo menos uma ordem em comum, estão disponíveis e são analisadas simultaneamente<sup>8,10,16,29,30</sup>. Duas ou mais matrizes têm uma ordem em comum quando compartilham seu espaço-linha ou espaço-coluna (possuem perfis de concentração ou espectrais em comum) e apresentam sincronização, isto é, possuem os mesmos índices nas linhas ou colunas das matrizes de dados.

A justaposição de matrizes na direção apropriada permite a criação de matrizes aumentadas que podem conter características úteis para a resolução. Por exemplo, suponha que o progresso da reação  $A + B \rightarrow C$  seja monitorado espectroscopicamente, onde todas as espécies são ativas. A reação global é assumida para ser de segunda ordem. Os perfis de concentração das espécies  $A$  e  $B$  têm a mesma forma, mas diferem em suas concentrações iniciais. A avaliação do posto dessa matriz indica que seu posto químico é dois. Uma vez que as três espécies são detectadas pela técnica utilizada, esperava-se que o valor de posto fosse três. Entretanto, devido à dependência linear entre os perfis de concentração das espécies  $A$  e  $B$ , o posto é, de fato, dois. Repetindo esse experimento com uma concentração inicial de  $B$  em excesso, essa reação seria de pseudo-primeira ordem e, basicamente, toda a variação existente nos dados seria relativa ao consumo da espécie  $A$  e formação da espécie  $C$ . As duas matrizes de dados assim obtidas, podem ser arranjadas para formar uma matriz aumentada que compartilhe o mesmo espaço-coluna, ou seja, apresentem os mesmos perfis

espectrais, mas que possam diferentes perfis de concentração. Esta configuração é representada na Figura 6a.



**Figura 6.** Possíveis configurações para matrizes com ordens em comum. a) matriz com uma ordem em comum e aumentada na direção das colunas; b) matriz com uma ordem em comum e aumentada na direção das linhas; c) matriz com duas ordens em comum, dados com estrutura “multi-way”.

Imagine agora outro exemplo no qual uma reação química é monitorada por diferentes técnicas analíticas. As matrizes individuais podem ser arranjadas de modo que o espaço-linha, referente aos mesmos tempos de reação, sejam comuns. Neste caso, as espécies químicas possuem os mesmos perfis de concentração, mas suas respostas espectrais serão completamente diferentes dependendo da técnica analítica considerada. Esta configuração é representada na Figura 6b.

A vantagem deste procedimento para matrizes com uma ordem em comum é que as boas características apresentadas por uma ou algumas das matrizes incluídas, sempre terão efeitos positivos na resolução, introduzindo condições de seletividade, de posto local favorável ou ainda, resolvendo problemas de posto<sup>31</sup>.

Para dados com duas ordens em comum, um arranjo “three-way”, conforme visualizado na Figura 6c, pode ser obtido. Neste caso, o espaço-linha e o espaço-coluna podem ter as mesmas bases para cada matriz, ou seja, os mesmos

## INTRODUÇÃO

---

conjuntos de perfis espectrais e de concentração descrevem a variação em mais de uma matriz, somente em diferentes proporções<sup>32</sup>.

Essa configuração particular é chamada trilinearidade e pode ser representada matematicamente pela equação:

$$\mathbf{D}_k = \mathbf{C} \mathbf{Z}_k \mathbf{S}^T + \mathbf{E} \quad \text{Eq.13}$$

onde  $\mathbf{Z}_k$  é uma matriz diagonal que guarda as proporções de cada componente em cada matriz  $\mathbf{D}_k$ .

Em 1944, Cattell<sup>33</sup> foi o primeiro a mostrar que a presença de perfis proporcionais paralelos deveria levar a uma decomposição livre de ambiguidade rotacional, fato que pode ser compreendido multiplicando a equação 13 pela matrizes de transformação  $\mathbf{X}$  e  $\mathbf{Y}$ . Assim,

$$\mathbf{D}_k = \mathbf{C} \mathbf{Z}_k \mathbf{S}^T + \mathbf{E} = \mathbf{C} \mathbf{X} \mathbf{X}^{-1} \mathbf{Z}_k \mathbf{Y} \mathbf{Y}^{-1} \mathbf{S}^T + \mathbf{E} \quad \text{Eq.14}$$

Esta expressão implica que ao invés de  $\mathbf{C}$ ,  $\mathbf{Z}_k$  e  $\mathbf{S}^T$  os pesos  $\mathbf{C} \mathbf{X}$ ,  $\mathbf{X}^{-1} \mathbf{Z}_k \mathbf{Y}$  e  $\mathbf{Y}^{-1} \mathbf{S}^T$  podem ser igualmente usados. Entretanto,  $\mathbf{X}^{-1} \mathbf{Z}_k \mathbf{Y}$  deve ser diagonal, já que  $\mathbf{Z}_k$  é uma matriz diagonal. Devido a esta condição, somente matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  muito especiais podem ser válidas para preservar a diagonalidade. Essas matrizes são conhecidas como matrizes de permutação ou de escala, e geram ambiguidades conhecidas como ambiguidades de permutação e de intensidade<sup>34</sup>.

Portanto, para um arranjo “three-way”, a trilinearidade é uma situação desejada, já que sob tal circunstância, a ambiguidade rotacional inerente aos métodos bilineares é totalmente resolvida e a recuperação dos perfis das espécies em ambas as ordens de medidas pode ser obtida<sup>35</sup>. Este fato tem importante consequência na análise de misturas, onde o problema de quantificação de um analito em uma amostra real, na presença de interferentes desconhecidos e não calibrados, pode ser resolvido sem tratamento prévio de amostras ou separação física do analito (vantagem de segunda ordem)<sup>36</sup>.

## INTRODUÇÃO

---

O cumprimento total da trilinearidade é uma exigência necessária. Entretanto, ela não é estritamente preenchida em muitas aplicações práticas. Por exemplo, em técnicas cromatográficas hífenadas, geralmente, a ordem referente aos tempos de eluição não é comum entre corridas, uma vez que as formas dos perfis de eluição de componentes coeluídos variam entre diferentes corridas. Já na espectroscopia de fluorescência multicomprimentos de onda, a trilinearidade é facilmente satisfeita, uma vez que existem duas ordens espectrais (emissão e excitação) altamente reproduzíveis<sup>32</sup>.

A trilinearidade pode ser checada através da análise de posto por SVD das matrizes aumentadas nas direções das colunas e linhas (Figuras 6a e 6b). Se os números de valores singulares significativos nas duas matrizes aumentadas forem iguais, então se pode inferir que a trilinearidade está sendo cumprida<sup>10,32,37</sup>.

Como comentado anteriormente, embora a ambiguidade rotacional pudesse ser completamente resolvida para matrizes com duas ordens em comum, seguindo estritamente a condição de trilinearidade, outros tipos de ambiguidades ainda se fazem presentes (ambiguidades de permutação e de intensidade). Estas também são encontradas para os casos de matrizes sem e com uma ordem em comum, mesmo existindo uma única matriz de transformação.

Por ambiguidade de permutação entende-se que existe uma incerteza na sequência dos perfis resolvidos. Já a ambiguidade de intensidade compreende a multiplicação de qualquer perfil de concentração por um escalar e a multiplicação do espectro correspondente pelo inverso do escalar. Isso não tem um efeito líquido na resolução<sup>11</sup>. Matematicamente, essa última ambiguidade pode ser ilustrada pela equação

$$\mathbf{c}_F \mathbf{s}_F^T = (\mathbf{c}_F r) (r^{-1} \mathbf{s}_F^T) \quad \text{Eq.15}$$

onde  $\mathbf{c}_F$  e  $\mathbf{s}_F^T$  correspondem aos perfis de concentração e espectral da  $F$ -ésima espécie e  $r$  é um escalar qualquer.

## INTRODUÇÃO

---

Devido à inevitável ambiguidade de intensidade, somente as formas dos perfis de concentração e espectrais podem ser determinadas diretamente pelos métodos de resolução de curvas multivariada, salvo quando os métodos MCR são aplicados simultaneamente a múltiplos conjuntos de dados com uma ou duas ordens em comum. Nesse caso, é possível recuperar a informação de escala (ou informação quantitativa) através da comparação da intensidade ou área da resolução para uma espécie em uma amostra desconhecida, com a intensidade ou a área da resolução de uma ou diversas amostras padrões do mesmo analito em concentrações conhecidas<sup>11</sup>.

Os métodos clássicos de resolução de curvas podem ser classificados em dois tipos: iterativos e não-iterativos. Muitos dos métodos não-iterativos baseiam-se na combinação de informações de pequenas seções do conjunto de dados (subespaços) para obter os perfis dos componentes puros. Estes subespaços podem ser janelas de concentrações ou regiões na matriz de dados que possuem propriedades particulares (presença ou ausência de componentes particulares). Os perfis de **C** ou **S** são recuperados uma única vez, e a matriz correspondente é obtida através de uma simples etapa de quadrados mínimos. “Window Factor Analysis” (WFA)<sup>38</sup>, “Subwindow Factor Analysis” (SFA)<sup>39</sup> ou “Heuristic Evolving Latent projections” (HELP)<sup>40-42</sup> são os primeiros e mais significativos métodos dentro desta categoria. A utilização de métodos não-iterativos ganhou força com o crescente uso de instrumentos hifenados que produzem vasta quantidade de dados.

Métodos iterativos de resolução são atualmente considerados os mais populares devido à flexibilidade para tratar com diferentes tipos de estruturas de dados e problemas químicos, além da habilidade de acomodar informação externa no processo de resolução. Todos eles compartilham uma etapa de otimização em comum (das matrizes **C** ou **S**), começando com estimativas iniciais de **C** ou **S** que evoluem para produzir perfis com formas quimicamente significativas, adaptados de acordo com a informação química ou matemática no processo de otimização, sob a forma de restrições. “Iterative Target Transformation Factor Analysis”

(ITFFA)<sup>17</sup> e “Multivariate Curve Resolution – Alternating Least Squares” (MCR-ALS)<sup>10,16</sup> foram os primeiros métodos, embora outras metodologias com diferentes princípios, tal como, “Resolving Factor Analysis” (RFA)<sup>43</sup> tenham aparecido depois.

Embora não incluídos na definição clássica de métodos de resolução de curvas, do qual apenas métodos bilineares fazem parte, algumas metodologias “multi-way”, tais como, PARAFAC<sup>34,44,45</sup>, PARAFAC2<sup>34,46</sup> e TUCKER<sup>34,47,48</sup> também são utilizadas para este fim, e podem ser classificadas como métodos iterativos.

Dentre os métodos de resolução de curvas, a utilização dos dois métodos mais populares, MCR-ALS e PARAFAC, será enfocada.

### **MCR-ALS**

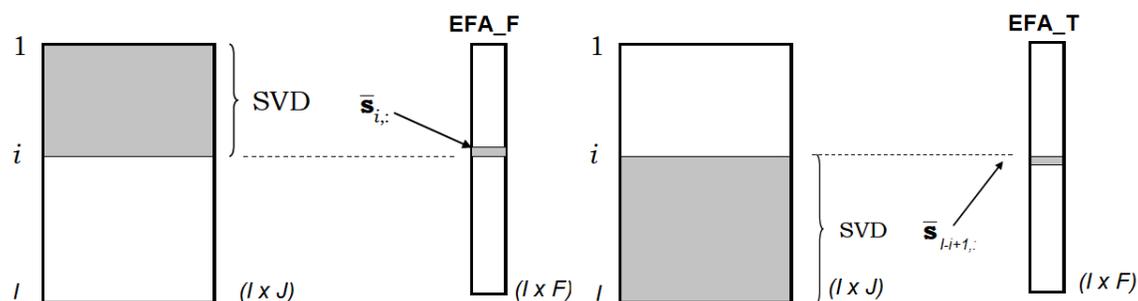
O método MCR-ALS foi proposto e desenvolvido pelo grupo do professor Tauler<sup>8-10,16,19,32</sup>. Este método resolve iterativamente a equação 8 por um algoritmo de quadrados mínimos alternados (ALS) restrito que calcula as matrizes de concentração e espectral, **C** e **S**, ajustando-as à matriz de dados **D**.

Essa otimização é realizada para um número proposto de componentes e utiliza estimativas iniciais para **C** ou **S**. A determinação do número de componentes em **D**, pode ser feita a partir da comparação dos respectivos valores singulares (via SVD) ou por conhecimento prévio do sistema em estudo. Já as estimativas iniciais das matrizes **C** ou **S** devem ser obtidas usando métodos de resolução orientada, como, por exemplo, os métodos de análise de posto local (EFA e suas derivações<sup>22,50-56</sup>), e/ou métodos de seleção de variáveis puras, como, SIMPLISMA e suas derivações<sup>49,57,58</sup>.

No primeiro caso, esses métodos realizam repetidas decomposições por SVD em pequenas partes ou submatrizes da matriz de dados, com o objetivo de conhecer como o número e a distribuição dos componentes evoluem. Essas janelas podem ter tamanhos gradualmente crescentes ou podem ser fixas, e em ambos os casos, movem-se ao longo da matriz de dados. Na análise EFA tradicional, estas submatrizes são formadas pelos *i* primeiros espectros de **D**, onde *i* aumenta de 1 até o número total de espectros *I*. Os valores singulares

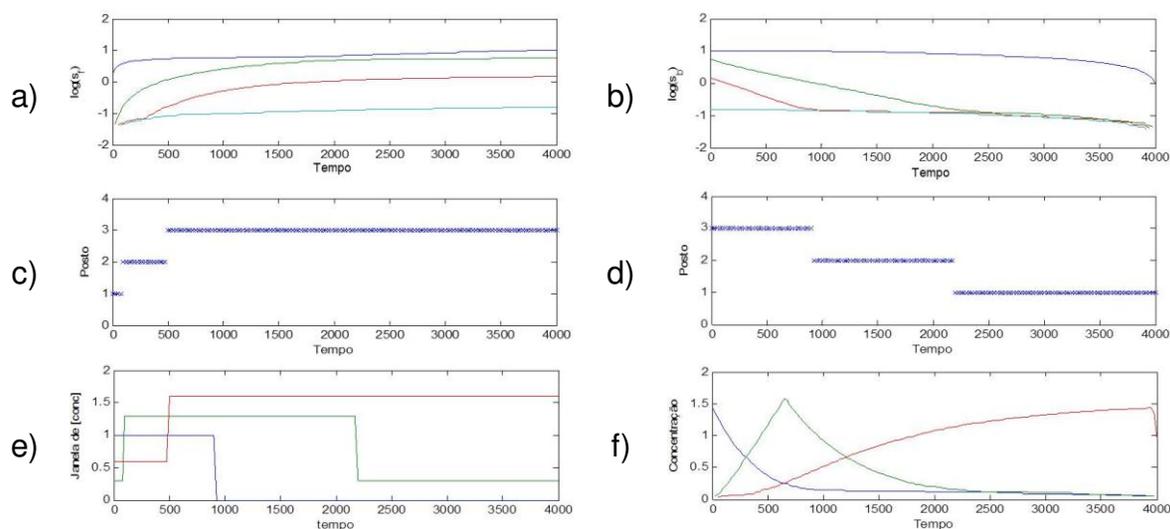
## INTRODUÇÃO

significativos são armazenados na matriz **EFA\_F**. O mesmo procedimento é novamente realizado na matriz **D**. Entretanto, as submatrizes são formadas pelos  $(l - i + 1)$  últimos espectros de **D**, aonde  $i$  vai de 1 até  $l$ . Nesse caso, os valores singulares significativos são armazenados na matriz **EFA\_T**. Este procedimento é explicado graficamente na Figura 7. As submatrizes de cor cinza são submetidas à análise SVD, e os  $F$  valores singulares significativos são armazenados em uma linha das matrizes **EFA\_F** e **EFA\_T**, respectivamente.



**Figura 7.** Representação esquemática da análise de fatores evolucionária (EFA) em ambas as direções da matriz **D**.

O logaritmo dos valores singulares armazenados nas matrizes **EFA\_F** e **EFA\_T** são plotados para facilitar a interpretação (Figuras 8a e 8b).



**Figura 8.** a) Evolução dos valores singulares significativos da matriz **EFA\_F**; b) evolução dos valores singulares significativos da matriz **EFA\_T**; c) evolução do posto da matriz **EFA\_F**; d) rebaixamento do posto da matriz **EFA\_T**; e) janela de concentração; f) perfis de concentração aproximados.

## INTRODUÇÃO

---

O aparecimento de um novo composto durante a aquisição do dado é indicado pelo surgimento de um novo valor singular significativo na matriz **EFA\_F**, ou seja, pelo aumento do posto nesta matriz (Figura 8c). Enquanto que o desaparecimento de um composto é indicado por um rebaixamento de posto na matriz **EFA\_T** (Figura 8d).

Os métodos de análise de posto local são particularmente relevantes em estudos de processos, onde os perfis de concentração de diferentes componentes evoluem, muitas vezes, seguindo um padrão sequencial, ou seja, o primeiro componente que surge é o primeiro a desaparecer e assim por diante. Essa propriedade permite estabelecer janelas de concentração (Figura 8e), ou seja, regiões ao longo do eixo do tempo onde os compostos existem, além de um perfil de concentração aproximado para os diferentes componentes (Figura 8f). Esse último é obtido pela combinação dos gráficos **EFA\_F** e **EFA\_T** utilizando o menor valor de cada par.

Outra maneira implementada no método MCR-ALS de se obter estimativas iniciais das matrizes **C** ou **S** consiste em utilizar métodos de seleção de variáveis puras (seletivas). Estes métodos encontram os perfis linha ou coluna mais representativos para diferentes componentes em um conjunto de dados. Muitos deles são úteis indiretamente para determinar o número de fatores de uma matriz e, quando condições seletivas são encontradas, os perfis linha ou coluna podem ser diretamente associados com perfis de concentração ou espectral de maneira que a resolução do sistema pode ser obtida.

Uma característica importante desses métodos é que, diferentemente dos métodos de posto local, eles não são afetados pela falta de ordem na direção das linhas ou colunas. Alguns deles trabalham no espaço abstrato das componentes principais, enquanto outros usam o espaço das medidas reais. Esses últimos são comumente mais utilizados e dentre eles, o método SIMPLISMA é o mais popular. Este método assume que cada componente em uma mistura tem uma variável com intensidade finita para um dado componente e intensidade zero para os demais. Ele utiliza um diagrama de variância e critérios de dissimilaridade para

## INTRODUÇÃO

---

encontrar as variáveis puras que em seguida são extraídas uma a uma da matriz de dados original até que esta só contenha ruído<sup>59</sup>.

Depois de definido o número de componentes e a estimativa inicial da matriz  $\mathbf{C}$  ou da matriz  $\mathbf{S}$ , o núcleo do método MCR-ALS trabalha para resolver, sob restrições adequadas, os dois seguintes problemas de quadrados mínimos:

$$\min_{\hat{\mathbf{C}}} \|\hat{\mathbf{D}}_{\text{PCA}} - \hat{\mathbf{C}}\hat{\mathbf{S}}^T\| \quad \text{Eq.16}$$

$$\min_{\hat{\mathbf{S}}^T} \|\hat{\mathbf{D}}_{\text{PCA}} - \hat{\mathbf{C}}\hat{\mathbf{S}}^T\| \quad \text{Eq.17}$$

Nessas duas equações, a norma de Frobenius dos resíduos é minimizada pelas estimativas de  $\mathbf{C}$  e  $\mathbf{S}$ ,  $\hat{\mathbf{C}}$  e  $\hat{\mathbf{S}}^T$ , obtidas no procedimento ALS.

O princípio do algoritmo ALS consiste simplesmente em dividir os parâmetros em diversos conjuntos e estimá-los por quadrados mínimos condicionando-os aos parâmetros restantes. As soluções por quadrados mínimos das equações 16 e 17 são:

$$\hat{\mathbf{S}}^T = \hat{\mathbf{C}}^+ \hat{\mathbf{D}}_{\text{PCA}} \quad \text{Eq.18}$$

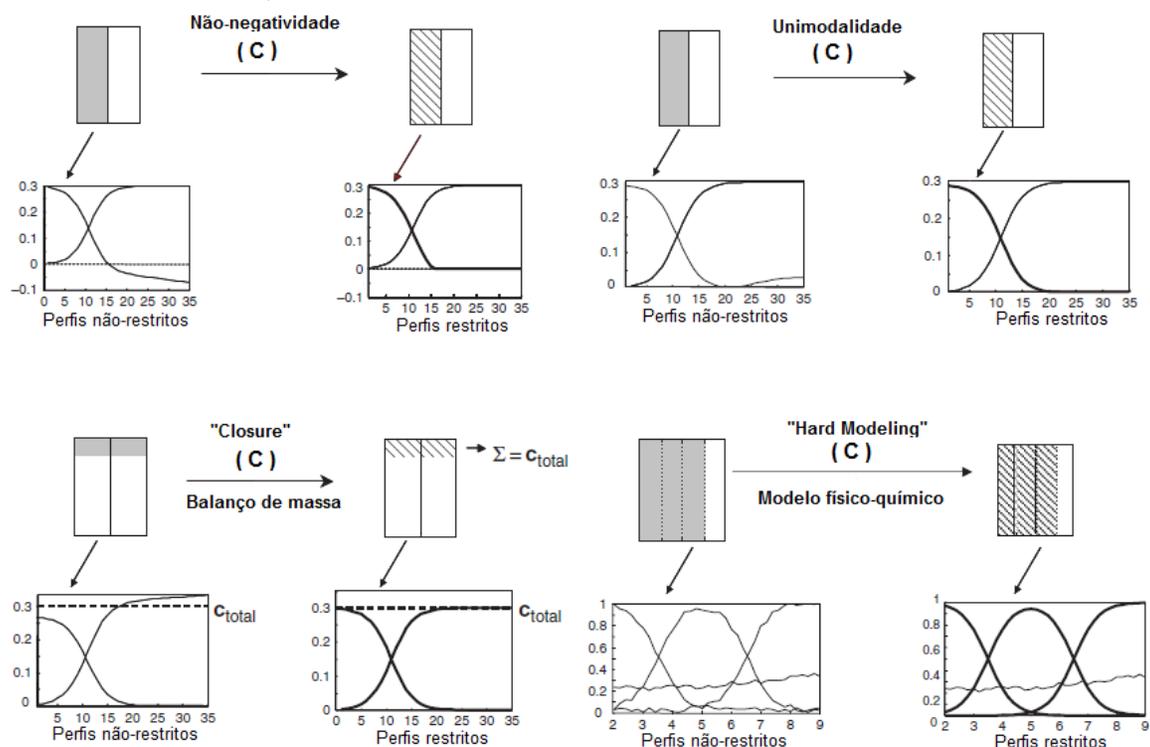
$$\hat{\mathbf{C}} = \hat{\mathbf{D}}_{\text{PCA}} (\hat{\mathbf{S}}^T)^+ \quad \text{Eq.19}$$

Elas são resolvidas sequencialmente em cada ciclo de forma que com uma estimativa inicial da matriz de concentração  $\mathbf{C}$  e a matriz  $\hat{\mathbf{D}}_{\text{PCA}}$ , é possível calcular uma estimativa para  $\mathbf{S}$  (Eq.18) e então utilizá-la para obter uma nova estimativa da matriz  $\mathbf{C}$ , e assim sucessivamente até que a convergência seja atingida. A convergência é obtida quando a diferença nos desvios padrões dos resíduos entre os valores experimentais e os calculados em dois ciclos consecutivos, é menor que um valor previamente selecionado, geralmente escolhido como 0,1%. Outra possibilidade consiste em determinar um número máximo de ciclos iterativos como critério de parada<sup>15</sup>.

Durante a otimização pelo algoritmo ALS, diversas restrições podem ser aplicadas para modelar a forma dos perfis em  $\mathbf{C}$  e  $\mathbf{S}$ , tais como não-negatividade,

## INTRODUÇÃO

unimodalidade, “closure”, seletividade, trilinearidade e restrições baseadas em modelos pré-definidos (“hard modeling”). A Figura 9 demonstra o efeito da aplicação de algumas restrições na matriz de concentração  $C$ .



**Figura 9.** Efeitos de algumas restrições aplicadas na correção dos perfis de concentração. Adaptado da referência 60.

Na restrição de não-negatividade, somente valores positivos para as concentrações têm significado físico. Assim, se um dado perfil apresenta elementos com valores negativos, estes devem ser forçados a serem iguais a zero ou serem maiores que um valor negativo muito pequeno. A restrição de unimodalidade faz com que os perfis tenham um único máximo. A restrição de “closure” força a soma dos perfis de concentração ser constante durante todo o processo. Já a restrição de “hard modeling” impõe que os perfis sigam modelos pré-definidos segundo as leis da físico-química. Observe ainda que estas restrições podem ser aplicadas apenas a alguns componentes.

## INTRODUÇÃO

---

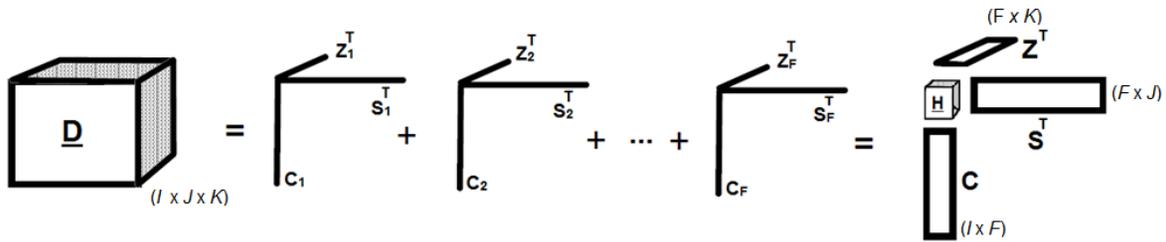
Outras restrições podem ser aplicadas quando diversas matrizes com pelo menos uma ordem em comum são analisadas simultaneamente. Estas restrições são: i) a correspondência entre os componentes em diferentes matrizes e ii) a trilinearidade.

A primeira permite fixar a sequência das espécies e selecionar quais delas estão presentes ou ausentes em cada submatriz. Esta presença/ausência é codificada em formato binário (zeros e uns), e introduzida no algoritmo MCR-ALS através de uma matriz com o número de linhas igual ao número de submatrizes e o número de colunas igual ao número total de componentes do sistema. A consequência imediata desta restrição é que quando um componente particular não existe em uma dada submatriz, seus elementos no perfil relacionado são igualados a zero, e isto contribui significativamente para eliminar a ambiguidade rotacional, permitindo em muitas circunstâncias alcançar condições de resolução única<sup>31</sup>.

Em geral, esta restrição é bastante útil quando a matriz de uma mistura complexa é aumentada com outra contendo somente a informação referente a um dos componentes da mistura. Como as condições de resolução para este analito em sua matriz pura são facilmente conseguidas sem ambiguidade, ela é estendida para a outra matriz analisada simultaneamente, possibilitando quantificar o referido analito mesmo na presença de interferentes desconhecidos. Embora este aspecto seja estritamente válido para sistemas trilineares, ele pode ser facilmente satisfeito para sistemas bilineares quando condições de posto local são obtidas<sup>31</sup>.

A segunda restrição é a de trilinearidade. Ela pode ser aplicada aos dados com duas ordens em comum (Figura 6c) nos quais perfis de concentração e espectrais idênticos são encontrados em cada submatriz. A expressão para um modelo trilinear já foi apresentada na equação 13 e a decomposição trilinear de um arranjo “three-way”  $\underline{\mathbf{D}}$  é representada graficamente na Figura 10.

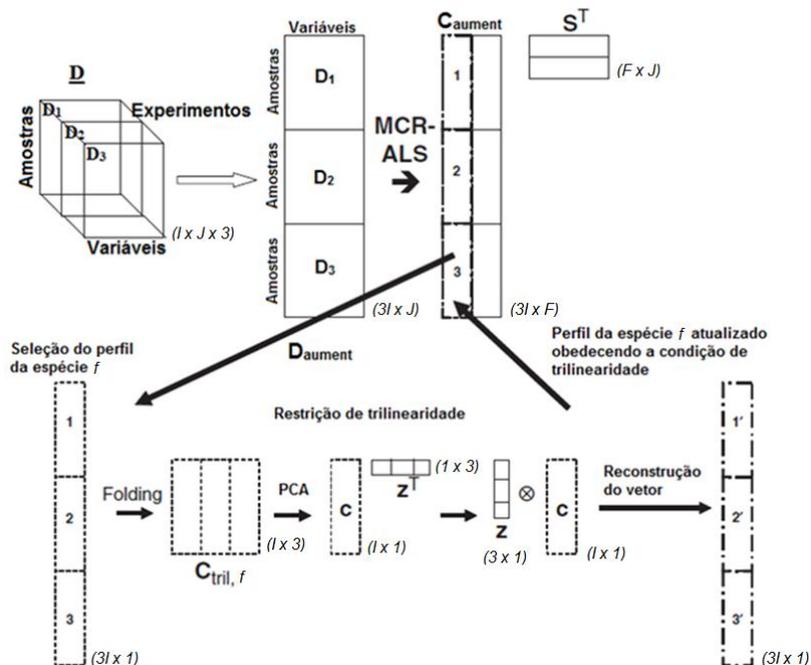
# INTRODUÇÃO



**Figura 10.** Decomposição de um conjunto de dados “three-way” de acordo com um modelo trilinear expresso através de matrizes  $C$ ,  $Z$  e  $S$ .  $H$  é um arranjo central binário de dimensões  $(F \times F \times F)$ .

Essa decomposição produz um conjunto de tríades referentes a cada componente que são armazenadas nas matrizes,  $C$ ,  $Z$  e  $S$ , conhecidas como matrizes de pesos do primeiro, segundo e terceiro modo, respectivamente.

Como descrito na equação 13,  $Z_k$  será a única matriz no modelo trilinear que muda na decomposição para diferentes matrizes  $D_k$ . O fato das matrizes  $C$  e  $S$  terem os mesmos perfis em todas as matrizes  $D_k$  é uma restrição muito forte que leva a decomposição trilinear ter uma solução única<sup>8,32</sup>. A Figura 11 mostra o procedimento executado ao se selecionar esta restrição.



**Figura 11.** Implementação da restrição de trilinearidade no método MCR-ALS. Nesta representação  $k=3$  e  $f = 1,2,\dots,F$ .

## INTRODUÇÃO

---

Os perfis de concentração do mesmo componente  $f$  em diferentes matrizes  $C_k$  são colocados um ao lado do outro para formar uma matriz do perfil de concentração dessa espécie, chamada  $C_{tril,f}$  com dimensões  $(I \times K)$ , onde  $I$  corresponde ao número de linhas das submatrizes e  $K$  o número de submatrizes. Essa última é então submetida a uma análise de PCA e somente a primeira componente será utilizada (equação 20).

$$C_{tril,f}(I, k) = \mathbf{c}(I, 1)\mathbf{z}^T(1, k) \quad \text{Eq.20}$$

onde  $\mathbf{c}$  é um vetor coluna que contém o perfil de concentração comum do componente  $f$  nas diferentes matrizes, e  $\mathbf{z}$  é um vetor linha com as quantidades relativas deste perfil de concentração nas diferentes matrizes.

Depois desta etapa, o vetor de concentração aumentado é reconstruído pelo produto de Kronecker entre os vetores  $\mathbf{z}$  e  $\mathbf{c}$ , e o vetor resultante que obedece à condição de trilinearidade é então utilizado para substituir o vetor original.

Vale ressaltar que os vetores  $\mathbf{c}$ ,  $\mathbf{s}$  e  $\mathbf{z}$  de cada fator estimam os mesmos perfis obtidos nas matrizes de peso do primeiro, segundo e terceiro modo,  $\mathbf{C}$ ,  $\mathbf{S}$  e  $\mathbf{Z}$ , quando um modelo de decomposição trilinear é utilizado. Portanto, os resultados obtidos com o método MCR-ALS, com restrição de trilinearidade aplicada a todos os fatores, produzirão resultados similares àqueles baseados em um modelo trilinear, tal como o método PARAFAC.

Este procedimento pode ser realizado para cada espécie  $f$  separadamente durante a otimização ALS, sendo esta a principal vantagem do uso do método MCR-ALS sobre os de resolução “three-way”, tais como PARAFAC, TUCKER3, já que ele permite a flexibilidade para abranger situações entre um modelo puramente bilinear e um completamente trilinear.

As figuras de mérito do procedimento de otimização são a percentagem de falta de ajuste, a percentagem de variância explicada e o desvio padrão dos resíduos com relação aos dados experimentais. A falta de ajuste é definida como a diferença entre a matriz de dados experimentais e os dados reproduzidos pelo

## INTRODUÇÃO

---

produto  $\mathbf{C S}^T$  em que  $\mathbf{C}$  e  $\mathbf{S}$  foram obtidos pelo método MCR-ALS. Este valor é calculado de acordo com a expressão:

$$\text{Falta de ajuste (\%)} = 100 * \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \quad \text{Eq.21}$$

onde  $d_{ij}$  representa o elemento na posição  $ij$  da matriz de dados experimentais  $\mathbf{D}$ , e  $e_{ij}$  corresponde ao respectivo elemento da matriz residual, obtido pela diferença entre os dados experimentais e os dados estimados pelo modelo. Dois valores diferentes de falta de ajuste são calculados. Eles diferem na definição da matriz de dados, ou seja, se os elementos  $d_{ij}$  são obtidos diretamente da matriz  $\mathbf{D}$  ou da matriz  $\hat{\mathbf{D}}_{\text{PCA}}$ .

A percentagem de variância explicada ( $R^2$ ) e o desvio padrão dos resíduos com relação aos dados experimentais ( $\sigma$ ) são calculados de acordo com as equações 22 e 23, respectivamente.

$$R^2(\%) = \left(1 - \frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}\right) * 100 \quad \text{Eq.22}$$

$$\sigma = \sqrt{\frac{\sum_{i,j} e_{ij}^2}{I * J}} \quad \text{Eq.23}$$

O método MCR-ALS se tornou um dos métodos mais populares para fazer a resolução de curvas devido à sua habilidade em tratar múltiplas matrizes simultaneamente, o que como já foi comentado, permite reduzir a ambiguidade rotacional intrínseca aos métodos bilineares e eliminar os problemas de deficiência de posto. Além disso, ele apresenta diversidade e flexibilidade de restrições que ajudam a melhorar os resultados da resolução. Por fim, o mesmo possui uma interface gráfica<sup>53</sup> bastante amigável que pode ser fácil e gratuitamente acessada em plataforma MATLAB.

## INTRODUÇÃO

---

O código MCR-ALS, tutoriais relacionados e conjuntos de dados para prática estão disponíveis em: [http://www.ub.edu/mcr/web\\_mcr/download.html](http://www.ub.edu/mcr/web_mcr/download.html). Um breve tutorial do programa MCR-ALS em português é apresentado nos anexos desta Tese.

### PARAFAC

PARAFAC (“Parallel Factor Analysis”) é um método de decomposição multilinear aplicado a arranjos de dados “multi-way”, cuja ideia central consiste em descrever as variações existentes nos dados, através do uso dos mesmos fatores latentes contidos nas diferentes matrizes em diferentes proporções<sup>34,44,45,61</sup>. Essa é exatamente a ideia por trás do princípio dos perfis proporcionais paralelos de Cattell<sup>33</sup>, que tem aplicação direta nos problemas de resolução de curvas e de calibração de segunda ordem.

No caso de uma análise “three-way”, a decomposição dos dados é feita em tríades ou componentes trilineares, conforme representado na Figura 10, e o modelo PARAFAC pode ser escrito através do produto de Khatri-Rao,  $\odot$ , segundo a equação 24.

$$\mathbf{D}(J \times IK) = \mathbf{C}(\mathbf{Z} \odot \mathbf{S})^T + \mathbf{E}(J \times IK) \quad \text{Eq.24}$$

$\mathbf{D}(J \times IK)$  é a matriz  $\mathbf{D}_{\text{augment}}$  representada na Figura 11.  $\mathbf{C}$ ,  $\mathbf{Z}$  e  $\mathbf{S}$  são as matrizes de pesos ilustradas na Figura 10.

Os parâmetros em  $\mathbf{C}$ ,  $\mathbf{Z}$  e  $\mathbf{S}$  podem ser estimados com diferentes algoritmos. Os mais utilizados são aqueles baseados no algoritmo ALS, devido à facilidade de implementação, simplicidade em incorporar restrições e garantia de convergência, embora esta algumas vezes seja demorada<sup>62</sup>.

A função de ajuste por quadrados mínimos utilizada é representada pela equação 25, e uma maneira de estimar  $\mathbf{C}$  condicionalmente a  $\mathbf{Z}$  e  $\mathbf{S}$  é apresentada na equação 26, onde  $\mathbf{W} = (\mathbf{Z} \odot \mathbf{S})$ .

$$\min_{\mathbf{C}, \mathbf{Z}, \mathbf{S}} \|\mathbf{D}(J \times IK) - \mathbf{C}(\mathbf{Z} \odot \mathbf{S})^T\| \quad \text{Eq.25}$$

## INTRODUÇÃO

---

$$\|\mathbf{D}(J \times IK) - \mathbf{C}\mathbf{W}^T\| \quad \text{Eq.26}$$

Dessa forma, uma estimativa por quadrados mínimos de  $\mathbf{C}$  pode ser encontrada fornecendo uma estimativa inicial para as matrizes  $\mathbf{Z}$  e  $\mathbf{S}$ , e conseqüentemente, para  $\mathbf{W}$ , de acordo com a equação 27.

$$\mathbf{C} = \mathbf{D}(J \times IK)(\mathbf{W}^T)^+ = \mathbf{D}(J \times IK)\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1} \quad \text{Eq.27}$$

Pela simetria do problema, as matrizes  $\mathbf{Z}$  e  $\mathbf{S}$  também podem ser estimadas de maneira similar através das equações 28 e 29.

$$\mathbf{W} = (\mathbf{C} \odot \mathbf{S}) \text{ e } \mathbf{Z} = \mathbf{D}(K \times IJ)\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1} \quad \text{Eq.28}$$

$$\mathbf{W} = (\mathbf{C} \odot \mathbf{Z}) \text{ e } \mathbf{S} = \mathbf{D}(I \times JK)\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1} \quad \text{Eq.29}$$

Este procedimento condicionado é repetido até que o critério de convergência (a menor mudança relativa no ajuste do modelo entre duas iterações consecutivas) seja alcançado. Em geral, o valor do critério de convergência é  $10^{-6}$ .

A etapa mais importante quando se emprega o método PARAFAC consiste em determinar o número apropriado de fatores ( $F$ ) no conjunto de dados. Este parâmetro pode ser determinado baseado em diversos critérios, como por exemplo, inspeção visual, variância explicada, diagnóstico de consistência de núcleo (CORCONDIA), dentre outros.

A variância explicada é calculada de acordo com a equação 30

$$\text{Variância Explicada} = \left(1 - \frac{\sum_{i,j,k} e_{ijk}^2}{\sum_{i,j,k} d_{ijk}^2}\right) * 100 \quad \text{Eq.30}$$

Para o cálculo do CORCONDIA<sup>63</sup>, o modelo PARAFAC deve ser assumido como um modelo TUCKER3 restrito representado pela equação 31

$$\mathbf{D}(J \times IK) = \mathbf{C}\mathbf{H}(F \times FF)(\mathbf{Z} \odot \mathbf{S})^T + \mathbf{E}(J \times IK) \quad \text{Eq.31}$$

## INTRODUÇÃO

onde  $\mathbf{H}(F \times F \times F)$  é o arranjo central binário ( $\mathbf{H}$ ), matricizado, que possui o valor 1 para todos os elementos da superdiagonal e o valor 0 para os demais (vide Figura 10).

Depois de ter ajustado o modelo PARAFAC, ou seja, calculado as matrizes  $\mathbf{C}$ ,  $\mathbf{Z}$  e  $\mathbf{S}$ , a verificação de que a estrutura trilinear é apropriada pode ser obtida calculando-se o arranjo central do modelo TUCKER3,  $\mathbf{G}$ , via quadrados mínimos, a partir de  $\mathbf{C}$ ,  $\mathbf{Z}$  e  $\mathbf{S}$ , isto é:

$$\min_{\mathbf{G}} \|\mathbf{D}(J \times I \times K) - \mathbf{CG}(F \times F \times F)(\mathbf{Z} \odot \mathbf{S})^T\| \quad \text{Eq.32}$$

Originalmente,  $\mathbf{G}$  pode ter diferentes dimensões em cada modo, ou seja,  $\mathbf{G}(D \times E \times F)$ . Os elementos desse arranjo definem como os fatores individuais nos diferentes modos interagem. Entretanto, como as matrizes  $\mathbf{C}$ ,  $\mathbf{Z}$  e  $\mathbf{S}$  obtidas por PARAFAC possuem o mesmo número de fatores ( $F$ ), o arranjo  $\mathbf{G}$  obtido na equação 32 passa a ter dimensões  $(F \times F \times F)$ .

Dessa forma, o monitoramento da distribuição dos elementos na superdiagonal de  $\mathbf{G}$  e fora dela, consiste em uma maneira simples de se verificar se a estrutura do modelo trilinear é adequada. Assim, se os elementos na superdiagonal são todos próximos de um, e os elementos fora da superdiagonal próximos de zero, então  $\mathbf{G}$  deve se parecer com  $\mathbf{H}$  e, portanto, o modelo trilinear é adequado<sup>63</sup>.

O valor do CORCONDIA pode então ser calculado a partir da equação 33.

$$\text{CORCONDIA} = \left( 1 - \frac{\sum_{i,j,k}^F (g_{ijk} - h_{ijk})^2}{\sum_{i,j,k}^F h_{ijk}^2} \right) * 100 \quad \text{Eq33}$$

Se o modelo PARAFAC é válido, então  $\mathbf{G}$  é semelhante a  $\mathbf{H}$ , e o valor de CORCONDIA é próximo a 100%. Se um dado não pode ser descrito aproximadamente por um modelo trilinear ou se muitos componentes são utilizados, então  $\mathbf{G}$  diferirá de  $\mathbf{H}$  e seu valor de CORCONDIA será próxima de zero

## INTRODUÇÃO

---

ou até mesmo negativo. Se este valor for próximo de 50%, o modelo é instável e a imposição de restrições pode auxiliar na sua estabilização<sup>63</sup>.

O algoritmo PARAFAC aqui utilizado pode ser obtido gratuitamente no site <http://www.models.life.ku.dk/source/nwaytoolbox/>.

Embora o método PARAFAC seja bastante utilizado em quimiometria devido: i) à sua semelhança estrutural com muitos modelos físicos de dados instrumentais, como os obtidos em cromatografia e fluorescência; ii) à sua propriedade de unicidade; o seu uso não é tão simples. Este envolve a execução de algumas etapas. A primeira consiste em obter o arranjo multidimensional. No caso da análise de dados “three-way”, a seguinte rotina pode ser utilizada.

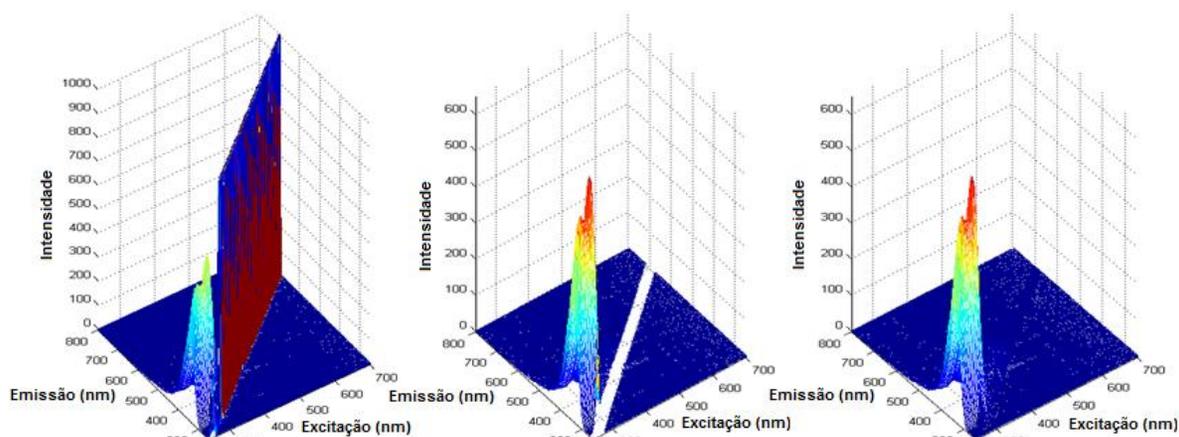
```
function cubo = montacubo(X,nmatrizes);  
% Nessa rotina o usuário entra com uma matriz (X) aumentada na direção  
das colunas, por exemplo, X=[X1;X2;...;Xk] e com o número de matrizes  
disponíveis (nmatrizes). A saída será o dado na forma de um arranjo  
three-way, aqui denominado cubo.  
[i,j] = size(X);  
l = i/nmatrizes;  
for k=1:nmatrizes  
    cubo(:, :, k) = X((k-1)*l+1:k*l, :);  
end
```

Essa rotina montará um arranjo “three-way” de dados. Depois do arranjo montado, o analista deve avaliar se o conjunto de dados necessita de algum pré-tratamento. Por exemplo, dados de fluorescência são geralmente afetados por efeitos de espalhamentos Rayleigh e Raman<sup>64</sup>. Uma vez que o método PARAFAC decompõe o sinal de fluorescência em uma série de estruturas trilineares e os picos de espalhamento não se comportam segundo essa premissa, isto é, a sua forma e posição mudam com o comprimento de onda de excitação, o sinal espalhado causa algumas dificuldades na decomposição<sup>64</sup>.

Existem diversas maneiras de contornar esse problema<sup>64</sup>. Uma delas consiste em substituir estas regiões por valores ausentes. Quando esse tratamento é aplicado, a intensidade da radiação é substituída pelo valor “NaN” que é o padrão adotado pelo software MATLAB. Pode-se também utilizar o

## INTRODUÇÃO

método de ajuste ponderado do PARAFAC, que adiciona pesos pequenos para a região da matriz que contém dados não-trilineares e pesos maiores para a região trilinear. Uma metodologia mais recente apresentada por Bahram et al.<sup>65</sup>, consiste em excluir a região de espalhamento através de uma janela com tamanho definido pelo usuário e completar essa área excluída com valores obtidos por interpolação. A aplicação dessa metodologia em um conjunto de dados de fluorescência é apresentada na Figura 12.



**Figura 12.** Remoção de picos de espalhamento de matrizes de emissão-excitação.

Este tratamento é feito utilizando o algoritmo EEMSCAT que também pode ser obtido gratuitamente no site [http://www.models.kvl.dk/EEM\\_correction](http://www.models.kvl.dk/EEM_correction).

Para ajustar um modelo PARAFAC ao arranjo “three-way” gerado e tratado previamente, basta escrever a linha de comando abaixo na janela de comando do MATLAB, indicando o arranjo (cubo), o número de fatores (Fac), alguns opcionais como o método de inicialização, critério de convergência, entre outros (Options), e algumas restrições (Const). Para mais informações digite “help parafac” na janela de comando.

```
[model,it,err,corcondia] = parafac(cubo,Fac,Options,Const);
```

Como resultado, os parâmetros do modelo **C**, **Z** e **S**, o número de iterações, o erro e o valor do CORCONDIA são obtidos.

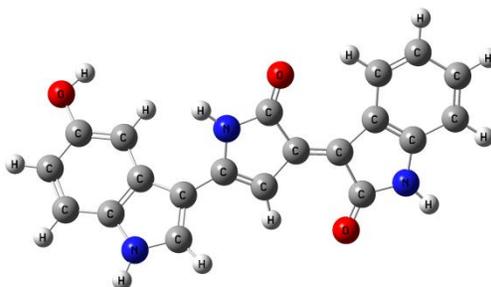
## INTRODUÇÃO

Recentemente, Bro e Vidal<sup>66</sup> divulgaram um novo algoritmo chamado EEMizer no qual todas as etapas da modelagem de dados de fluorescência por PARAFAC foram automatizadas neste único pacote. Ele pode ser obtido no site <http://www.models.life.ku.dk/EEMizer>.

### A violaceína e sua biossíntese

A violaceína é um derivado indólico sintetizado por diversas linhagens de bactérias Gram-negativas, principalmente do gênero *Chromobacterium*, que tem se destacado devido ao seu amplo espectro de atividade antibactericida<sup>67-69</sup>, além de outras propriedades tais como tripanocida<sup>67,68,70</sup>, tumoricida<sup>70-73</sup> e antiviral<sup>70,74</sup>.

Este composto (Figura 13) é um metabólito secundário com massa molecular 343,3 u, constituído pelos grupos 5-hidroxiindol, 2-pirrolidona e 2-oxindol e que apresenta intensa banda de absorção na região do visível devido à estrutura eletrônica altamente ressonante, o que confere coloração violeta às suas colônias.



**Figura 13.** Estrutura química da violaceína.

Estudos sobre sua biossíntese iniciaram-se em 1934 com Tobie<sup>75</sup>, quando se observou que a oxigenação de uma cultura da *Chromobacterium violaceum* reduzia significativamente o tempo necessário para a produção da violaceína. Vinte e cinco anos mais tarde, DeMoss e Evans<sup>76,77</sup> descobriram que para sintetizar violaceína, a bactéria necessitava de oxigênio molecular e L-triptofano. De 1987 a 1990, Hoshino e colaboradores demonstraram que i) o esqueleto carbônico do grupo pirrolidona era formado pela condensação da cadeia lateral de duas moléculas de L-triptofano<sup>78</sup>; ii) todos os átomos de carbono, nitrogênio e

## INTRODUÇÃO

---

hidrogênio do grupo pirrolidona eram exclusivamente provenientes do L-triptofano e que os átomos de oxigênio eram oriundos do oxigênio molecular<sup>79</sup>; iii) o 5-hidroxi-L-triptofano era um provável intermediário da biossíntese<sup>80</sup>.

Até então, a biossíntese da violaceína tinha sido estudada somente através de experimentos de incorporação de radioisótopos. Entretanto, em 1991, Pemberton *et al.*<sup>81</sup> iniciaram seu estudo em nível genético. Eles isolaram um fragmento de 14.5 kb da *C. violaceum* que continha os genes que codificavam a biossíntese da violaceína.

De 1993 a 2000, em uma série de artigos, Hoshino e colaboradores descobriram i) que o ácido cromopirrólico (ACP) foi produzido independentemente da biossíntese da violaceína<sup>82</sup>; ii) proviolaceína, prodeoxiviolaceína e pseudoviolaceína foram identificados como possíveis intermediários desse bioprocessos, e que a oxigenação na posição 2 do anel indólico ocorreu na etapa final da biossíntese<sup>83</sup>; iii) o envolvimento de oxigenases e a necessidade do cofator NADPH<sup>84</sup>; iv) um rearranjo 1,2 do anel indólico através de um processo intramolecular durante a formação da violaceína<sup>85</sup>.

Ainda na década de 90, McClean *et al.*<sup>86</sup> reportaram que a biossíntese desse pigmento estava vinculada a um mecanismo regulatório, dependente da densidade celular, chamado de “quorum sensing”. Neste mecanismo, a interação de uma molécula sinalizadora, geralmente n-acil homoserinas lactonas, com um ativador transcricional de proteína, daria início à expressão de um fenótipo específico, dependente da densidade populacional.

Em 2000, August *et al.*<sup>87</sup> mostraram que os genes necessários à síntese de violaceína estavam organizados em um “*operon*” conhecido como VioABCD. Reunindo todas as informações existentes até então, esses autores propuseram uma rota hipotética para essa biossíntese que consistia de: i) modificação de um L-triptofano por VioD formando 5-hidroxi-L-triptofano e desaminação oxidativa do outro por VioA produzindo ácido 3-indol pirúvico (AIP); ii) condensação descarboxilativa dessas duas espécies derivadas do L-triptofano por VioB,

## INTRODUÇÃO

---

resultando em prodeoxiviolaceína; iii) oxigenação desse último por VioC produzindo violaceína.

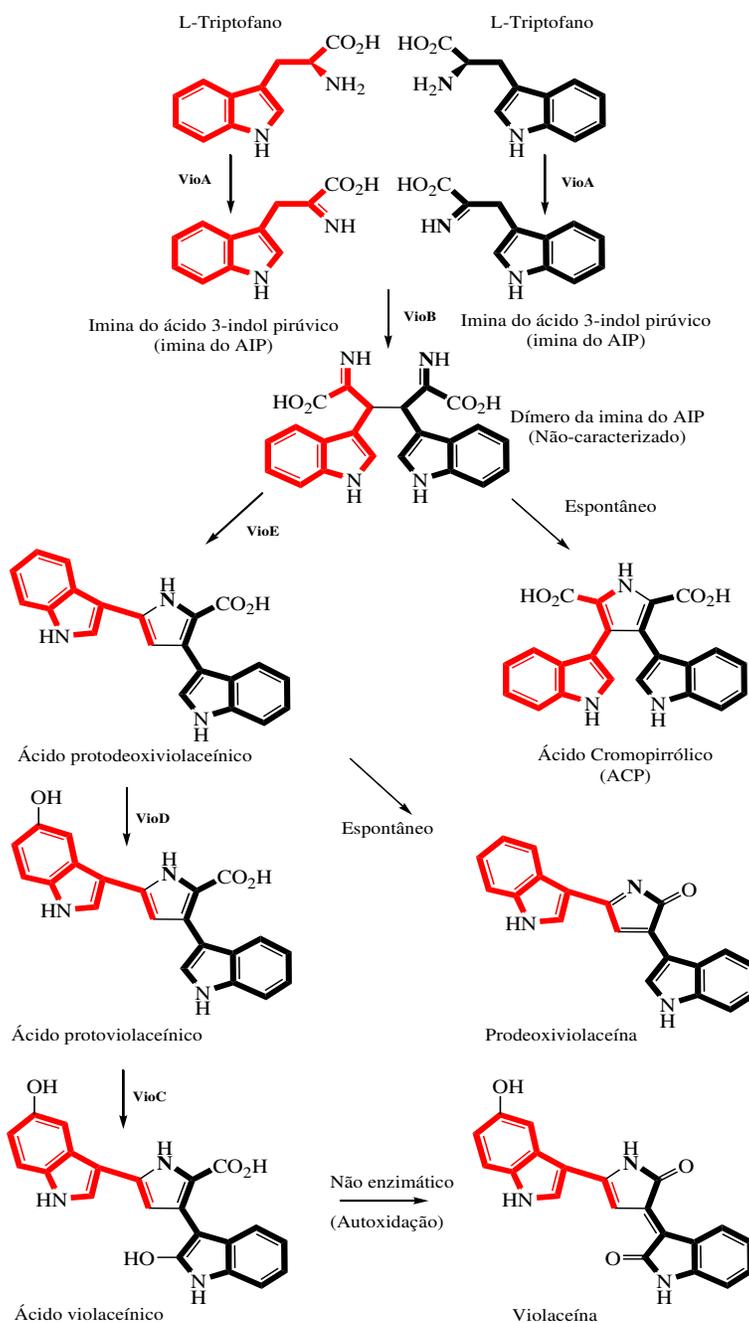
Em 2005, Howard-Jones e Walsh<sup>88</sup> observaram a similaridade entre as biossínteses da violaceína e de indolocarbazóis. Em ambos, a fusão descarboxilativa de espécies derivadas do L-triptofano foi observada. Além disso, os autores encontraram significantes níveis de identidade entre VioB e RebD (uma heme proteína necessária para a formação do ACP envolvida na biossíntese de indolocarbazóis).

Baseado nessa similaridade, Sánchez *et al.*<sup>89</sup>, em 2006, tentaram obter indolocarbazóis oxigenados através da construção de rotas híbridas, pela co-expressão de genes envolvidos nas duas biossínteses. Essa tentativa não produziu indolocarbazóis oxigenados, mas forneceu importantes informações para a biossíntese da violaceína; i) o par de genes VioAB foi responsável pelas etapas iniciais da biossíntese da violaceína e se mostrou equivalente ao par homólogo envolvido na biossíntese de indolocarbazóis (RebOD); ii) adicionalmente ao *operon* VioABCD, um quinto gene (VioE) foi apontado como essencial para a biossíntese da violaceína; iii) o ACP não foi um produto intermediário; iv) tanto VioC quanto VioD atuaram nas etapas finais da biossíntese da violaceína. Estes dois últimos resultados foram contraditórios aos reportados por Hoshino e colaboradores<sup>80,82,85</sup> e, conseqüentemente, ao mecanismo proposto por August *et al.*<sup>87</sup>.

Em 2006, Balibar e Walsh<sup>90</sup> reconstruíram toda a rota biossintética da violaceína através de experimentos *in vitro*. Seus resultados demonstraram que i) L-triptofano e não 5-hidroxi-L-triptofano foi o único precursor da violaceína; ii) VioB foi responsável pelo acoplamento oxidativo de duas moléculas de imina do AIP para a formação do centro pirrólico; iii) um quinto gene, VioE, foi responsável pelo rearranjo 1,2 do anel indólico resultando na formação de prodeoxiviolaceína ao invés de ACP; iv) a formação da violaceína ocorreu pela ação sequencial de VioD e VioC. Estes resultados corroboraram os resultados *in vivo* publicados por Sánchez *et al.*<sup>89</sup>.

## INTRODUÇÃO

Em 2007, Shinoda *et al.*<sup>91</sup> reportaram a identificação de um intermediário verdadeiro da biossíntese da violaceína, chamado de ácido protoviolaceínico. Este foi obtido pela incubação do L-triptofano com VioABDE na presença de NADPH. Adicionalmente, os referidos autores propuseram uma rota para a biotransformação do L-triptofano em violaceína (Figura 14).



**Figura 14.** Mecanismo de reação para a biossíntese da violaceína.

## INTRODUÇÃO

---

A complexidade dessa rota é notória. Ela se inicia com a oxidação do L-triptofano à imina do AIP pela ação da VioA. Então, o acoplamento oxidativo de duas moléculas de imina do AIP por VioB produz um intermediário não caracterizado, proposto para ser um dímero da imina do AIP. O esqueleto desse composto é cataliticamente arranjado pela VioE, através de um rearranjo intramolecular no anel indólico, produzindo ácido protodeoxiviolaceínico. Este último pode ser espontaneamente convertido, por auto-oxidação, em prodeoxiviolaceína ou pode sofrer uma hidroxilação catalisada pela VioD na posição 5 do anel indólico do lado esquerdo, formando ácido protoviolaceínico. O ácido protoviolaceínico é hidroxilado pela VioC na posição 2 do anel indólico do lado direito para produzir ácido violaceínico. A subsequente conversão a violaceína envolve um processo não enzimático de descarboxilação oxidativa.

A investigação mecanística desta complexa rota biossintética em nível químico, bioquímico e genético envolveu muitas etapas. Experimentos de marcação isotópica acoplados a técnicas de separação e elucidação estrutural, ferramentas moleculares e métodos recombinantes foram utilizados. Entretanto, uma propriedade especial deste sistema não foi explorada. Na biossíntese da violaceína, tanto o precursor (triptofano), o produto (violaceína) e, provavelmente, os intermediários são compostos fluorescentes. Portanto, informações sobre esta biossíntese podem ser obtidas quando a espectroscopia de fluorescência for utilizada para detectar compostos que são consumidos e produzidos durante o cultivo.

Nesse sentido, uma nova e complementar metodologia para obter informações sobre a rota biossintética de compostos naturais fluorescentes é proposta neste trabalho. Ela consiste no emprego da espectroscopia de fluorescência multicomprimento de onda e métodos de resolução de curvas para detectar e identificar (através de resolução espectral) fluoróforos que são consumidos e produzidos pela bactéria.

De fato, como previamente testado em nosso grupo<sup>92</sup>, a obtenção dessas informações mostra-se plausível, sendo que fluoróforos consumidos e produzidos

## INTRODUÇÃO

---

durante a biossíntese puderam ser detectados e suas variações ao longo do tempo avaliadas. Entretanto, esbarrou-se na dificuldade de identificá-los através da comparação de espectros de emissão e excitação resolvidos com aqueles disponíveis em base de dados.

Na tentativa de caracterizar melhor essas espécies e obter informações temporais mais confiáveis, também se investigou a possibilidade de analisar simultaneamente, através do método MCR-ALS, dados provenientes de diversas técnicas analíticas, a saber, fluorescência, absorção molecular no UV-VIS e cromatografia líquida de ultra eficiência com detecção por espectrometria de massas (UPLC-MS).

Os detalhes experimentais e os resultados discutidos serão apresentados a seguir.

### OBJETIVOS

O objetivo geral desta tese consiste em divulgar para comunidade científica brasileira a potencialidade dos métodos de resolução de curvas, MCR-ALS e PARAFAC, em extrair informações dos componentes puros sem necessidade de separação física.

Como objetivos específicos têm-se:

- Investigar a combinação da espectroscopia de fluorescência multicomprimento com os métodos PARAFAC e MCR-ALS como uma nova ferramenta para a análise da biossíntese de produtos naturais fluorescentes;
- Comparação dos resultados obtidos por estes dois métodos;
- Explorar a biossíntese da violaceína sob o ponto de vista da análise simultânea de dados experimentais gerados por diferentes técnicas usando o método MCR-ALS;
- Relacionar os resultados obtidos com o mecanismo outrora proposto na literatura.



### PARTE EXPERIMENTAL

#### **Organismo, meio de cultura e condições de cultivo**

Uma colônia da *C. violaceum* CCT 3496 foi inoculada em um Erlenmeyer de 250 mL contendo 50 mL de meio de cultura esterilizado (D-glicose 0,50%, peptona bacteriológica 0,50%, extrato de levedura 0,25% e L-triptofano 0,03%) e crescida por 24 horas a 33 °C em um “shaker” a 200 rpm. Em uma etapa posterior, 10 mL desta cultura de bactérias foram transferidos para um bioreator BioFlo de 1500 mL (New Brunswick Scientific) contendo 1000 mL de meio de cultura esterilizados e os parâmetros temperatura, agitação e fluxo de ar foram ajustados para 33 °C, 200 rpm e 1,0 L min<sup>-1</sup>, respectivamente, e mantidos constantes durante toda fermentação (aproximadamente 36 horas).

#### **Coleta das amostras**

A cada 2 horas, alíquotas de 10 mL eram retiradas do bioreator com uma seringa de vidro, armazenadas em tubos Falcon de 15 mL e estocadas a -20 °C. Em um estágio posterior, cada alíquota foi descongelada em banho termostatizado a 30 °C por 5 minutos, e então centrifugada a 7000 rpm por 10 minutos. O sobrenadante foi eliminado, 5 mL de etanol absoluto foram adicionados ao tubo e centrifugados a 7000 rpm por 10 minutos para romper as células bacterianas. Em seguida, coletou-se o sobrenadante que foi passado através de um filtro com tamanho de poro 0,45 µm. Ao todo, 18 alíquotas correspondentes ao extrato etanólico da *C. violaceum* em diferentes estágios da biossíntese foram obtidas.

#### **Espectroscopia de fluorescência**

Medidas de fluorescência foram realizadas usando um espectrofotômetro de fluorescência Varian modelo Cary Eclipse, no modo de varredura. Alíquotas de 0,5 mL dos extratos etanólicos foram diluídas em 2,5 mL de etanol absoluto na própria cubeta para reduzir efeitos de supressão, filtros internos e processos de transferência de energia. Espectros de luminescência total, também conhecidos como matrizes de emissão-excitação (MEE), foram coletados para cada amostra na região espectral de 250-620 nm para excitação, e 270-800 nm para emissão. A velocidade de varredura para cada MEE foi 1200 nm min<sup>-1</sup>. Os incrementos para

## PARTE EXPERIMENTAL

---

os comprimentos de onda de excitação e emissão foram de 5 e 2 nm, respectivamente. As larguras das fendas de excitação e emissão foram ajustadas para 5 nm. Os filtros foram definidos no modo “auto” para ambos os monocromadores. Isto significa que o filtro rotativo é automaticamente movido para uma posição apropriada de acordo com o comprimento de onda excitação/emissão selecionado. A voltagem da fotomultiplicadora foi 600 V. As medidas foram realizadas em geometria de 90°.

### **Espectroscopia de absorção molecular na região UV-VIS**

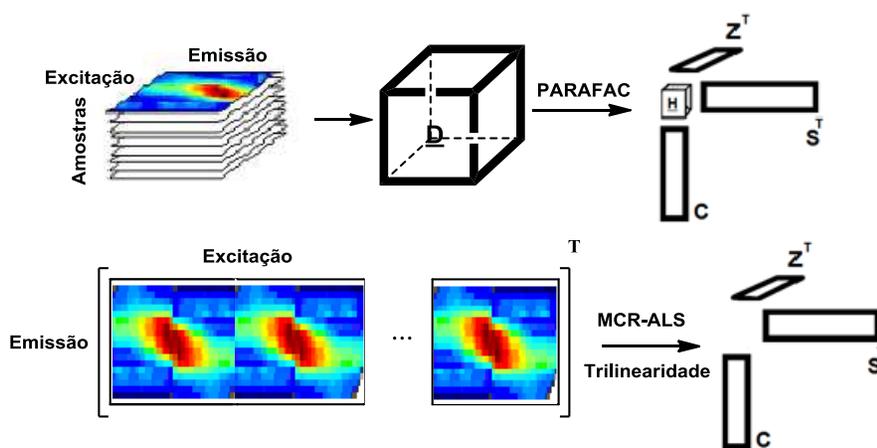
Os espectros de absorção na região ultravioleta-visível foram obtidos em um espectrofotômetro HP8453 com arranjo de diodos (Hewlett-Packard) na região espectral de 205-800 nm, com resolução espectral de 1 nm. Alíquotas de 0,5 mL dos extratos etanólicos foram diluídos em 2,5 mL de etanol absoluto na própria cubeta.

### **Cromatografia líquida de ultra eficiência-detector de massas (UPLC-MS)**

As amostras também foram analisadas empregando um sistema de cromatografia líquida de ultra eficiência da Waters Acquity<sup>®</sup>, acoplado a um espectrômetro de massas. A separação foi realizada em uma coluna Acquity<sup>®</sup> UPLC BEH C18 (2,1 x 50 mm), com diâmetro de partícula de 1,7 µm, empregando o seguinte gradiente de eluição: inicialmente com 95% (v/v) da fase A (água Milli-Q + ácido fórmico 0,1%) e 5% (v/v) da fase B (acetonitrila + ácido fórmico 0,1%) com rampa linear para 70% da fase B em 4 min e permanecendo assim por mais 4 min. No minuto seguinte, a composição foi restabelecida, perfazendo um tempo de análise de 9 min. O fluxo da fase móvel foi mantido a 350 µL min<sup>-1</sup> e o volume de injeção foi de 5 µL. A ionização por “electrospray” no modo positivo foi escolhida como método de ionização nas seguintes condições: voltagem do cone, 45 V; voltagem do capilar, 2,5 kV; temperatura da fonte, 150 °C; temperatura de dessolvatação, 400 °C. A faixa de aquisição foi de 100-1000 m/z.

### Modelagem dos dados

Inicialmente, as MEE foram empilhadas para formar um arranjo  $\underline{D}$  com três dimensões e, em seguida, o método PARAFAC foi aplicado a estes dados. O número de componentes foi avaliado com base em dois critérios: a variância explicada e o diagnóstico CORCONDIA. Posteriormente, a matriz aumentada,  $\underline{D}_{\text{aument}}$ , foi explorada por MCR-ALS com restrições de trilinearidade e não-negatividade, usando como estimativas iniciais os perfis de emissão obtidos pelo PARAFAC e perfis obtidos por EFA. A representação gráfica dessas decomposições é apresentada na Figura 15.



**Figura 15.** Representação gráfica da decomposição de  $\underline{D}$  por PARAFAC e de  $\underline{D}_{\text{aument}}$  por MCR-ALS.

Os parâmetros estimados pelos dois métodos foram comparados baseados no cálculo do ângulo entre dois vetores ( $\alpha$ ), por exemplo,  $\mathbf{c}_{1,\text{parafac}}$  e  $\mathbf{c}_{1,\text{MCR-ALS}}$ , de acordo com a equação 34.

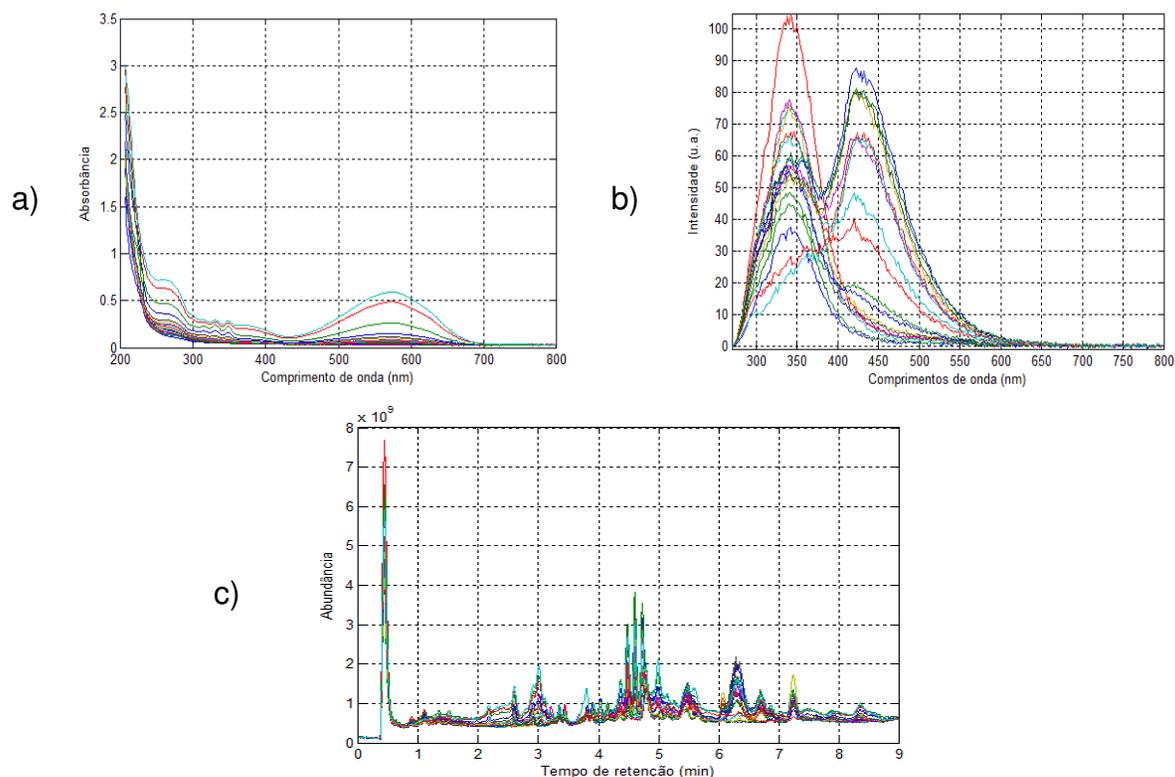
$$\alpha = \arccos\left(\frac{\mathbf{c}_{1,\text{parafac}}^T * \mathbf{c}_{1,\text{MCR-ALS}}}{\|\mathbf{c}_{1,\text{parafac}}\| * \|\mathbf{c}_{1,\text{MCR-ALS}}\|}\right) \quad \text{Eq34}$$

Uma vez que as mesmas amostras, extratos etanólicos da biossíntese da violaceína, foram monitoradas por diferentes técnicas analíticas, pensou-se em uma forma de arranjar estes dados e analisá-los simultaneamente. Como os dados de absorção molecular na região do UV-VIS são de primeira ordem,

## PARTE EXPERIMENTAL

decidiu-se transformar os dados de segunda ordem em dados de primeira, e assim explorar a biossíntese da violaceína conjuntamente. No caso dos dados de fluorescência, três matrizes, cada uma referente a um comprimento de onda de excitação específico (280, 310 e 350 nm), foram obtidas coletando-se os espectros de emissão em todos os tempos de reação. Para os dados de UPLC-MS, a conversão de uma matriz em um vetor pode se dar de duas formas: i) soma nas linhas (todos os tempos de retenção numa dada massa – espectro de massas totais); ii) soma nas colunas (todas as massas em um dado tempo de retenção – cromatograma de íons totais- TIC). Neste trabalho, a segunda forma foi utilizada, sendo que os TIC's foram arranjados em uma matriz  $D_{TIC}$  de acordo com a ordem evolutiva da reação.

Alguns desses dados são representados graficamente na Figura 16.

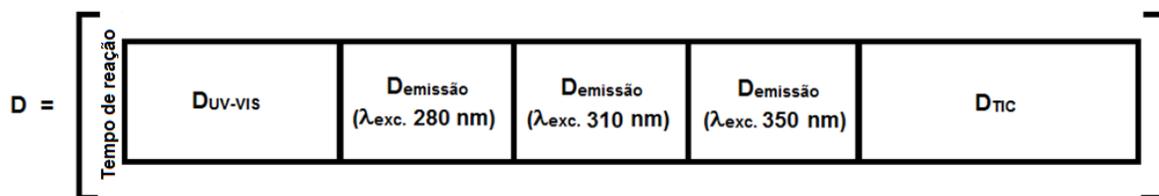


**Figura 16.** (a) Espectros de absorção molecular na região UV-VIS dos extratos etanólicos; (b) Espectros de emissão dos extratos etanólicos quando excitados em 280 nm; (c) Cromatogramas de íons totais (TIC's) dos extratos etanólicos.

## PARTE EXPERIMENTAL

Antes de concatenar estes dados, alguns pré-tratamentos foram aplicados. Inicialmente, devido às diferentes ordens de grandeza na escala das intensidades de cada uma dessas técnicas, conforme observado na Figura 16, os dados de fluorescência e UPLC-MS foram convertidos para a mesma escala de intensidade dos dados de absorção. Este procedimento foi conduzido dividindo essas matrizes por fatores de escalamento apropriados baseado em seu valor de máxima intensidade. A seguir, a linha de base dos cromatogramas foi corrigida utilizando o algoritmo AsLS<sup>93</sup> (“Asymmetric Least Squares”) e o sinal referente ao solvente (em torno de 0,5 min) foi excluído.

A vetorização das respostas instrumentais permitiu criar uma matriz **D** onde o modo das linhas, relacionado ao tempo de reação, foi conservado em cada técnica analítica (Figura 17). Isso equivale a monitorar a biossíntese da violaceína utilizando, simultaneamente, dados de absorção molecular, fluorescência e UPLC-MS.



**Figura 17.** Matriz de dados obtida para monitorar a biossíntese da violaceína.

Para extrair a informação dos componentes puros que explicam a variância nessa matriz de dados, o método de resolução de curvas multivariada por quadrados mínimos alternados (MCR-ALS) foi utilizado.

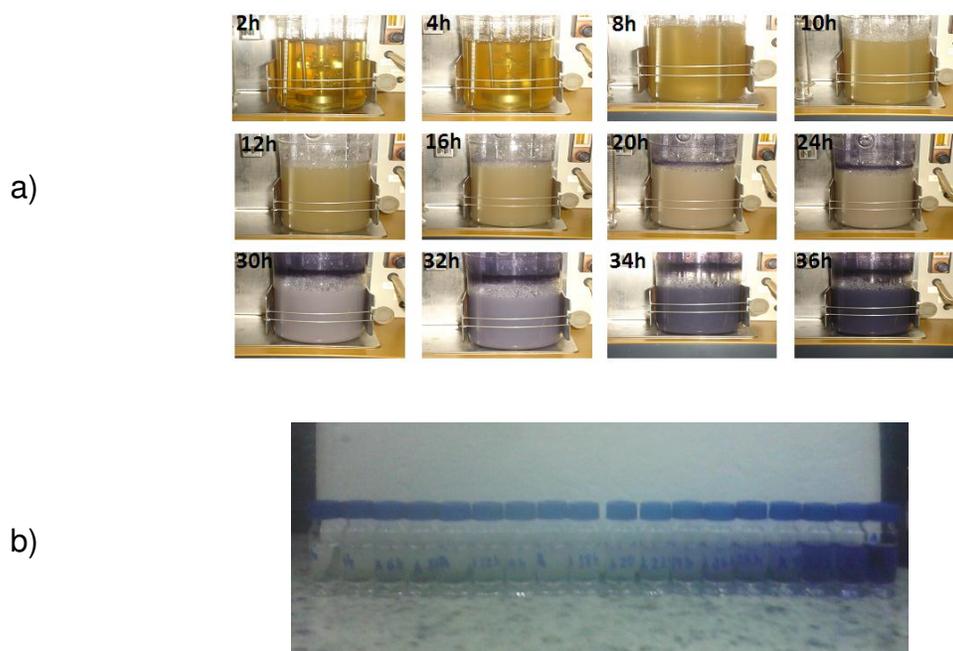


### RESULTADOS E DISCUSSÃO

A apresentação dos resultados será dividida em duas etapas. A primeira contemplará o monitoramento da biossíntese da violaceína através da espectroscopia de fluorescência e a resolução de curvas por PARAFAC e MCR-ALS. A segunda abordará os resultados referentes à fusão de todas as técnicas analíticas utilizadas para monitorar esta biossíntese, e a modelagem desses dados por MCR-ALS.

#### Fases de crescimento da bactéria

No intuito de mostrar as alterações visuais ocorridas no meio reacional durante a biossíntese da violaceína, uma série de imagens de amostras não-tratadas e tratadas foram registradas e são apresentadas na Figura 18.



**Figura 18.** Alterações visuais ocorridas no meio reacional durante a biossíntese da violaceína. a) amostras não-tratadas; b) amostras tratadas.

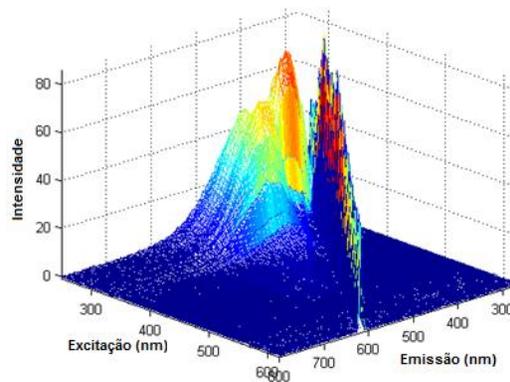
Fazendo um comparativo entre a Figura 18a e uma típica curva de crescimento bacteriano é possível observar que as duas primeiras horas após o inóculo correspondem à fase *lag*, período de adaptação da bactéria caracterizado por intensa atividade metabólica e ausência de divisão celular. A seguir, inicia-se a

## RESULTADOS E DISCUSSÃO

fase *log* ou *exponencial* caracterizada pela adaptação plena da bactéria ao meio e crescimento celular acelerado, a qual pode ser facilmente visualizada pela turvação do meio de cultura na Figura 18a. Por fim, percebe-se o surgimento da fase *estacionária* onde os nutrientes estão se tornando escassos e os produtos tóxicos provenientes do metabolismo mais abundantes. É na fase estacionária que são sintetizados vários metabólitos secundários e, nesse caso, caracterizada pela produção de violaceína<sup>94</sup>.

### Dados de Fluorescência

A espectroscopia de fluorescência multicomprimentos de onda envolve a aquisição sucessiva de espectros de emissão em diversos comprimentos de onda de excitação, criando MEEs que fornecem um mapa detalhado das propriedades fluorescentes da amostra. O gráfico da superfície tridimensional de uma MEE referente a um extrato etanólico da *C. violaceum* é mostrado na Figura 19.



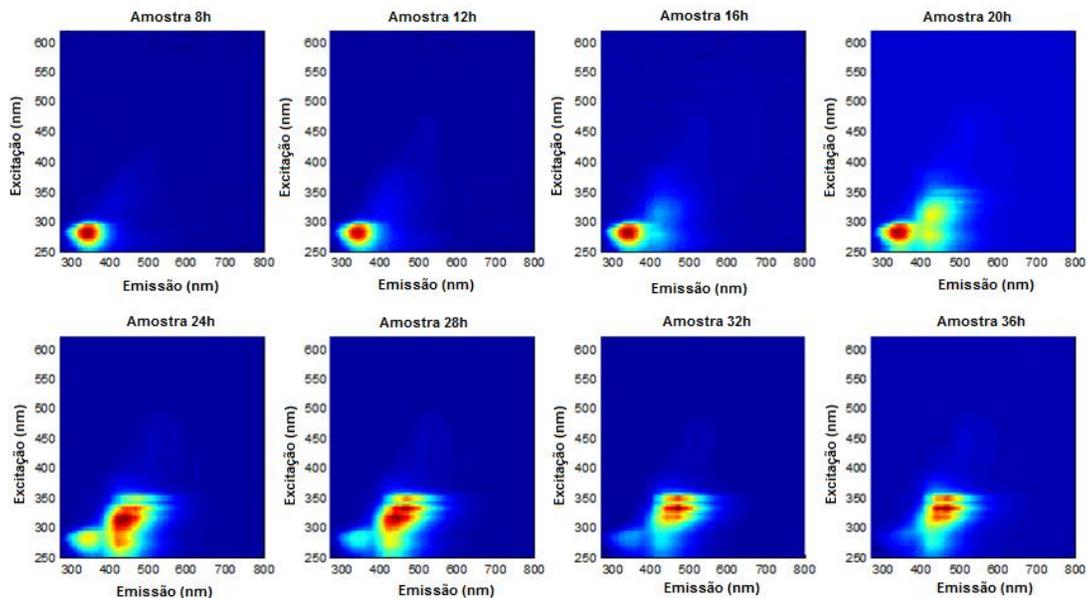
**Figura 19.** Típica MEE do extrato etanólico da *C. violaceum*.

É possível observar que existem três picos na região espectral de maior energia. Além disso, podem-se visualizar picos de espalhamento Rayleigh situados diagonalmente nessa MEE. A região de espalhamento não contém nenhuma informação das propriedades fluorescentes de uma amostra e não segue um comportamento trilinear (a forma e posição dos picos espalhados mudam com o comprimento de onda de excitação)<sup>64</sup>. Essa região deve, portanto, ser evitada na modelagem por PARAFAC, já que o sinal de fluorescência é decomposto em uma

## RESULTADOS E DISCUSSÃO

série de estruturas trilineares. O algoritmo EEMSCAT foi utilizado para essa finalidade<sup>65</sup>.

Ao excluir a região de espalhamento das MEE, é possível visualizar melhor o comportamento dinâmico dos fluoróforos envolvidos na biossíntese da violaceína, conforme apresentado na Figura 20.



**Figura 20.** Evolução temporal dos fluoróforos envolvidos na biossíntese da violaceína.

O primeiro fluoróforo apresenta máximo de emissão entre 330-360 nm quando excitado entre 270-290 nm. Este composto está sendo consumido durante a fermentação uma vez que as intensidades dessa banda se reduzem com o tempo de reação. O segundo fluoróforo mostra dois máximos de excitação (270-285 nm e 295-320 nm) e um máximo de emissão entre 400-450 nm. O terceiro fluoróforo apresenta três máximos de excitação (315-320 nm, 325-340 nm e 345-355 nm) e um máximo de emissão entre 440-480 nm. Essas múltiplas absorções podem ser compreendidas como excitações para diferentes níveis vibracionais do primeiro estado excitado de mesma multiplicidade do fluoróforo.

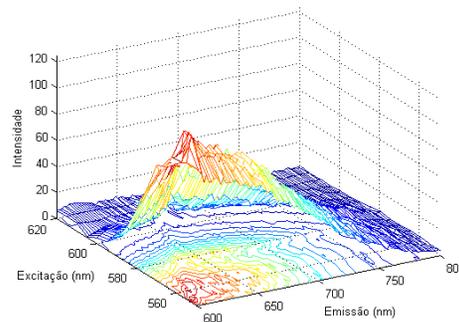
A molécula de violaceína possui máximo de emissão em 675 nm quando excitada em 575 nm<sup>95</sup>. Entretanto, como pode ser observado na Figura 20, não se

## RESULTADOS E DISCUSSÃO

visualiza nenhum sinal nessa região. A cor violeta do meio de cultura (Figura 18) e a espectroscopia de absorção no UV-VIS (técnica analítica muito menos sensível que a fluorescência) confirmam que este composto está sendo sintetizado.

Essa aparente contradição pode ser explicada: i) pela dependência da eficiência de alguns componentes do fluorímetro com a região espectral ii) por um baixo rendimento quântico da violaceína. Para explorar este problema, novas configurações instrumentais foram testadas. Os filtros em ambos os monocromadores foram ajustados para o modo aberto, o que significa que nenhum filtro foi utilizado, e a voltagem da fotomultiplicadora foi aumentada para 800 V. Os filtros de emissão e excitação são usados para reduzir os efeitos da radiação espalhada e o aumento na voltagem da fotomultiplicadora acentua o sinal de saída para uma dada quantidade de luz.

Com essa configuração foi possível detectar sinal de fluorescência na região de menor energia (550-620 nm para excitação e 600-800 nm para emissão), conforme pode ser visualizado na Figura 21.



**Figura 21.** Superfície 3D e gráfico de contorno de uma MEE na região F.

Baseado nesse resultado decidiu-se dividir a faixa espectral estudada em regiões menores e ajustar condições instrumentais específicas para cada uma delas que são apresentadas na Tabela 1.

## RESULTADOS E DISCUSSÃO

Tabela 1. Condições de medidas para cada região espectral.

Região	Excitação (nm)	Emissão (nm)	Filtro excitação	Filtro emissão	PMT* (V)	Fenda excitação (nm)	Fenda emissão (nm)
B	250-320	300-600	Auto	Auto	600	5	5
C	325-395	375-675	Auto	Auto	600	5	5
D	400-470	450-750	Aberto	Aberto	800	5	5
E	475-545	525-800	Aberto	Aberto	800	5	5
F	550-620	600-800	Aberto	Aberto	800	5	5

\*PMT = fotomultiplicadora

Para visualizar o comportamento dinâmico dos fluoróforos em toda a região de menor energia, novas MEEs foram criadas pela justaposição das MEEs pré-processadas D, E e F. O resultado dessa estratégia é apresentado na Figura 22.

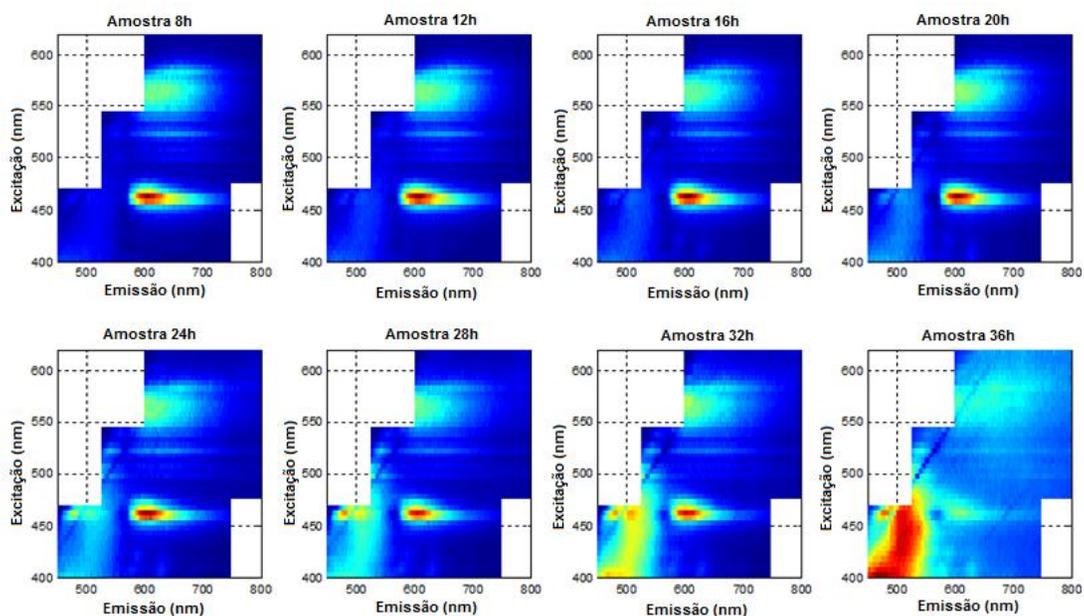


Figura 22. Evolução temporal dos fluoróforos envolvidos na biossíntese da violaceína detectados na região espectral de menor energia.

O fluoróforo com banda de emissão entre 590-620 nm parece apresentar três máximos de excitação (450-470 nm, 525-535 nm e 550-570 nm). Nesse caso, o pico de absorção de menor energia deve-se à excitação do fluoróforo para o

primeiro estado excitado de mesma multiplicidade. O pico na região de maior energia é devido uma excitação do mesmo fluoróforo para um segundo estado excitado de mesma multiplicidade. A única banda de emissão observada indica a ocorrência do processo de conversão interna, onde as moléculas excitadas do fluoróforo relaxam rapidamente para o nível vibracional de menor energia do primeiro estado excitado. Ainda nessa Figura, é possível observar que outro fluoróforo é produzido durante a fermentação, apresentando largas bandas de excitação (400-460 nm) e emissão (450-540 nm).

### **Modelagem dos dados de fluorescência utilizando o método PARAFAC**

A sobreposição das bandas de fluorescência impossibilita a identificação exata dos fluoróforos. Dessa forma, o método PARAFAC foi empregado para resolução de curvas. Esse método realiza uma deconvolução nos dados, proporcionando a separação dos espectros de emissão e excitação de cada espécie química presente nas amostras.

Como outrora mencionado, um ponto importante na modelagem por PARAFAC consiste na escolha do número de fatores. Definir o número de fatores significa determinar, nesse caso, o número de fluoróforos existentes no conjunto das amostras analisadas e, assim, estimar os espectros de emissão e excitação dessas espécies. Aqui, essa definição foi feita baseada na variância explicada e no teste de diagnóstico de consistência do núcleo (CORCONDIA).

Para a obtenção de perfis espectrais com significado físico, a restrição de não-negatividade foi aplicada nos três modos em todos os modelos.

Os resultados dos modelos ajustados para descrever os dados de fluorescência por PARAFAC para todas as regiões estudadas são apresentados na Tabela 2.

## RESULTADOS E DISCUSSÃO

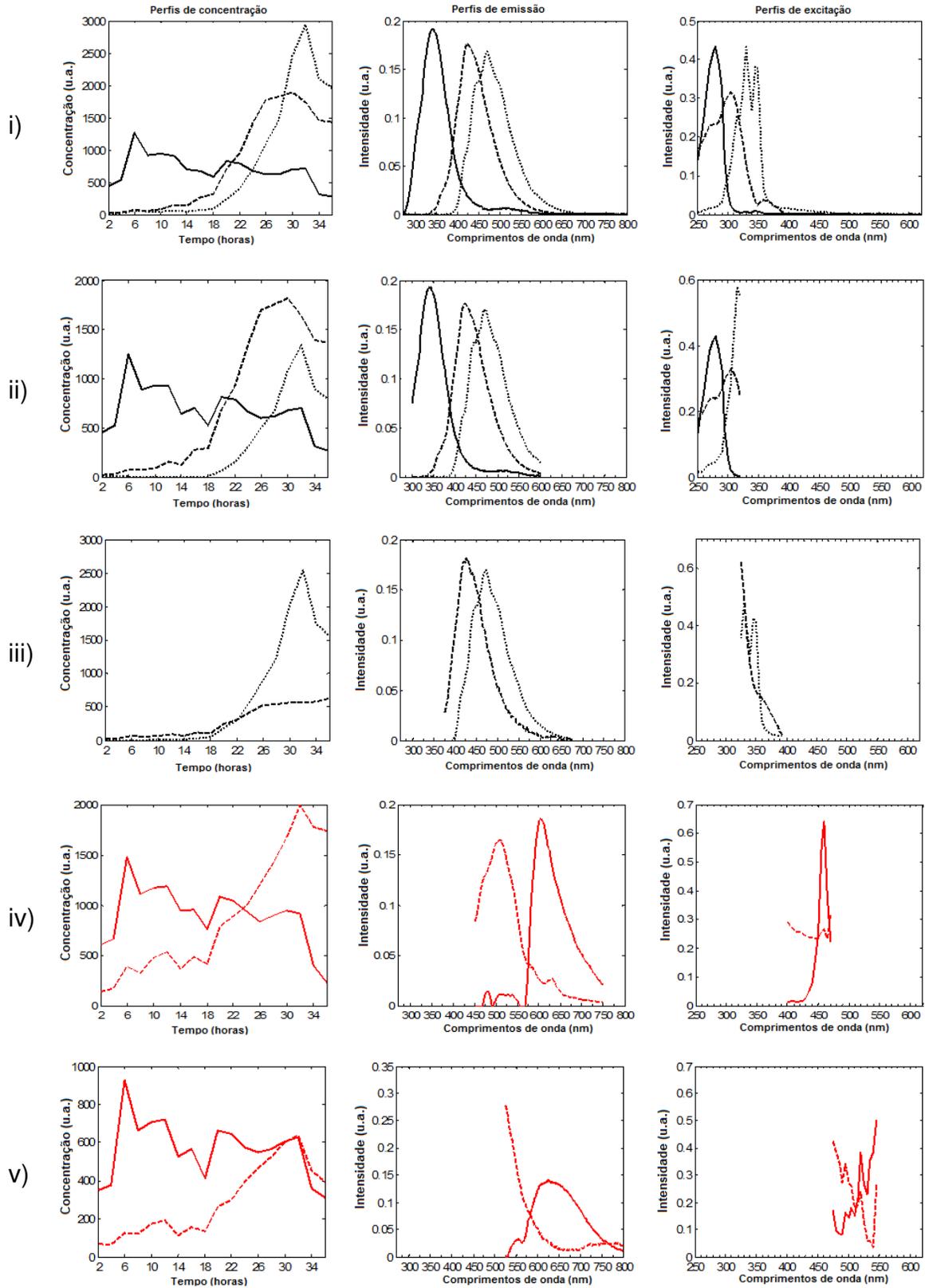
**Tabela 2.** Variância explicada e CORCONDIA calculados para os modelos trilineares de 1 a 6 fatores com o método PARAFAC.

Nº Fatores	1	2	3	4	5	6
<b>Toda a região espectral</b>						
Variância Explicada (%)	82,30	96,03	99,56	99,68	99,75	99,83
CORCONDIA (%)	100,0	99,94	99,58	70,64	11,54	41,30
<b>Região B</b>						
Variância Explicada (%)	80,85	98,35	99,70	99,83	99,87	99,93
CORCONDIA (%)	100	99,97	99,01	72,69	25,83	13,31
<b>Região C</b>						
Variância Explicada (%)	98,91	99,71	99,89	99,95	99,96	99,97
CORCONDIA (%)	100	99,93	22,65	11,25	-0,23	-0,05
<b>Região D</b>						
Variância Explicada (%)	75,03	96,51	98,16	98,83	99,21	99,63
CORCONDIA (%)	100	99,77	51,39	33,75	12,41	4,07
<b>Região E</b>						
Variância Explicada (%)	86,65	97,13	97,96	98,50	98,87	99,14
CORCONDIA (%)	100	99,98	35,39	42,35	7,64	6,14
<b>Região F</b>						
Variância Explicada (%)	96,18	98,60	99,44	99,63	99,67	99,71
CORCONDIA (%)	100	97,72	55,56	27,05	11,73	-1,21

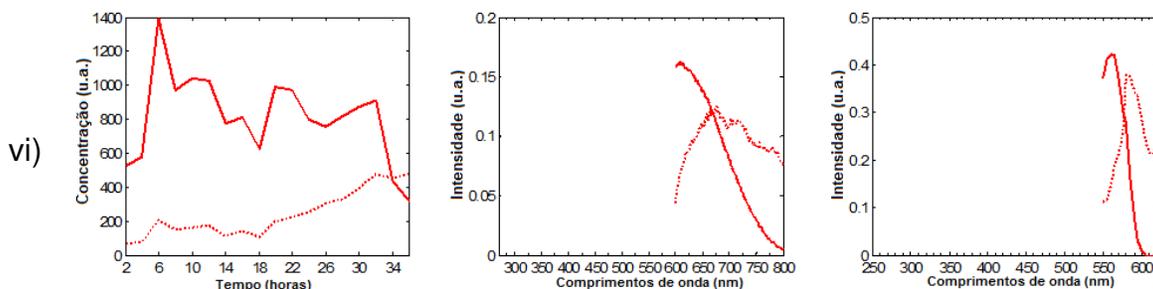
Analisando os resultados da Tabela 2 para toda a região espectral e para a região B, pode-se inferir que a variância explicada indica que um modelo com 3 fatores é adequado uma vez que o aumento obtido neste parâmetro com mais que 3 fatores é muito pequeno. O valor de CORCONDIA também sugere que o número de fatores deve ser 3. Para as regiões C, D, E e F, embora a variância explicada aponte para um número maior de fatores, o valor de CORCONDIA indica que 2 é o número ideal.

Os perfis de concentração, emissão e excitação obtidos por PARAFAC para cada região espectral são apresentados na Figura 23.

# RESULTADOS E DISCUSSÃO



## RESULTADOS E DISCUSSÃO



**Figura 23.** Estimativas dos perfis de concentração, emissão e excitação obtida por PARAFAC. i) toda a região espectral; ii) região B; iii) região C; iv) região D; v) região E; vi) região F.

Ao todo, seis espécies químicas provavelmente envolvidas na biossíntese da violaceína puderam ser detectadas com esta metodologia. Na Figura 23i, o primeiro perfil de excitação tem máximo em 280 nm e máximo de emissão em 342 nm. Os perfis espectrais resolvidos por PARAFAC são similares aos espectros de emissão e absorção do triptofano<sup>96</sup>. Este fato corrobora o mecanismo apresentado na Figura 14 no qual o triptofano é o precursor da biossíntese. O perfil de concentração deste composto apresenta tendência de diminuição gradual ao longo da biossíntese, indicando que o mesmo está sendo consumido. Fazendo um comparativo entre este perfil de concentração e a avaliação visual de seu sinal de fluorescência representado na Figura 20, é possível afirmar que parece haver uma inconsistência entre eles. Por exemplo, nos gráficos das MEE da Figura 20, referentes aos tempos de 32 e 36 horas, é possível visualizar que o sinal dessa espécie praticamente desaparece. Entretanto, seu perfil de concentração resolvido (Figura 23i) não é zero, indicando a existência de sinal para este composto. Depois de diversas análises, percebeu-se que a resolução está correta. Entretanto, o que acontece é que, como a intensidade de fluorescência dos outros compostos aumenta significativamente, o sinal desta espécie, que está diminuindo com o tempo, é camuflado pelos sinais dos compostos cujas concentrações estão aumentando.

O segundo fluoróforo tem um pico com máximo de emissão em 423 nm quando excitado em  $\lambda_{\text{max}}$  280 nm e 308 nm. Esta dupla absorção demonstra a estrutura vibracional do estado excitado. O perfil de concentração indica que este

## RESULTADOS E DISCUSSÃO

---

fluoróforo alcança sua máxima concentração em torno de 30 horas e então começa a ser consumido. Não foi possível identificar esse composto através da comparação de espectros de emissão e excitação/absorção molecular disponíveis na literatura ou em base de dados.

O terceiro fluoróforo apresenta banda de excitação e emissão com estruturas vibrônicas, o que indica que, tanto no estado fundamental quanto excitado, o arranjo geométrico dos núcleos é o mesmo. O perfil de concentração indica que esse fluoróforo começa a ser produzido depois de 18 horas de fermentação e alcança sua máxima concentração em 32 horas.

A Figura 23ii apresenta os mesmos perfis de emissão e concentração que a Figura 23i. Como a faixa de excitação é menor, os perfis de excitação aparecem cortados.

Na Figura 23iii, uma vez que a faixa de excitação não abrange o comprimento de onda de excitação do primeiro fluoróforo, ele não é observado na figura. A justaposição dos perfis de excitação obtidos na Figura 23ii e 23iii recuperam os perfis de excitação obtidos para toda a região espectral.

Os perfis de emissão e excitação são os mesmos nas Figuras 23i, 23ii e 23iii. Isso mostra que a forma dos espectros de emissão independe da energia (comprimento de onda) de excitação, devido à rápida conversão interna de estados excitados ou níveis vibracionais mais altos para o menor nível de energia vibracional do primeiro estado excitado.

O fluoróforo representado pela linha vermelha contínua nas figuras 23iv, 23v e 23vi apresenta perfil de concentração similar ao triptofano. Este composto parece ter três máximos de excitação (460, 525 e 560 nm) e um máximo de emissão em torno de 605 nm (ver Figura 20). Como outrora mencionado, essa múltipla absorção pode ser entendida como transições eletrônicas para diferentes estados excitados e/ou como transições eletrônicas entre níveis vibracionais no mesmo estado excitado. A mudança na forma do perfil de emissão desse composto nas figuras 23iv-v gera dúvidas se estes se tratam da mesma espécie.

Asamizu et al.<sup>97</sup> obtiveram o espectro de absorção da StaD, uma enzima envolvida na biossíntese da estaurosporina, que é homóloga a VioB. Este espectro apresentou três máximos de absorção em (430, 530 e 565 nm) e, portanto, similar ao espectro de excitação do fluoróforo descrito no parágrafo anterior.

O composto representado pela linha tracejada vermelha na Figura 23iv mostra ampla banda de excitação (400-470 nm) e pico de emissão ao redor de 512 nm. Seu perfil de concentração indica que ele está sendo produzido durante a biossíntese da violaceína. Por fim, o perfil de excitação na Figura 23vi com máximo centrado em 570 nm e o perfil de emissão com máximo em 670 nm são similares aos respectivos espectros de absorção e emissão da violaceína<sup>95</sup>.

### **Modelagem dos dados de fluorescência utilizando o método MCR-ALS**

Em uma etapa posterior, os dados de fluorescência foram arranjados em uma matriz aumentada ( $D_{\text{augment}}$ ) e explorados com o método MCR-ALS aplicando as restrições de trilinearidade e não-negatividade para o mesmo número de fatores outrora determinados pelo valor de CORCONDIA. Os perfis de emissão obtidos por PARAFAC e perfis de concentração obtidos por EFA foram utilizados como estimativas iniciais para o procedimento iterativo. Observou-se que ambas as estimativas levaram à mesma solução.

Os modelos obtidos pelo método MCR-ALS para os dados de fluorescência foram comparados àqueles previamente obtidos por PARAFAC através da falta de ajuste e percentagem de variância explicada (Tabela 3).

## RESULTADOS E DISCUSSÃO

**Tabela 3.** Figuras de mérito obtidas para os modelos obtidos pelos métodos MCR-ALS e PARAFAC.

	Falta de ajuste		R <sup>2</sup>	
	MCR-ALS	PARAFAC	MCR-ALS	PARAFAC
<b>Toda região espectral</b>	7,15	6,61	99,49	99,56
<b>B</b>	5,68	5,49	99,68	99,69
<b>C</b>	5,56	5,38	99,69	99,71
<b>D</b>	18,76	18,69	96,48	96,51
<b>E</b>	16,99	16,94	97,11	97,13
<b>F</b>	13,33	11,82	98,22	98,60

Para todos os modelos, o método PARAFAC apresentou os menores resultados de falta de ajuste e as maiores percentagem de variância explicada. Entretanto, estes valores são bastante próximos dos obtidos pelo método MCR-ALS. Esses resultados indicam que não existem sérios desvios da trilinearidade nos dados, condição requerida pelo método PARAFAC.

Os parâmetros estimados pelos dois métodos foram comparados baseados no cálculo do ângulo entre vetores para cada fator. Se os perfis recuperados por PARAFAC e MCR-ALS forem idênticos, o ângulo entre estes será zero. Quanto mais distante de zero, maior a dissimilaridade entre os perfis. Estes resultados são apresentados na Tabela 4.

## RESULTADOS E DISCUSSÃO

**Tabela 4.** Grau de similaridade entre os perfis recuperados por PARAFAC e MCR-ALS obtidos através do cálculo do arccosseno entre os dois vetores. Unidades em graus ( $^{\circ}$ ).

	Modo 1	Modo 2	Modo 3
<b>Toda a região espectral</b>			
Composto 1	0,20	0,45	0,43
Composto 2	1,96	1,12	0,69
Composto 3	2,69	0,26	1,43
<b>Região B</b>			
Composto 1	0,14	1,06	0,34
Composto 2	1,30	0,24	0,96
Composto 3	2,09	0,72	1,69
<b>Região C</b>			
Composto 2	2,55	2,38	2,01
Composto 3	1,32	0,90	0,65
<b>Região D</b>			
Composto 4	0,87	0,29	0,67
Composto 5	0,20	0,97	0,42
<b>Região E</b>			
Composto 4	0,75	0,32	0,45
Composto 5	0,84	0,74	0,74
<b>Região F</b>			
Composto 4	2,87	3,38	4,08
Composto 6	14,20	8,56	14,57

Aqui é possível observar que os ângulos variaram entre 0,14 e 14,57 graus. Com exceção da região F, as resoluções pelos dois métodos mostraram-se bastante similares. A discrepância observada para a região F é justificada pelas diferentes faltas de ajustes obtidas pelos métodos PARAFAC e MCR-ALS, conforme apresentado na Tabela 3.

De acordo com o exposto acima, a aplicação dos métodos PARAFAC e MCR-ALS em dados de fluorescência permitiu obter informações qualitativas de fluoróforos que parecem estar envolvidos na biossíntese da violaceína. Entretanto, a grande dificuldade encontrada consistiu na identificação de tais espécies, uma

vez que poucos espectros de fluorescência de produtos naturais são encontrados na literatura.

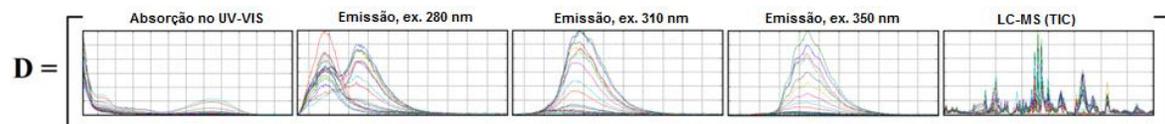
No sentido de melhor caracterizar essas espécies e obter informações temporais mais confiáveis do processo, investigou-se a possibilidade de se explorar essa biossíntese avaliando, simultaneamente, dados provenientes de diversas técnicas analíticas, através do método de resolução de curvas MCR-ALS.

### Modelagem dos dados fundidos utilizando o método MCR-ALS

As matrizes de absorção no UV-VIS, fluorescência e de ULPC-MS dos dezoito extratos etanólicos coletados durante diferentes estágios da biossíntese da violaceína foram analisadas em detalhes.

Seguindo o procedimento descrito na parte experimental, uma matriz de dados **D** de dimensão (18 x 2187) foi obtida, onde as linhas dessa matriz descrevem o processo biossintético (tempo de reação) e as colunas trazem a informação do sinal processado em diferentes canais (comprimentos de onda e razão massa/carga).

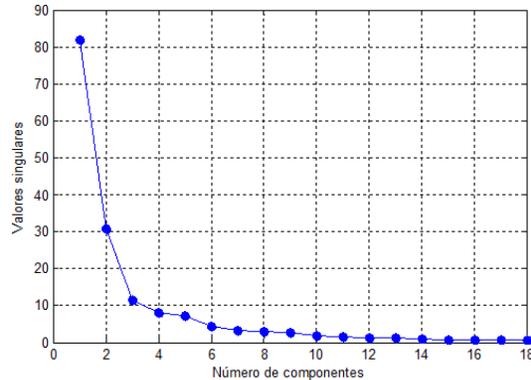
A visualização gráfica dessa matriz de dados é apresentada na Figura 24.



**Figura 24.** Visualização gráfica da matriz de dados formada pela fusão de diferentes técnicas analíticas.

A determinação do número de componentes foi baseada na comparação dos valores singulares desta matriz (Figura 25)

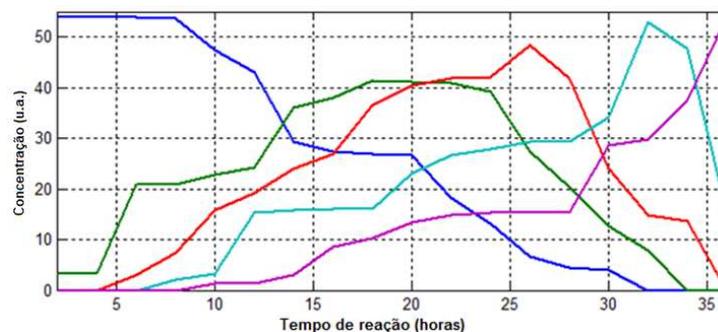
## RESULTADOS E DISCUSSÃO



**Figura 25.** Gráfico dos valores singulares

Pelo gráfico da Figura 25 é possível inferir que cinco é, provavelmente, o número de componentes químicos dessa matriz, uma vez que seus valores singulares diminuem acentuadamente até o quinto valor e então estabilizam lentamente para os outros restantes, que contêm principalmente ruído experimental.

Os perfis de concentração dos componentes estimados por EFA são apresentados na Figura 26. Estes perfis foram usados como estimativas iniciais na otimização por quadrados mínimos alternados, aplicando restrição de não-negatividade tanto no modo de concentração quanto no modo espectral.

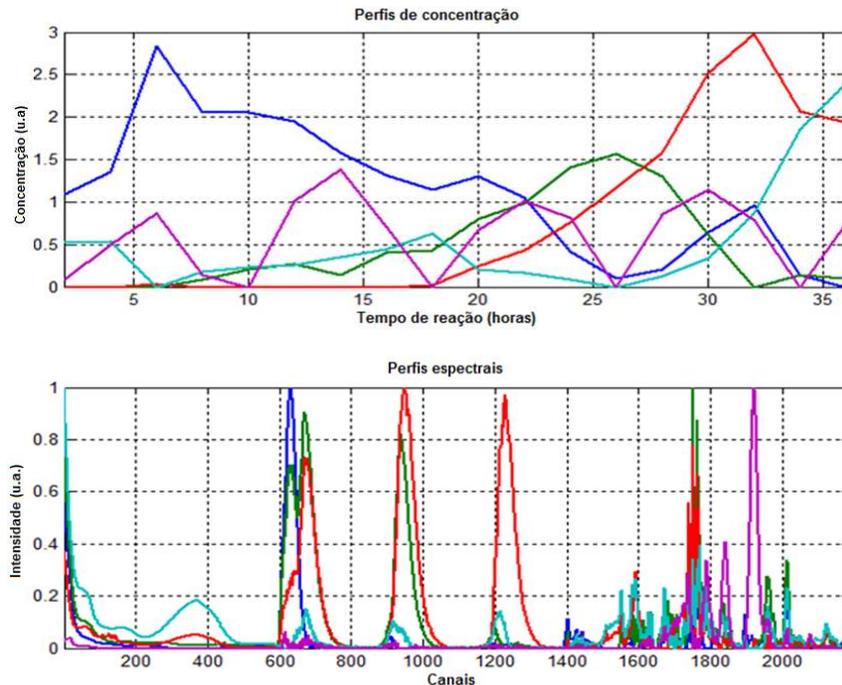


**Figura 26.** Perfis de concentração dos componentes obtida por EFA.

Os perfis de concentração e espectros puros obtidos com MCR-ALS são mostrados na Figura 27. Este modelo final com 5 componentes apresentou

## RESULTADOS E DISCUSSÃO

valores de falta de ajuste e porcentagem de variância explicada de 8,94% e 99,20%, respectivamente. Tendo em vista a complexidade dos mesmos, esses valores apresentaram-se bastante satisfatórios.



**Figura 27.** Perfis de concentração e espectrais obtidos com MCR-ALS sob restrição de não-negatividade.

Embora obtidos sem utilizar nenhuma informação prévia sob o sistema químico em estudo, estes resultados já descrevem aproximadamente as principais tendências observadas quando este bioprocessos foi monitorado por fluorescência.

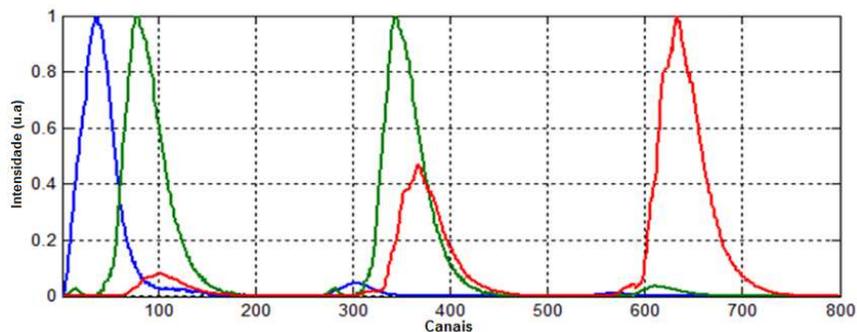
A inclusão de conhecimento adicional deve ser usada para melhorar as formas dos perfis de concentração e espectrais resolvidos e reduzir a ambiguidade rotacional associada aos modelos de decomposição bilineares. Nesse sentido, sabe-se que na espectroscopia de fluorescência, a forma dos espectros de emissão independe do comprimento de onda de excitação e vice-versa. Esta característica é consistente com a premissa de cumprimento do modelo trilinear em dados “three-way”. Portanto, a aplicação do método MCR-ALS com restrição adicional de trilinearidade é proposta para melhorar os espectros de emissão

## RESULTADOS E DISCUSSÃO

resolvidos e seus perfis de concentração correspondentes. Assim, a restrição de trilinearidade foi aplicada, especificamente, a matriz de emissão aumentada formada pelos espectros de emissão obtidos em três diferentes comprimentos de onda de excitação.

Seguindo essa estratégia, diferentes modelos MCR-ALS com 3, 4 e 5 componentes foram testados na análise dessa matriz. O modelo com 3 componentes pareceu ser mais adequado, mostrando uma falta de ajuste de 5,1% e uma percentagem de variância explicada de 99,73%.

Os espectros puros de emissão assim obtidos são mostrados na Figura 28.



**Figura 28.** Perfis espectrais obtidos por MCR-ALS com restrições de não-negatividade e trilinearidade nas matrizes de fluorescência.

Basicamente, é possível inferir que o primeiro componente representado pela linha azul mostrou fluorescência efetiva quando excitado em 280 nm. O segundo componente representado pela linha verde apresentou fluorescência quando excitado em 280 e 310 nm. O terceiro componente representado pela linha vermelha mostrou fluorescência efetiva quando excitado em 310 e 350 nm. Adicionalmente, é possível verificar que os espectros de emissão de uma dada espécie apresentam a mesma forma, independente do comprimento de onda de excitação.

Utilizando essa informação, uma restrição de seletividade adicional nos perfis espectrais foi então implementada para melhorar a resolução da matriz de

## RESULTADOS E DISCUSSÃO

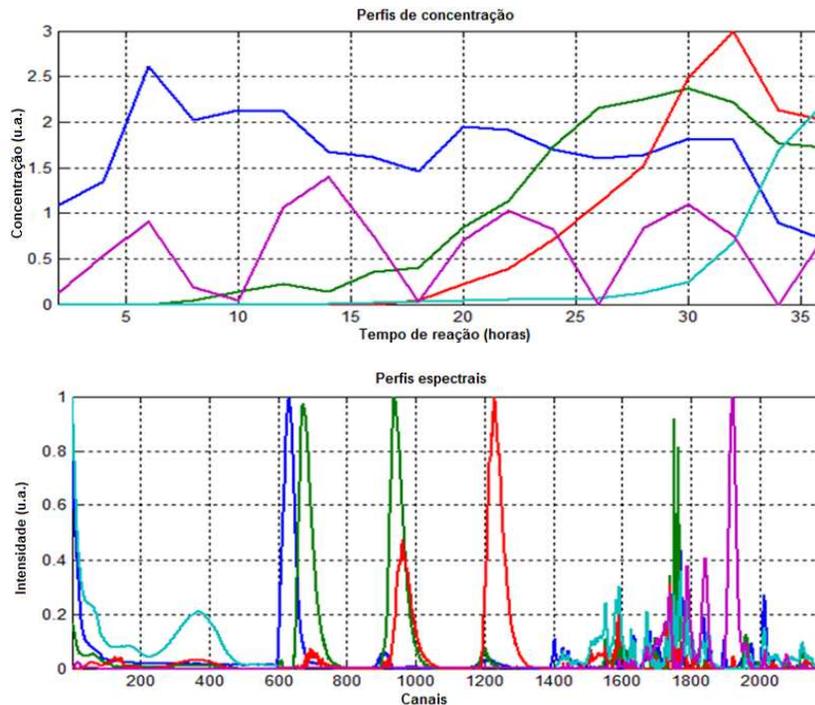
---

dados fundida. Neste caso, uma matriz  $S_{sel}$  com as mesmas dimensões da matriz de perfis espectrais  $S$  ( $5 \times 2187$ ) foi gerada utilizando valores não definidos, representados pela notação NaN. Alguns elementos da matriz  $S_{sel}$  referentes aos componentes que não produzem sinal num dado canal devem ser restritos. Nesse caso, os valores de NaN relativos a estes canais devem ser substituídos por algum valor muito pequeno, como por exemplo,  $10^{-9}$ .

De acordo com a solução obtida na Figura 28, apenas 3 dos 5 componentes estão presentes na região referente aos dados de fluorescência. Assim, os dois últimos componentes da matriz de perfis espectrais, representados pelas linhas de cor ciano e magenta na Figura 27, foram restritos para a região de fluorescência. Essa imposição produziu um modelo onde a falta de ajuste foi de 10,77% e a variância total explicada foi de 98,84%.

As mudanças relativamente pequenas quando se comparou os modelos restrito e não restrito indicaram que esta restrição pode ser aceita, uma vez que sua aplicação não levou a um aumento significativo na falta de ajuste (aumentou somente 1,83%), nem a uma diminuição significativa na variância explicada (diminuiu somente 0,36%). Os resultados do modelo MCR-ALS com a restrição de seletividade são apresentados na Figura 29.

## RESULTADOS E DISCUSSÃO



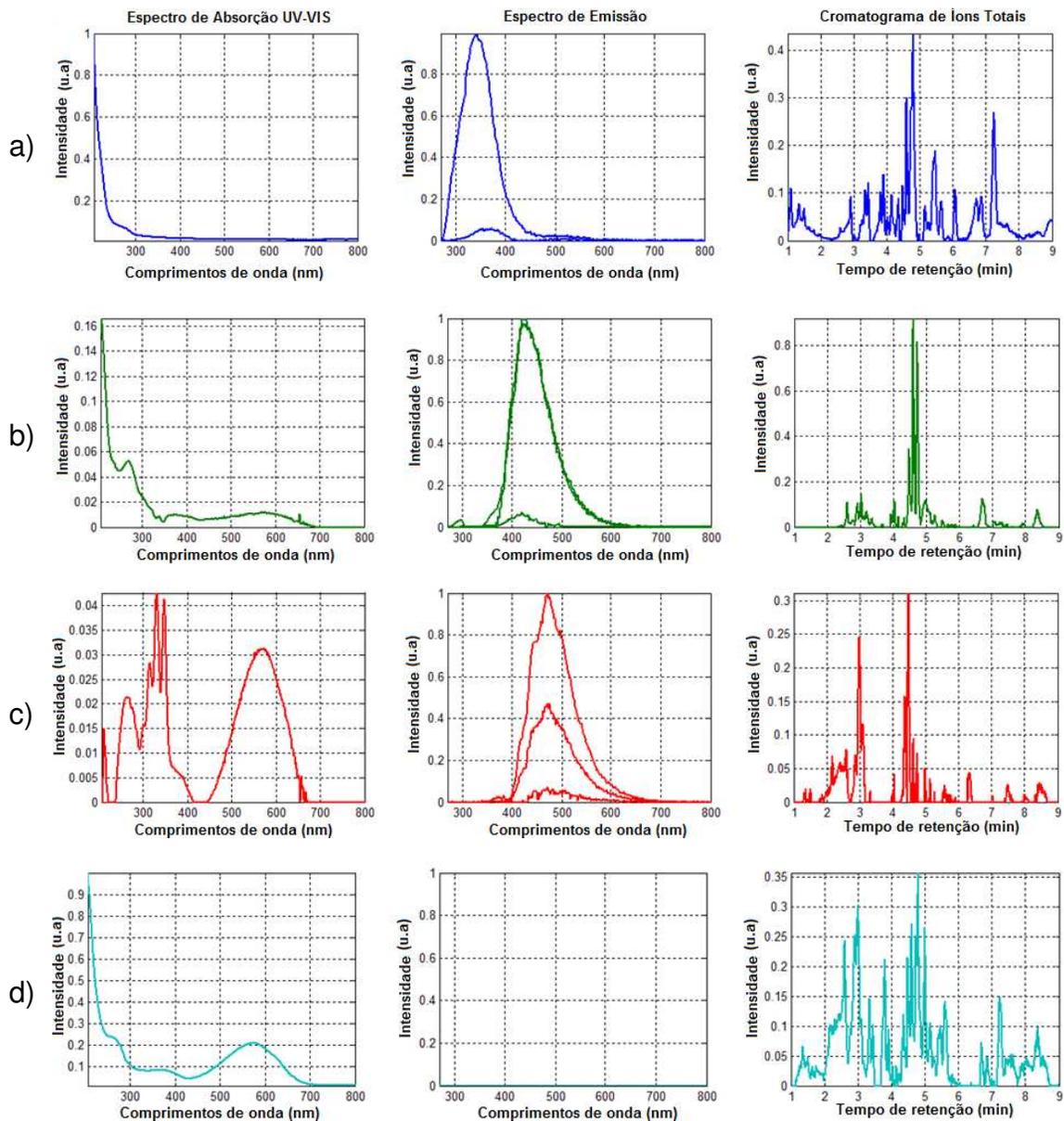
**Figura 29.** Perfis de concentração e espectrais resolvidos por MCR-ALS sob restrições de não-negatividade e seletividade.

Os resultados obtidos desta forma são mais satisfatórios e adequados do ponto de vista químico. O perfil de concentração representado pela linha de cor azul tem uma tendência em diminuir com o tempo de reação. O perfil representado pela linha de cor verde indica que esta espécie começa a ser produzida após seis horas de reação e alcança seu máximo de concentração após aproximadamente 30 horas. O perfil de concentração representado pela linha de cor vermelha mostra que o terceiro composto começa a ser produzido aproximadamente, após 15 horas e alcança seu máximo de concentração após 32 horas. O perfil da linha ciano mostra que a concentração dessa quarta espécie aumenta progressivamente com o tempo de reação. A forma não usual do perfil de concentração representado pela linha violeta descreve uma provável contribuição de um sinal espectral de fundo ou “background”. Por fim, os perfis de concentração representados pelas linhas de cores azul, verde e vermelho são similares àqueles anteriormente reportados obtidos quando somente a espectroscopia de fluorescência e métodos de resolução de curvas foram

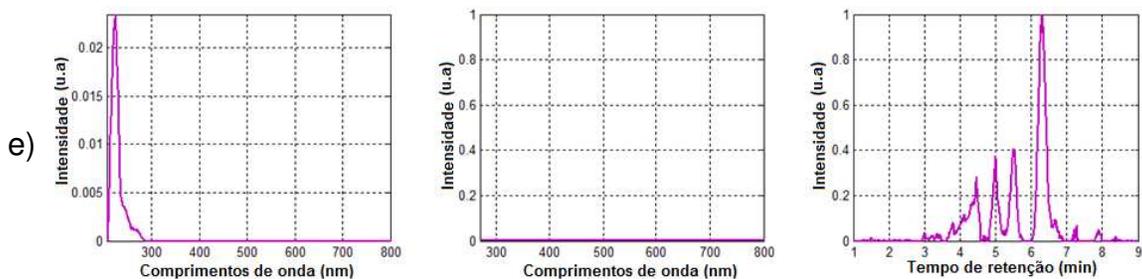
## RESULTADOS E DISCUSSÃO

utilizados. Nos perfis espectrais, a aplicação desta restrição de seletividade levou a eliminação de sinais referentes aos quarto e quinto componentes espectrais nas matrizes de emissão e permitiu descrever os três primeiros espectros de emissão com forma independente do comprimento de onda de excitação.

A Figura 30 mostra os perfis espectrais puros dos componentes individuais.



## RESULTADOS E DISCUSSÃO



**Figura 30.** Perfis espectrais obtidos com MCR-ALS.

Os espectros de absorção resolvidos nas Figuras 30a-d parecem corresponder a espécies químicas que possuem grupos indólicos em sua estrutura devido às bandas de absorção ao redor de 270 nm. Embora o espectro de absorção da terceira espécie, 30c, tenha duas bandas de absorção com máximos em 270 e 570 nm, seu espectro de excitação, previamente reportado na resolução dos dados de fluorescência por PARAFAC e MCR-ALS, corrobora a larga banda com estrutura vibracional localizada entre 310-360 nm. Portanto, o espectro resolvido desse terceiro componente parece ser a soma de espectros puros de diferentes espécies químicas. Este fato pode ser explicado devido à presença de dependência linear nos perfis de concentração das espécies correspondentes. O espectro de absorção representado na figura 30d mostra três bandas de absorção com máximos em 270, 370 e 350 nm. Este espectro é bastante similar ao espectro de absorção da violaceína<sup>95</sup>. O quinto espectro de absorção resolvido, 30e, tem uma banda de absorção com máximo em 225 nm.

Os espectros de emissão são consistentes com aqueles representados previamente na Figura 28 obtidos por MCR-ALS com restrição de trilinearidade. O espectro de emissão, representado na Figura 30a, é similar ao espectro de emissão do triptofano puro<sup>96</sup>. As espécies 30d e 30e não são fluorescentes quando excitados em 280, 310 e 350 nm.

A interpretação direta dos perfis de eluição cromatográficos resolvidos não é tão fácil, pois diversos picos são alocados para cada composto resolvido. Isto leva a questões tais como: Por que mais de um pico estão situados em cada perfil

## RESULTADOS E DISCUSSÃO

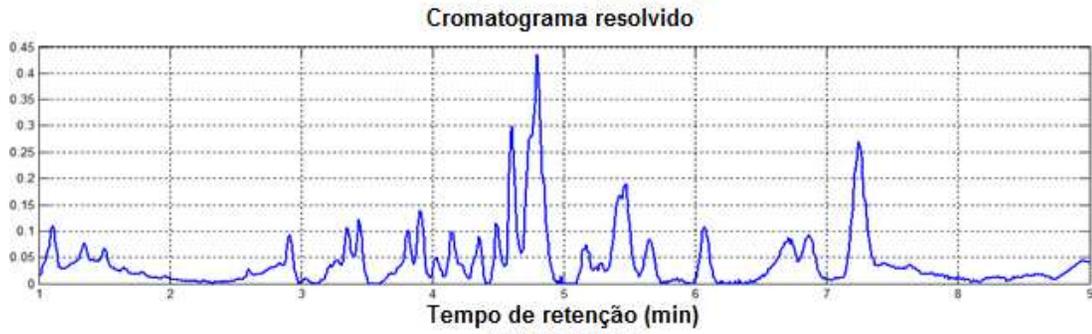
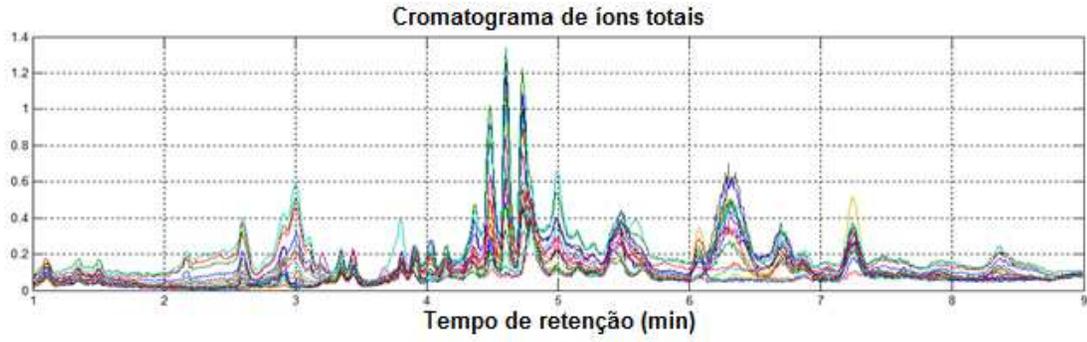
---

de eluição? Sabe-se que o sistema de detecção por massas possibilita a determinação de um maior número de espécies que as técnicas de absorção e fluorescência molecular. Assim, a estimativa do número total de espécies químicas presentes em um sistema pode ser difícil se estas espécies não contribuem diferentemente e significativamente para a variância observada nos dados. Portanto, nessas situações, é compreensível que estes compostos extras sejam alocados em algum lugar durante a resolução, dando mais que um pico para cada espécie.

Na tentativa de relacionar tempos de retenção individuais com os cinco compostos previamente resolvidos, uma análise de correlação foi feita entre a matriz  $D_{TIC}$  e cada perfil de concentração obtido na Figura 29. Os resultados dessa proposta são mostrados na Figura 31.

Para melhor interpretar estes resultados, primeiro é necessário propor um critério para a identificação das espécies. Aqui, este critério é baseado no produto cruzado (elemento por elemento) entre os cromatogramas resolvidos e os correlogramas calculados. Assim, a espécie mais provável é aquela que possui o maior valor positivo para este produto.

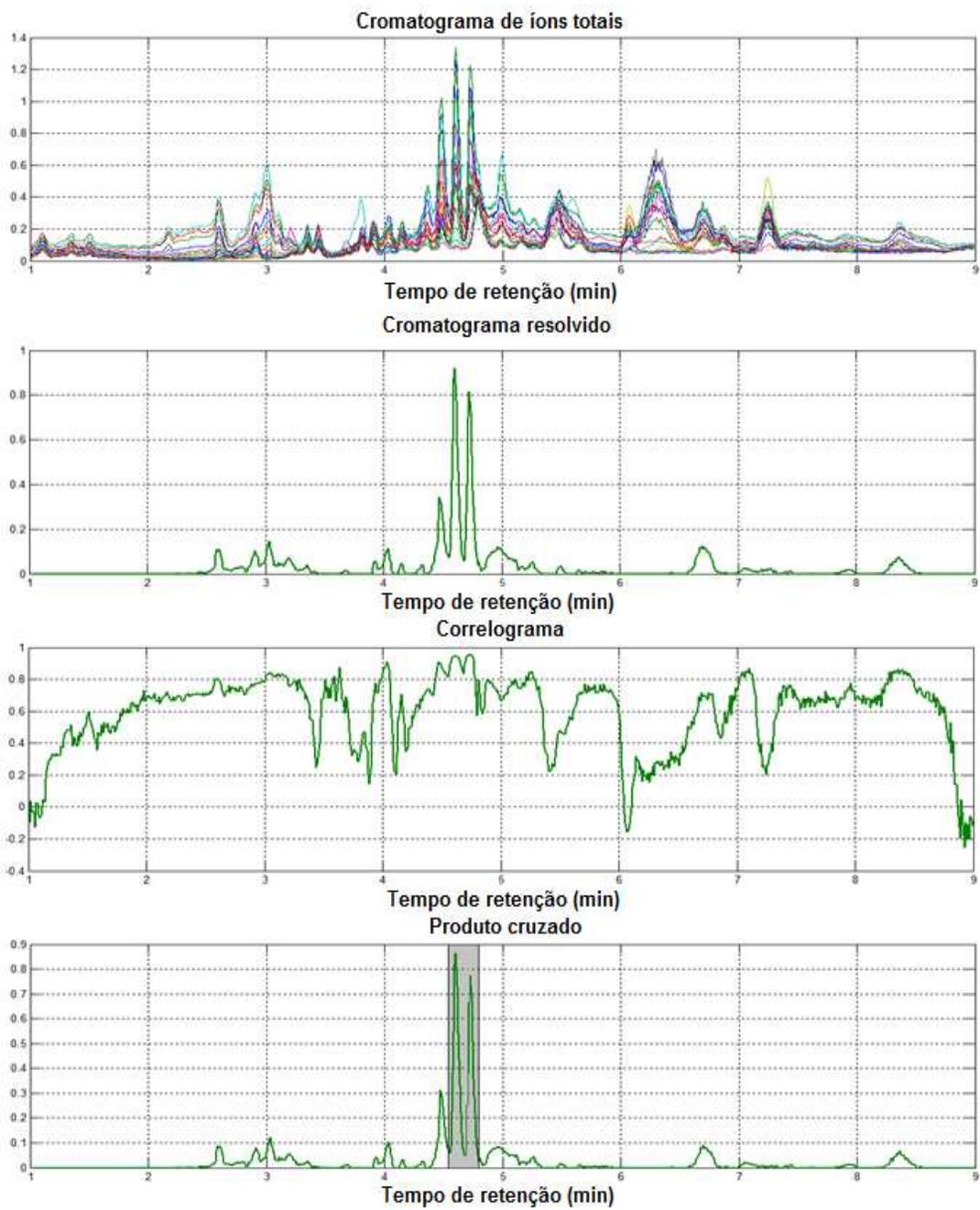
## RESULTADOS E DISCUSSÃO



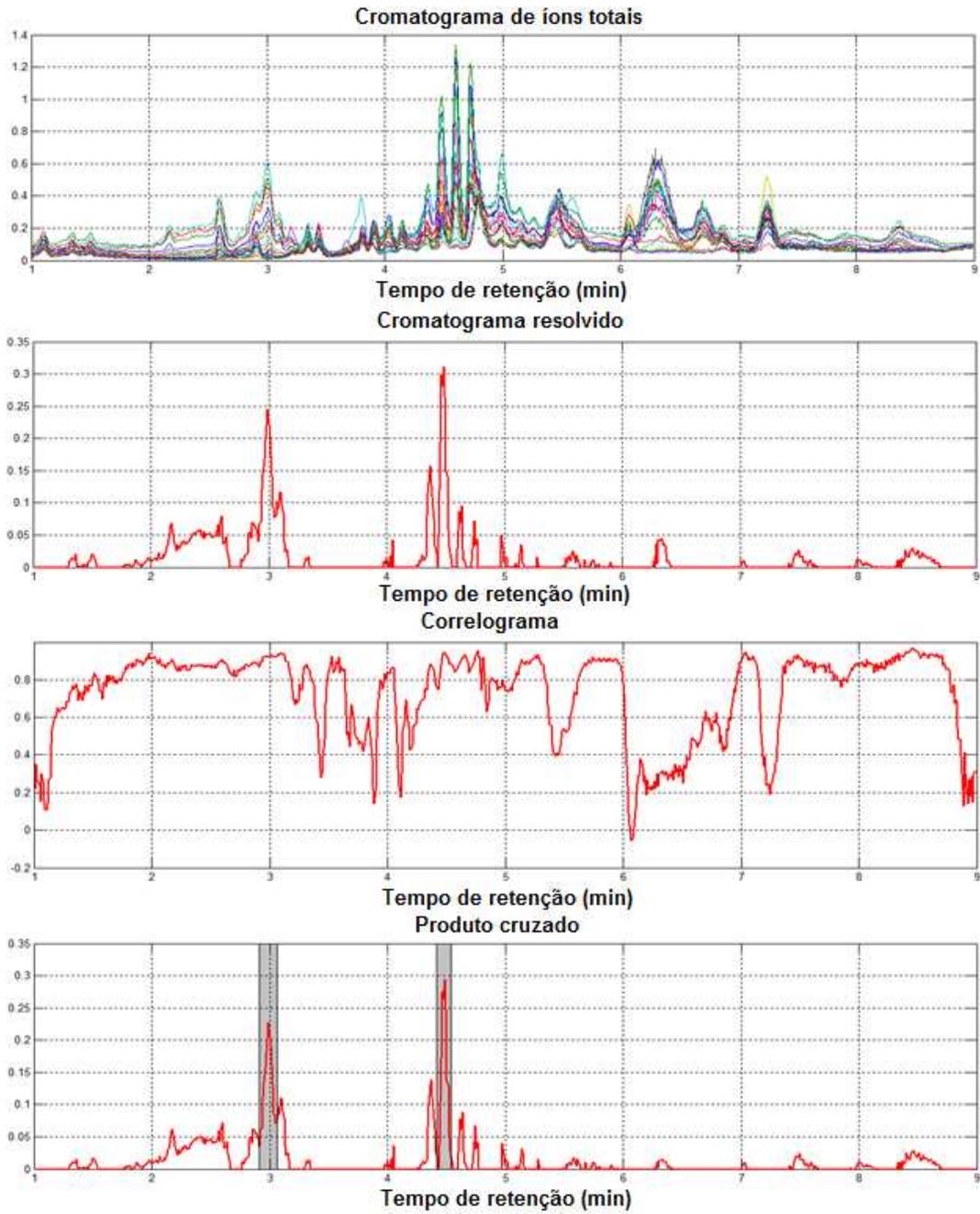
a)



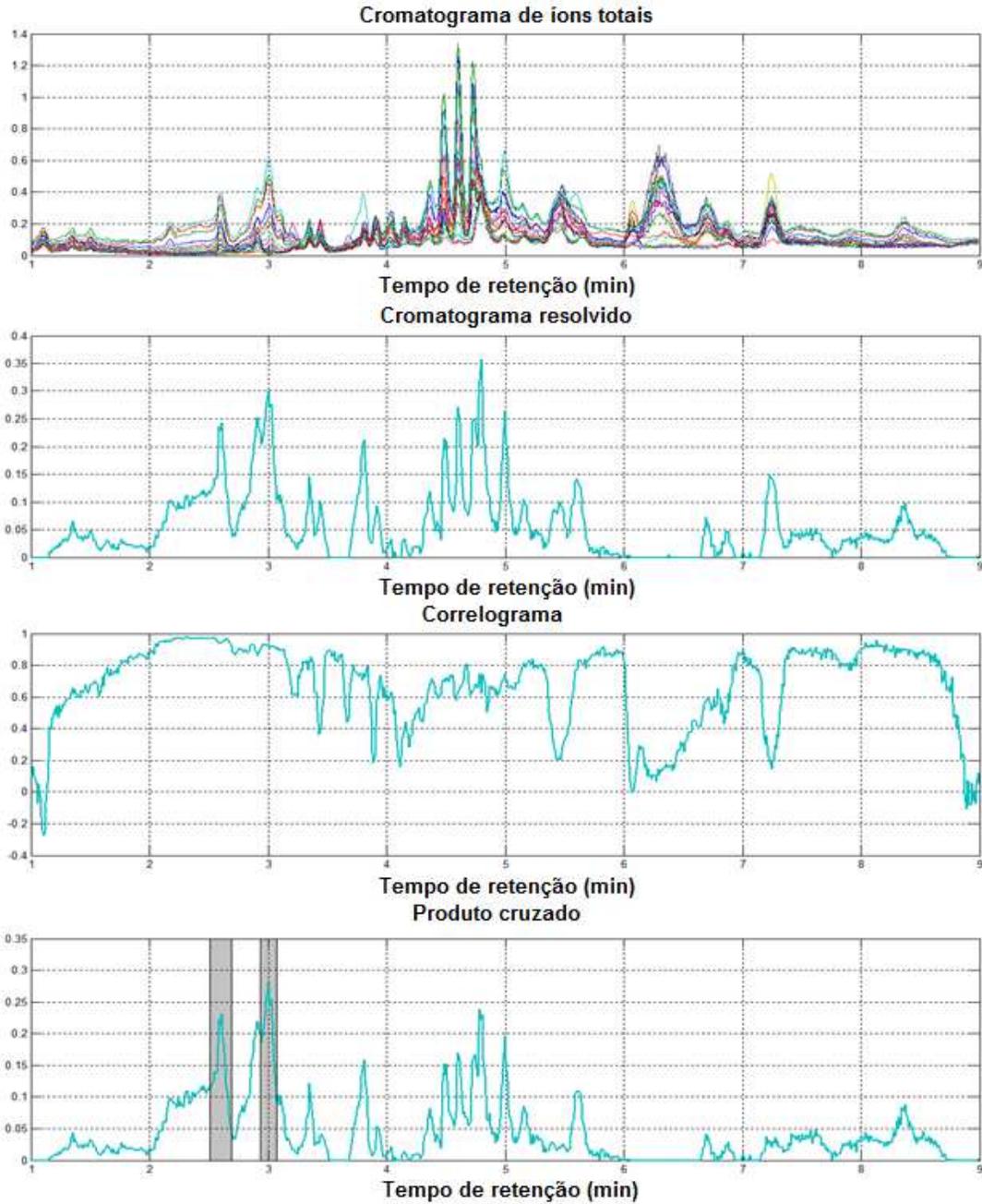
## RESULTADOS E DISCUSSÃO



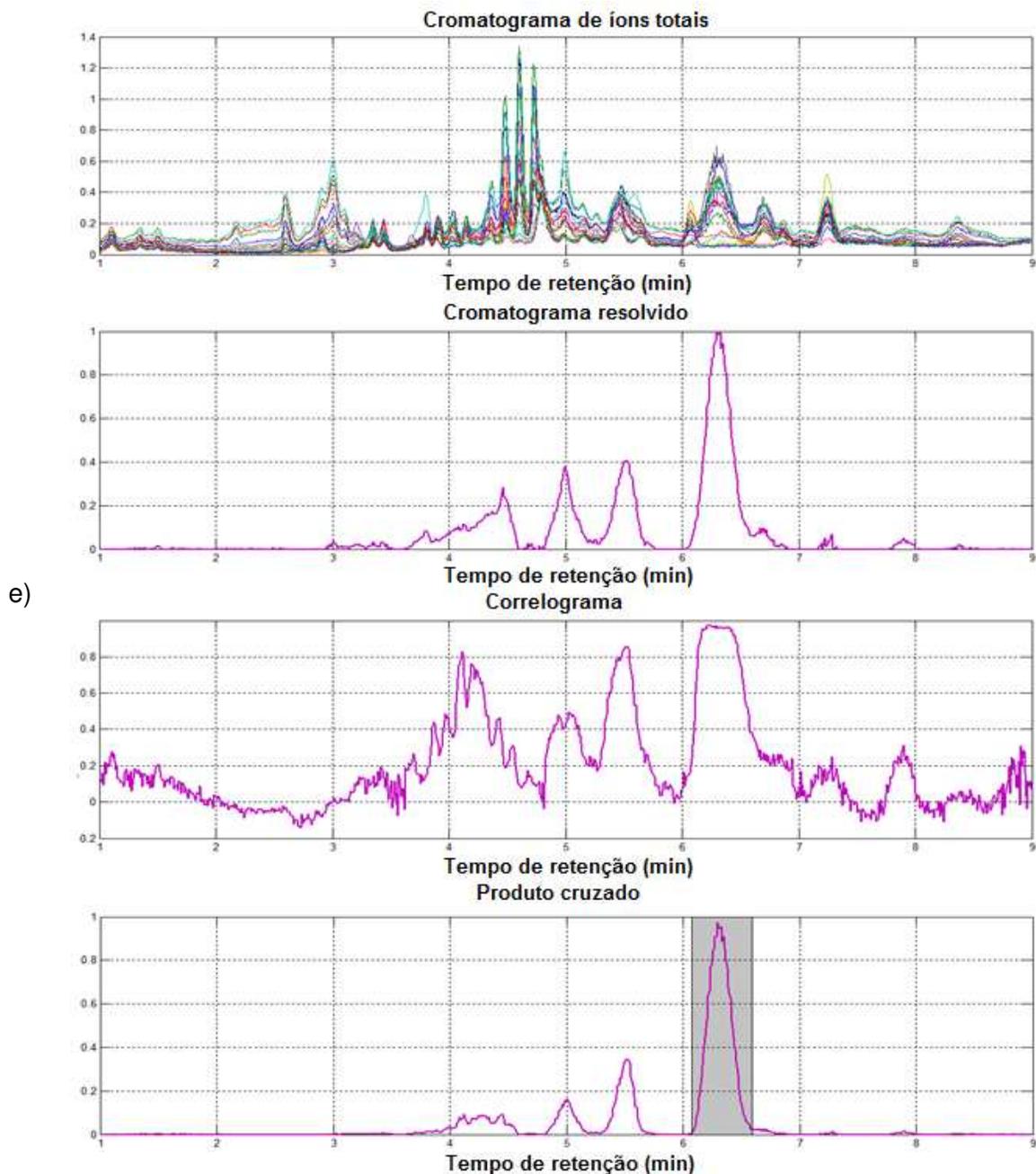
## RESULTADOS E DISCUSSÃO



## RESULTADOS E DISCUSSÃO



## RESULTADOS E DISCUSSÃO

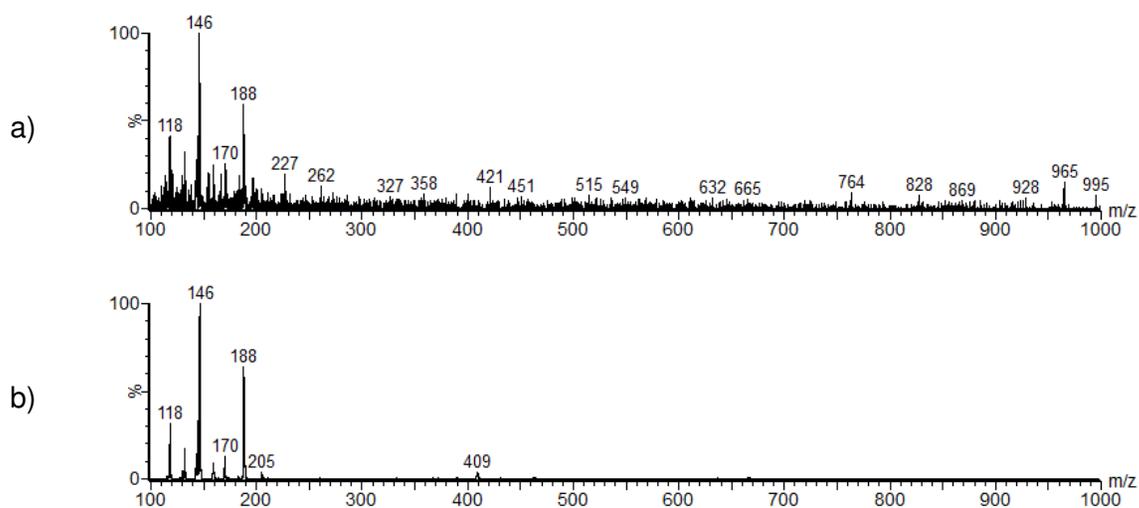


**Figura 31.** Gráficos superiores, cromatogramas de íons totais; gráficos medianos, cromatograma de íons totais resolvidos por MCR-ALS; gráficos inferiores, correlogramas e produto cruzado dos cinco componentes modelados.

Na Figura 31a, diversos compostos químicos foram alocados no primeiro cromatograma resolvido, mas usando o critério acima mencionado, o composto

## RESULTADOS E DISCUSSÃO

mais provável está localizado entre 1,08-1,12 min. Esta espécie foi previamente identificada como triptofano de acordo com seu espectro de emissão. Para verificar se esta espécie era de fato o triptofano, o espectro de massas de uma amostra no tempo de retenção de 1,1 min foi comparado com o espectro de massas de uma solução padrão de L-triptofano (*Vide* Figura 32).

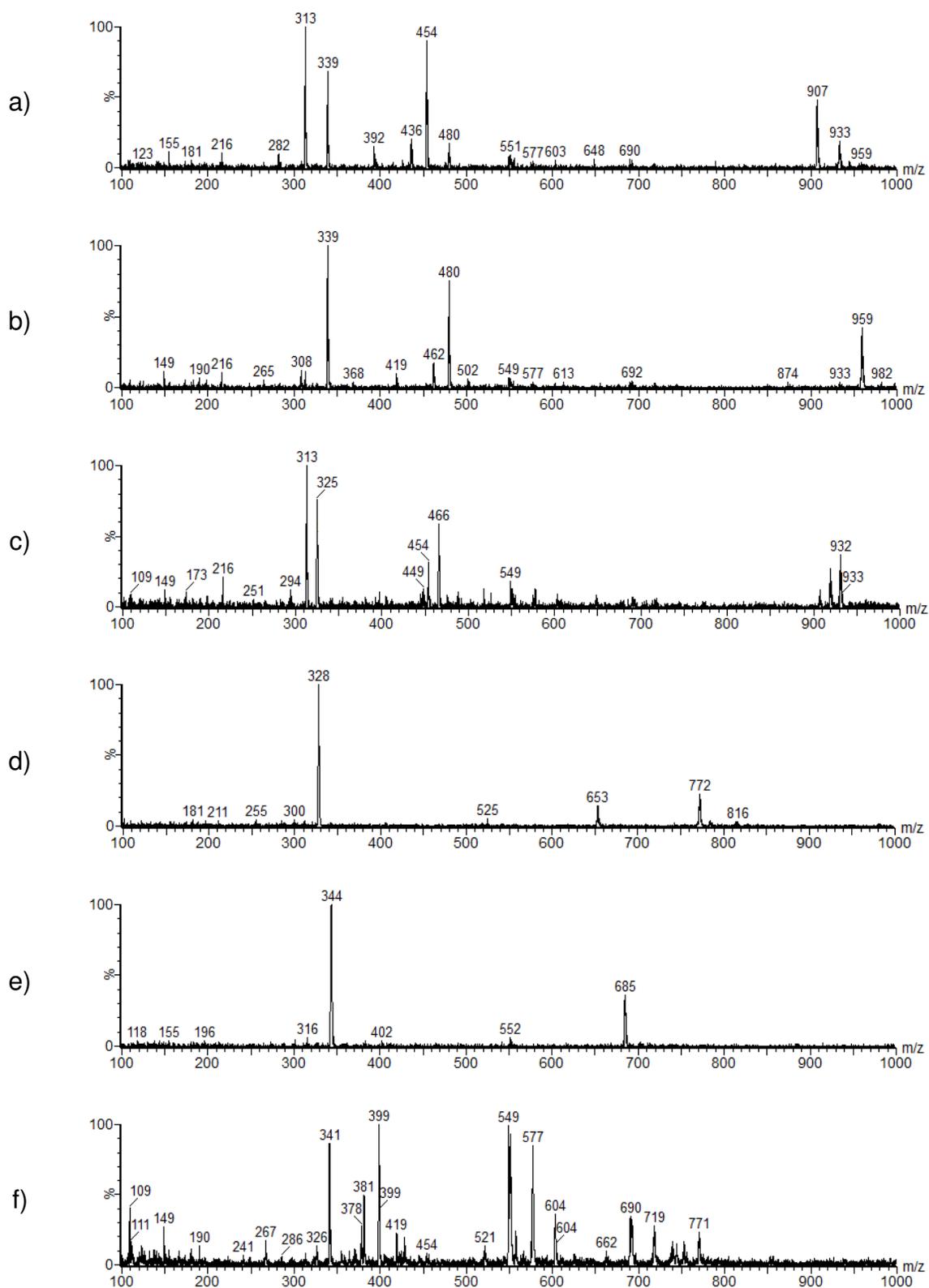


**Figura 32.** Espectros de massas obtidos no modo positivo via ionização por “electrospray” a) extrato etanólico da *C. violaceum* em 1,1 minuto. B) solução padrão de L-triptofano.

Como pode ser visto na Figura 32, a comparação dos íons gerados entre a amostra e o padrão confirma que esta espécie é o triptofano e este resultado corrobora o mecanismo apresentado na Figura 14, onde este composto é o precursor da biossíntese da violaceína.

Na Figura 31b é possível ver que tanto o cromatograma resolvido quanto o correlograma indicam que os tempos de retenção mais prováveis estão localizados entre 4,55-4,78 min. Dois compostos são observados nessa faixa. Seus espectros de massas são mostrados na Figura 33a-b.

## RESULTADOS E DISCUSSÃO



**Figura 33.** Espectros de massas de uma amostra de extrato etanólico da *C. violaceum* nos tempos de retenções a) 4,60 min; b) 4,73 min; c) 4,47 min; d) 3,01 min; e) 2,61 min; f) 6,33 min.

Os íons com  $m/z$  313, 339, 392, 436, 454, 480 e 907 podem ser visualizados no espectro representado na Figura 33a. O monitoramento da fragmentação dos íons 454 e 907 com um segundo analisador de massas permitiu inferir que este último é a forma dimérica do íon molecular protonado ( $[M+H]^+$ ) com  $m/z$  454. Portanto, a massa molecular desta espécie é 453 u, indicando um número ímpar de átomos de nitrogênio presentes nessa estrutura. Adicionalmente, o íon com  $m/z$  313 é derivado do  $[M+H]^+$  e os íons com  $m/z$  436 e 392 foram identificados como  $[M+H-H_2O]^+$  e  $[M+H-H_2O-CO_2]^+$ , respectivamente. Os íons com  $m/z$  339 e 480 não foram derivados do  $[M+H]^+$ , sugerindo que outro composto coeluiu com este.

Na Figura 33b, o espectro de massas apresenta um íon com  $m/z$  339 dominante. O monitoramento da fragmentação dos íons 480 e 959 com um segundo analisador de massas permitiu inferir que o sinal em  $m/z$  959 é devido à forma dimérica do íon molecular protonado ( $[M+H]^+$ ) com  $m/z$  480. Portanto, a massa molecular desta espécie é 479 u. Novamente, um número ímpar de átomos de nitrogênio está presente nessa estrutura. O íon com  $m/z$  462 foi atribuído a  $[M+H-H_2O]^+$  e o íon com  $m/z$  339 é derivado do íon molecular protonado.

Na Figura 31c é possível ver que os tempos de retenção mais prováveis estão localizados entre 4,44-4,54 e 2,94-3,04 min. Nessas regiões, dois compostos são observados. O primeiro, representado pelo espectro de massas obtido em 4,47 min é apresentado na Figura 33c. Aqui os íons com  $m/z$  313, 325, 466 e 932 podem ser visualizados. O monitoramento da fragmentação dos íons 466 e 931 com um segundo analisador de massas permitiu inferir que o sinal em  $m/z$  931 é devido à forma dimérica do íon molecular protonado com  $m/z$  466. Portanto, a massa molecular dessa espécie é 465 u. O íon base com  $m/z$  313 não é derivado do  $[M+H]^+$ , sugerindo que a espécie com massa 453 u representada na Figura 33a coeluiu com esta.

## RESULTADOS E DISCUSSÃO

---

As espécies representadas nas Figuras 33a-c diferem por 26, 14 e 12 unidades de massas, respectivamente e a perda de um fragmento com  $m/z$  141 (454-313  $m/z$ , 480-339  $m/z$  e 466-325  $m/z$ ) é comum para as três espécies.

Uma vez que todos os compostos representados na Figura 14 apresentam massas menores que 400 u, especula-se que os compostos com massas moleculares de 453, 465 e 479 u, aqui detectados, sejam intermediários dessa biossíntese. Infelizmente, estudos de elucidação estrutural para estas espécies não foram realizados neste trabalho.

O segundo composto representado pelo espectro de massa obtido em 3,01 min é apresentado na Figura 33d. O íon com  $m/z$  328 foi identificado como o íon molecular protonado ( $[M+H]^+$ ) e, portanto, com massa molecular de 327 u. A literatura reporta três isômeros com essa massa e composição molecular  $C_{20}N_3O_2H_{13}$  envolvidos na biossíntese da violaceína, a saber: deoxiviolaceína, pseudodeoxiviolaceína e proviolaceína. Este resultado corrobora a discussão sobre o espectro de absorção na Figura 30c, interpretado como uma combinação linear dos espectros puros de duas espécies que apresentam dependência linear entre seus perfis de concentração. Portanto, a banda com estrutura vibracional localizada entre 310-360 nm pode ser associada ao composto com massa molecular 465 u e os picos em 280 e 570 nm relacionados ao composto com massa molecular de 328 u, o qual apresenta estrutura similar à violaceína.

Na Figura 31d é possível ver que tanto o cromatograma resolvido quanto o correlograma indicam que os tempos de retenção mais prováveis estão localizados entre 2,94-3,04 e 2,54-2,66 min. Dois compostos são observados nessas regiões. O primeiro composto, representado pelo espectro de massa obtido em 3,01 min e apresentado na Figura 33d, é o mesmo discutido acima. O segundo é caracterizado pelo espectro de massa obtido em 2,61 min, representado na Figura 33e. O íon com  $m/z$  344 foi atribuída à forma protonada da violaceína, que possui massa molecular de 343 u e fórmula  $C_{20}N_3O_3H_{13}$ .

## RESULTADOS E DISCUSSÃO

---

Na Figura 31e é possível inferir que a região mais provável está localizada na faixa entre 6,1-6,5 min, sendo que um único pico pôde ser visualizado. Diversos íons com  $m/z$  341, 381, 399, 549 e 577 são vistos no espectro de massa representado na Figura 33f. A atribuição do íon molecular neste espectro é dificultada pela existência desses íons. Especula-se que estes se devem aos aglomerados de moléculas do solvente. Entretanto, não se sabe justificar porque seu perfil de concentração apresenta comportamento quase cíclico, conforme apresentado na Figura 29.

### CONCLUSÕES

O uso da espectroscopia de fluorescência em combinação com métodos quimiométricos de resolução de curvas (PARAFAC e MCR-ALS) foi introduzido como uma nova e complementar metodologia aos experimentos de incorporação isotópica, para se obter informações sobre a biossíntese de produtos naturais fluorescentes. O poder dessa metodologia foi demonstrado utilizando a biossíntese da violaceína.

O comportamento dinâmico dos fluoróforos envolvidos na biossíntese desse metabólito foi caracterizado através dos mapas de fluorescência total, ou matrizes de emissão excitação, onde a dependência da sensibilidade da fotomultiplicadora com o comprimento de onda utilizado na excitação demonstrou a necessidade de utilização de diferentes configurações instrumentais para detectar sinais nas regiões espectrais de menor energia.

Dos seis compostos detectados com essa metodologia, dois (triptofano e violaceína) foram identificados por comparação dos espectros resolvidos com dados existentes na literatura.

A comparação das resoluções obtidas por PARAFAC e MCR-ALS mostrou uma ligeira vantagem do método PARAFAC em se ajustar melhor a este conjunto de dados, devido a não existência de sérios desvios da trilinearidade nos mesmos. Este fato comprova a eficiência da metodologia aqui utilizada para excluir a região de espalhamento que não segue um comportamento trilinear.

A dificuldade de identificação de alguns compostos foi a etapa crítica deste procedimento, levando a proposição da fusão de dados de diferentes técnicas analíticas e a sua análise por MCR-ALS como uma tentativa de identificação dessas espécies.

Além de triptofano e violaceína que novamente foram bem caracterizados, esta metodologia permitiu identificar outros compostos com massas moleculares

## CONCLUSÕES

---

de 453, 465 e 479 u, até então não reportados na literatura como intermediários desta biossíntese.

Por fim, a aplicação dos métodos de resolução de curvas MCR-ALS e PARAFAC a dados experimentais gerados por diferentes técnicas analíticas provou ser uma poderosa ferramenta para investigar a biossíntese de produtos naturais.

Como perspectiva futura, pretende-se explorar a biossíntese da violaceína maneira quantitativa, utilizando a vantagem dos métodos de segunda ordem, que permitem a determinação quantitativa de espécies de interesse na presença de interferentes, através do uso de padrões. Além disso, almeja-se aplicar esta metodologia a outros sistemas biológicos como, por exemplo, na síntese de indolcarbazois.

### REFERÊNCIAS BIBLIOGRÁFICAS

1. Kiralj, R., and Ferreira, M. M. C. The past, present, and future of chemometrics worldwide: some etymological, linguistic, and bibliometric investigations, *Journal of Chemometrics* **2006**, *20*, 247-272.
2. Keithley, R. B., Mark Wightman, R., and Heien, M. L. Multivariate concentration determination using principal component regression with residual analysis, *TrAC Trends in Analytical Chemistry* **2009**, *28*, 1127-1136.
3. Booksh, k. S., and Kowalski, B. R. Theory of analytical chemistry, *Analytical Chemistry* **1994**, *7*, 782A-719A.
4. Khatri, C. G. R., R. Solutions to some functional equations and their applications to characterization of probability distributions, *Sankhya: The Indian Journal of Statistics* **1968**, *30*, 167-180.
5. Wold, S., Esbensen, K., and Geladi, P. Principal component analysis, *Chemometrics and Intelligent Laboratory Systems* **1987**, *2*, 37-52.
6. Borgen, O. S., and Kowalski, B. R. An extension of the multivariate component-resolution method to three components, *Analytica Chimica Acta* **1985**, *174*, 1-26.
7. Lawton, W. H., and Sylvestre, E. A. Self modeling curve resolution, *Technometrics* **1971**, *13*, 617-633.
8. Tauler, R. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution, *Journal of Chemometrics* **2001**, *15*, 627-646.
9. Tauler, R., Kowalski, B., and Fleming, S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial-process, *Analytical Chemistry* **1993**, *65*, 2040-2047.
10. Tauler, R., Smilde, A., and Kowalski, B. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, *Journal of Chemometrics* **1995**, *9*, 31-58.
11. Rutan, S. C., de Juan, A., and Tauler, R. Introduction to multivariate curve resolution, In: Brown, S., Tauler, R., and Walczak, B. (Eds), *Comprehensive Chemometrics*, Elsevier, Oxford, **2009**.
12. Malinowski, E. R. *Factor analysis in chemistry*, 2nd ed., Wiley & Sons, New York, **1991**.
13. Watkins, D. *Fundamentals of matrix computations*, 2nd ed., Wiley & Sons, New York, **2002**.
14. Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi, P. J., Smeyers-Verbeke, J. *Handbook of chemometrics and qualimetrics: Part A*, Elsevier, Amsterdam, **1997**.
15. de Juan, A., Rutan, S. C., and Tauler, R. Two-way data analysis: multivariate curve resolution-iteative resolution methods, In: Brown, S., Tauler, R., and Walczak, B. (Eds), *Comprehensive Chemometrics*, Elsevier, Oxford, **2009**.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

16. Tauler, R. Multivariate curve resolution applied to second order data, *Chemometrics and Intelligent Laboratory Systems* **1995**, *30*, 133-146.
17. Vandeginste, B. G. M., Derks, W., and Kateman, G. Multicomponent self-modelling curve resolution in high-performance liquid chromatography by iterative target transformation analysis, *Analytica Chimica Acta* **1985**, *173*, 253-264.
18. Gemperline, P. J. Target transformation factor-analysis with linear inequality constraints applied to spectroscopic chromatographic data, *Analytical Chemistry* **1986**, *58*, 2656-2663.
19. Tauler, R., Izquierdo-Ridora, A., and Casassas, E. Simultaneous analysis of several spectroscopic titrations with self-modeling curve resolution, *Chemometrics and Intelligent Laboratory Systems* **1993**, *18*, 293-300.
20. Diaz-Cruz, J. M., Tauler, R., Grabaric, B. S., Esteban, M., and Casassas, E. Application of multivariate curve resolution to voltammetric data .1. Study of zn(ii) complexation with some polyelectrolytes, *J. Electroanalytical Chemistry* **1995**, *393*, 7-16.
21. Gampp, H., Maeder, M., Meyer, C. J., and Zuberbuehler, A. D. Quantification of a known component in an unknown mixture, *Analytica Chimica Acta* **1987**, *193*, 287-293.
22. Maeder, M., and Zuberbuehler, A. D. The resolution of overlapping chromatographic peaks by evolving factor-analysis, *Analytica Chimica Acta* **1986**, *181*, 287-291.
23. Bezemer, E., and Rutan, S. C. Multivariate curve resolution with non-linear fitting of kinetic profiles, *Chemometrics and Intelligent Laboratory Systems* **2001**, *59*, 19-31.
24. Bijlsma, S., and Smilde, A. K. Application of curve resolution based methods to kinetic data, *Analytica Chimica Acta* **1999**, *396*, 231-240.
25. Bijlsma, S., and Smilde, A. K. Estimating reaction rate constants from a two-step reaction: a comparison between two-way and three-way methods, *Journal of Chemometrics* **2000**, *14*, 541-560.
26. de Juan, A., Maeder, M., Martinez, M., and Tauler, R. Combining hard- and soft-modelling to solve kinetic problems, *Chemometrics and Intelligent Laboratory Systems* **2000**, *54*, 123-141.
27. Diewok, J., de Juan, A., Maeder, M., Tauler, R., and Lendl, B. Application of a combination of hard and soft modeling for equilibrium systems to the quantitative analysis of pH-modulated mixture samples, *Analytical Chemistry* **2003**, *75*, 641-647.
28. Geladi, P., and Wold, S. Local principal component models, rank maps and contextuality for curve resolution and multi-way calibration inference, *Chemometrics and Intelligent Laboratory Systems* **1987**, *2*, 273-281.
29. Amrhein, M., Srinivasan, B., Bonvin, D., and Schumacher, M. M. On the rank deficiency and rank augmentation of the spectral measurement matrix, *Chemometrics and Intelligent Laboratory Systems* **1996**, *33*, 17-33.
30. Saurina, J., Hernandez-Cassou, S., Tauler, R., and Izquierdo-Ridora, A. Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions, *Journal of Chemometrics* **1998**, *12*, 183-203.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

31. Tauler, R., Marcel., M., and de Juan, A. Multiset data analysis: extended multivariate curve resolution, In: Brown, S., Tauler, R., and Walczak, B. (Eds), *Comprehensive Chemometrics*, Elsevier, Oxford, **2009**.
32. Tauler, R., Marqués, I., and Casassas, E. Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid–base titration of salicylic acid at three excitation wavelengths, *Journal of Chemometrics* **1998**, *12*, 55-75.
33. Cattell, R. B. Parallel proportional profiles and others principles for determining the choice of factors by rotation, *Psychometrika* **1944**, *9*, 267-283.
34. Bro, R. Multi-Way analysis in the food industry. Models, algorithm and applications, Ph.D Thesis, University of Amsterdam, Amsterdam, **1998**.
35. Sanchez, E., and Kowalski, B. R. Tensorial resolution: a direct trilinear decomposition, *Journal of Chemometrics* **1990**, *4*, 29-45.
36. Sanchez, E., and Kowalski, B. R. Tensorial calibration: II. second-order calibration, *Journal of Chemometrics* **1988**, *2*, 265-280.
37. Ferreira, M. M. C., Brandes, M. L., Ferreira, I. M. C., Booksh, K. S., Dolowy, W. C., Gouterman, M., and Kowalski, B. R. Chemometric study of the fluorescence of dental calculus by trilinear decomposition, *Applied Spectroscopy* **1995**, *49*, 1317-1325.
38. Malinowski, E. R. Window Factor Analysis -Theoretical derivation and application to flow-injection analysis data, *Journal of Chemometrics* **1992**, *6*, 29-40.
39. Manne, R., Shen, H. L., and Liang, Y. Z. Subwindow factor analysis, *Chemometrics and Intelligent Laboratory Systems* **1999**, *45*, 171-176.
40. Kvalheim, O. M., and Liang, Y. Z. Heuristic evolving latent projections - resolving 2-way multicomponent data. 1.Selectivity, latent-projective graph, datascope, local rank, and unique resolution, *Analytical Chemistry* **1992**, *64*, 936-946.
41. Liang, Y. Z., and Kvalheim, O. M. Heuristic evolving latent projections - resolving hyphenated chromatographic profiles by component stripping, *Chemometrics and Intelligent Laboratory Systems* **1993**, *20*, 115-125.
42. Liang, Y. Z., Kvalheim, O. M., Keller, H. R., Massart, D. L., Kiechle, P., and Erni, F. Heuristic evolving latent projections - resolving 2-way multicomponent data. 2.Detection and resolution of minor constituents, *Analytical Chemistry* **1992**, *64*, 946-953.
43. Mason, C., Maeder, M., and Whitson, A. Resolving factor analysis, *Analytical Chemistry* **2001**, *73*, 1587-1594.
44. Harshman, R. A. Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis, *UCLA Working Papers in Phonetics* **1970**, *16*, 1-84.
45. Harshman, R. A., and Lundy, M. E. PARAFAC-parallel factor analysis, *Computational Statistics & Data Analysis* **1994**, *18*, 39-72.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

46. Harshman, R. A. PARAFAC2: Mathematical and technical notes, *UCLA Working Papers in Phonetics* **1972**, *22*, 30-47.
47. Tucker, L. R. Some mathematical notes on 3-mode factor analysis, *Psychometrika* **1966**, *31*, 279.
48. Kroonenberg, P. M. *Three-Mode Principal Component Analysis*, DSWO Press, Leiden, **1983**.
49. Windig, W., and Guilment, J. Interactive self-modeling mixture analysis, *Analytical Chemistry* **1991**, *63*, 1425-1432.
50. Maeder, M. Evolving factor-analysis for the resolution of overlapping chromatographic peaks, *Analytical Chemistry* **1987**, *59*, 527-530.
51. Maeder, M., and Zilian, A. Evolving factor-analysis, a new multivariate technique in chromatography, *Chemometrics and Intelligent Laboratory Systems* **1988**, *3*, 205-213.
52. Gampp, H., Maeder, M., Meyer, C. J., and Zuberbuhler, A. D. Calculation of equilibrium-constants from multiwavelength spectroscopic data. 3. Model-free analysis of spectrophotometric and electron-spin-resonance titrations, *Talanta* **1985**, *32*, 1133-1139.
53. Jaumot, J., Gargallo, R., de Juan, A., and Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemometrics and Intelligent Laboratory Systems* **2005**, *76*, 101-110.
54. Keller, H. R., Massart, D. L., and Debeer, J. O. Window evolving factor-analysis for assessment of peak homogeneity in liquid-chromatography, *Analytical Chemistry* **1993**, *65*, 471-475.
55. Whitson, A. C., and Maeder, M. Exhaustive evolving factor analysis (E-EFA), *Journal of Chemometrics* **2001**, *15*, 475-484.
56. Zeng, Z. D., Xu, C. J., Liang, Y. Z., and Li, B. Y. Sectional moving window factor analysis for diagnosing elution chromatographic patterns, *Chemometrics and Intelligent Laboratory Systems* **2003**, *69*, 89-101.
57. Windig, W., Antalek, B., Lippert, J. L., Batonneau, Y., and Bremard, C. Combined use of conventional and second-derivative data in the SIMPLISMA self-modeling mixture analysis approach, *Analytical Chemistry* **2002**, *74*, 1371-1379.
58. Windig, W., Gallagher, N. B., Shaver, J. M., and Wise, B. M. A new approach for interactive self-modeling mixture analysis, *Chemometrics and Intelligent Laboratory Systems* **2005**, *77*, 85-96.
59. Sena, M. M., Poppi, R. J., Frighetto, R. T. S., and Valarini, P. J. Avaliação do uso de métodos quimiométricos em análise de solos, *Química Nova* **2000**, *23*, 547-556.
60. de Juan, A., Tauler, R. Multivariate curve resolution (MCR) from 2000: progress in concepts and applications, *Critical Reviews in Analytical Chemistry* **2006**, *36*, 163-176.
61. Bro, R. PARAFAC. Tutorial and applications, *Chemometrics and Intelligent Laboratory Systems* **1997**, *38*, 149-171.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

62. Faber, N. M., Bro, R., and Hopke, P. K. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review, *Chemometrics and Intelligent Laboratory Systems* **2003**, *65*, 119-137.
63. Bro, R., and Kiers, H. A. L. A new efficient method for determining the number of components in PARAFAC models, *Journal of Chemometrics* **2003**, *17*, 274-286.
64. Andersen, C. M., and Bro, R. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data, *Journal of Chemometrics* **2003**, *17*, 200-215.
65. Bahram, M., Bro, R., Stedmon, C., and Afkhami, A. Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation, *Journal of Chemometrics* **2006**, *20*, 99-105.
66. Bro, R., and Vidal, M. EEMizer: Automated modeling of fluorescence EEM data, *Chemometrics and Intelligent Laboratory Systems* **2011**, *106*, 86-92.
67. Durán, N., Antonio, R. V., Haun, M., and Pilli, R. A. Biosynthesis of a trypanocide by *Chromobacterium violaceum*, *World Journal of Microbiology and Biotechnology* **1994**, *10*, 686-690.
68. Leon, L. L., Miranda, C. C., De Souza, A. O., and Duran, N. Antileishmanial activity of the violacein extracted from *Chromobacterium violaceum*, *Journal Antimicrobial Chemotherapy* **2001**, *48*, 449-450.
69. Lichstein, H. C., and Vandesand, V. F. Violacein, an antibiotic pigment produced by *Chromobacterium violaceum*, *Journal of Infectious Diseases* **1945**, *76*, 47-51.
70. Durán, N., and Menck, C. F. M. *Chromobacterium violaceum*: a review of pharmacological and industrial perspectives, *Critical Reviews in Microbiology* **2001**, *27*, 201-222.
71. Ferreira, C. V., Bos, C. L., Versteeg, H. H., Justo, G. Z., Durán, N., and Peppelenbosch, M. P. Molecular mechanism of violacein mediated human leukemia cell death, *Blood* **2004**, *104*, 1459-1464.
72. Melo, P. S. J., G.Z.; de Azevedo, M.B.M.; Duran, N.; Haun M. Violacein and its beta-cyclodextrin complexes induce apoptosis and differentiation in HL60 cells, *Toxicology* **2003**, *186*, 217-225.
73. Saraiva, V. S., Marshall, J.-C., Cools-Lartigue, J., and Burnier, M. N. J. Cytotoxic effects of violacein in human uveal melanoma cell lines, *Melanoma Research* **2004**, *14*, 421-424.
74. Andrighetti-Frohner, C. R., Antonio, R. V., Creczynski-Pasa, T. B., Barardi, C. R. M., and Simoes, C. M. O. Cytotoxicity and potential antiviral evaluation of violacein produced by *Chromobacterium violaceum*, *Memórias do Instituto Oswaldo Cruz* **2003**, *98*, 843-848.
75. Tobie, W. C. The Pigment of *Bacillus violaceus*: I. The Production, extraction, and purification of violacein, *Journal of Bacteriology* **1935**, *29*, 223-227.
76. DeMoss, R. D., and Evans, N. R. Physiological aspects of violacein biosynthesis in nonproliferating cells, *Journal of Bacteriology* **1959**, *78*, 583-588.
77. DeMoss, R. D., and Evans, N. R. Incorporation of C14-labeled substrates into violacein, *Journal of Bacteriology* **1960**, *79*, 729-733.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

78. Hoshino, T., Kondo, T., Uchiyama, T., and Ogasawara, N. Studies on the biosynthesis of violacein. 1. Biosynthesis of violacein - a novel rearrangement in tryptophan-metabolism with a 1,2-shift of the indole ring, *Agricultural and Biological Chemistry* **1987**, *51*, 965-968.
79. Hoshino, T., Takano, T., Hori, S., and Ogasawara, N. Studies on the biosynthesis of violacein. 2. Biosynthesis of violacein - origins of hydrogen, nitrogen and oxygen-atoms in the 2-pyrrolidone nucleus, *Agricultural and Biological Chemistry* **1987**, *51*, 2733-2741.
80. Hoshino, T., and Ogasawara, N. Studies on the biosynthesis of violacein. 3. Biosynthesis of violacein - evidence for the intermediacy of 5-hydroxy-L-tryptophan and the structure of a new pigment, oxyviolacein, produced by the metabolism of 5-hydroxytryptophan, *Agricultural and Biological Chemistry* **1990**, *54*, 2339-2346.
81. Pemberton, J. M., Vincent, K. M., and Penfold, R. J. Cloning and heterologous expression of the violacein biosynthesis gene-cluster from *chromobacterium violaceum*, *Current Microbiology* **1991**, *22*, 355-358.
82. Hoshino, T., Kojima, Y., Hayashi, T., Uchiyama, T., and Kaneko, K. Studies on the biosynthesis of violacein. 4. A new metabolite of tryptophan, chromopyrrolic acid, produced by *chromobacterium violaceum*, *Bioscience Biotechnology Biochemistry* **1993**, *57*, 775-781.
83. Hoshino, T., Hayashi, T., and Odajima, T. Biosynthesis of violacein. 1. Oxygenation at the 2-position of the indole ring and structures of proviolacein, prodeoxyviolacein and pseudoviolacein, the plausible biosynthetic intermediates of violacein and deoxyviolacein, *Journal of the Chemical Society, Perkin Transactions* **1995**, *1*, 1565-1571.
84. Hoshino, T., and Yamamoto, M. Studies on the biosynthesis of violacein, part VIII. Conversion from tryptophan precursor into violacein pigments by a cell-free system from *Chromobacterium violaceum*, *Bioscience Biotechnology Biochemistry* **1997**, *61*, 2134-2136.
85. Momen, A., and Hoshino, T. Biosynthesis of violacein: intact incorporation of the tryptophan molecule on the oxindole side, with intramolecular rearrangement of the indole ring on the 5-hydroxyindole side, *Bioscience Biotechnology Biochemistry* **2000**, *64*, 539-549.
86. McClean, K. H., Winson, M. K., Fish, L., Taylor, A., Chhabra, S. R., Camara, M., Daykin, M., Lamb, J. H., Swift, S., Bycroft, B. W., Stewart, G., and Williams, P. Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones, *Microbiology* **1997**, *143*, 3703-3711.
87. August, P. R., Grossman, T. H., Minor, C., Draper, M. P., MacNeil, I. A., Pemberton, J. M., Call, K. M., Holt, D., and Osburne, M. S. Sequence analysis and functional characterization of the violacein biosynthetic pathway from *Chromobacterium violaceum*, *Journal of Molecular Microbiology Biotechnology* **2000**, *2*, 513-519.
88. Howard-Jones, A. R., and Walsh, C. T. Enzymatic generation of the chromopyrrolic acid scaffold of rebeccamycin by the tandem action of RebO and RebD, *Biochemistry* **2005**, *44*, 15652-15663.
89. Sanchez, C., Brana, A. F., Mendez, C., and Salas, J. A. Reevaluation of the violacein biosynthetic pathway and its relationship to indolocarbazole biosynthesis, *ChemBioChem* **2006**, *7*, 1231-1240.

## REFERÊNCIAS BIBLIOGRÁFICAS

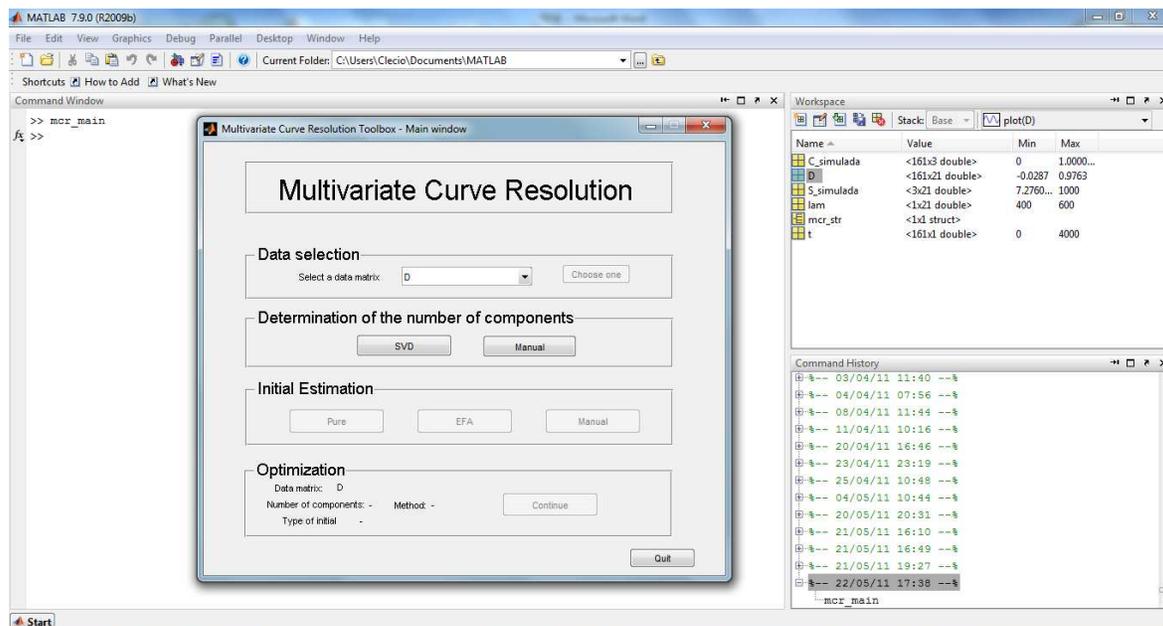
---

90. Balibar, C. J., and Walsh, C. T. In vitro biosynthesis of violacein from L-tryptophan by the enzymes VioA-E from *Chromobacterium violaceum*, *Biochemistry* **2006**, *45*, 15444-15457.
91. Shinoda, K., Hasegawa, T., Sato, H., Shinozaki, M., Kuramoto, H., Takamiya, Y., Sato, T., Nikaidou, N., Watanabe, T., and Hoshino, T. Biosynthesis of violacein: a genuine intermediate, protoviolaceinic acid, produced by VioABDE, and insight into VioC function, *Chemical Communication* **2007**, 4140-4142.
92. Guterres, M. V. Análise multivariada em N-modos: metodologias para o monitoramento de reações químicas, Tese de doutorado, Universidade Estadual de Campinas, Campinas, **2006**.
93. Eilers, P. H. C. Parametric time warping, *Analytical Chemistry* **2004**, *76*, 404-411.
94. Tortora, G. J., Funke, B. R., and Case, C. L. *Microbiology: an introduction*, 10th ed., Benjamin Cummings, San Francisco, **2009**.
95. Rettori, D. Estudos físico-químicos e biológicos da violaceína : Um pigmento produzido pela *Chromobacterium violaceum*, Tese de doutorado, Universidade Estadual de Campinas, Campinas, **2000**.
96. Du, H., Fuh, R.-C. A., Li, J., Corkan, L. A., and Lindsey, J. S. PhotochemCAD: A Computer-Aided Design and Research Tool in Photochemistry, *Photochemistry and Photobiology* **1998**, *68*, 141-142.
97. Asamizu, S., Kato, Y., Igarashi, Y., and Onaka, H. VioE, a prodeoxyviolacein synthase involved in violacein biosynthesis, is responsible for intramolecular indole rearrangement, *Tetrahedron Letters* **2007**, *48*, 2923-2926.



## ANEXOS

Um breve tutorial do programa MCR-ALS em português é apresentado a seguir. A rotina principal é chamada *mcr\_main* e é executada escrevendo-se este nome na janela de comando do MATLAB. A primeira caixa de diálogo que aparece é apresentada na Figura A1.

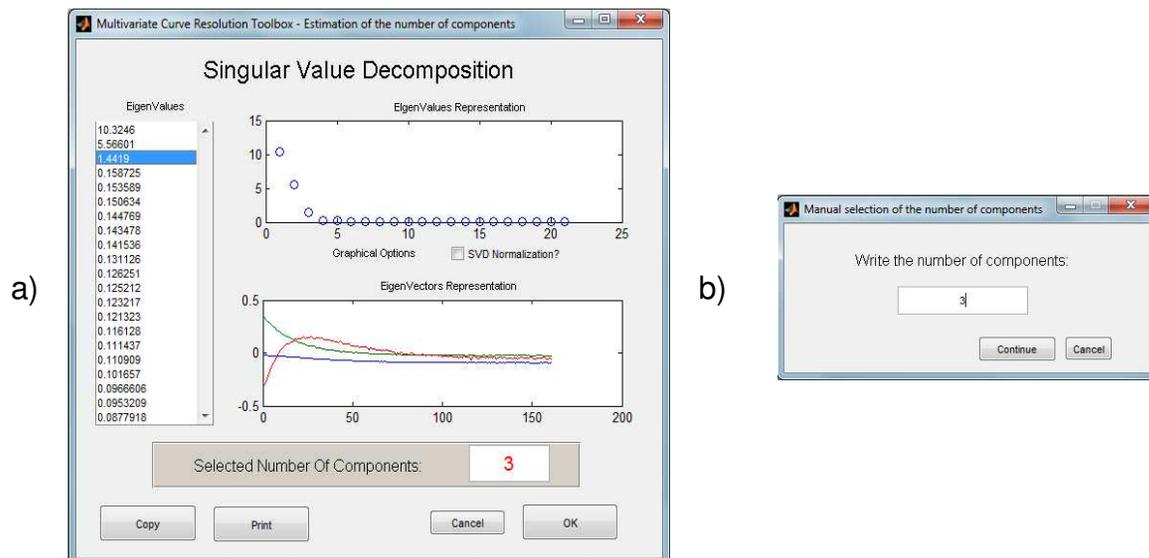


**Figura A1.** Janela de seleção do conjunto de dados utilizando MCR-ALS na plataforma MATLAB.

Nesta janela, o usuário deve selecionar o conjunto de dados a ser resolvido, determinar o número de fatores do modelo e fornecer uma estimativa inicial para os perfis de concentração ou espectrais dos componentes selecionados. Para demonstrar as diversas etapas envolvidas no emprego do método MCR-ALS, a matriz de dados representada graficamente na Figura 4 desta Tese, constituída de 161 espectros de absorção molecular, registrados durante 60 min de reação em 21 comprimentos de onda selecionados, será utilizada como exemplo.

A determinação do número de fatores ou o posto químico da matriz de dados **D** pode ser feita a partir da comparação dos respectivos valores singulares (via SVD) ou por conhecimento prévio do sistema em estudo (via manual). Ao

selecionar o botão SVD na janela da Figura A1, uma nova janela será aberta (Figura A2a), onde os valores singulares de  $\mathbf{D}$  e os autovetores da matriz  $(\mathbf{D}\mathbf{D}^T)$ , que estão armazenados em  $\bar{\mathbf{U}}$ , serão apresentados numérica e/ou graficamente. Observe nesta figura que a distribuição dos valores singulares diminui acentuadamente até o quarto valor e então se estabiliza para os outros restantes, indicando que o número de valores singulares significativos é três.



**Figura A2.** a) Janela de seleção do número de componentes por SVD; b) seleção do número de componentes manualmente.

Ainda nesta figura, vale a pena esclarecer um pequeno equívoco. No título da representação numérica e gráfica dos valores singulares, encontra-se escrito o termo autovalores (“eigenvalues”), que são definidos como o quadrado dos valores singulares e, conforme teste realizado, o que estão representados são os valores singulares e não os autovalores.

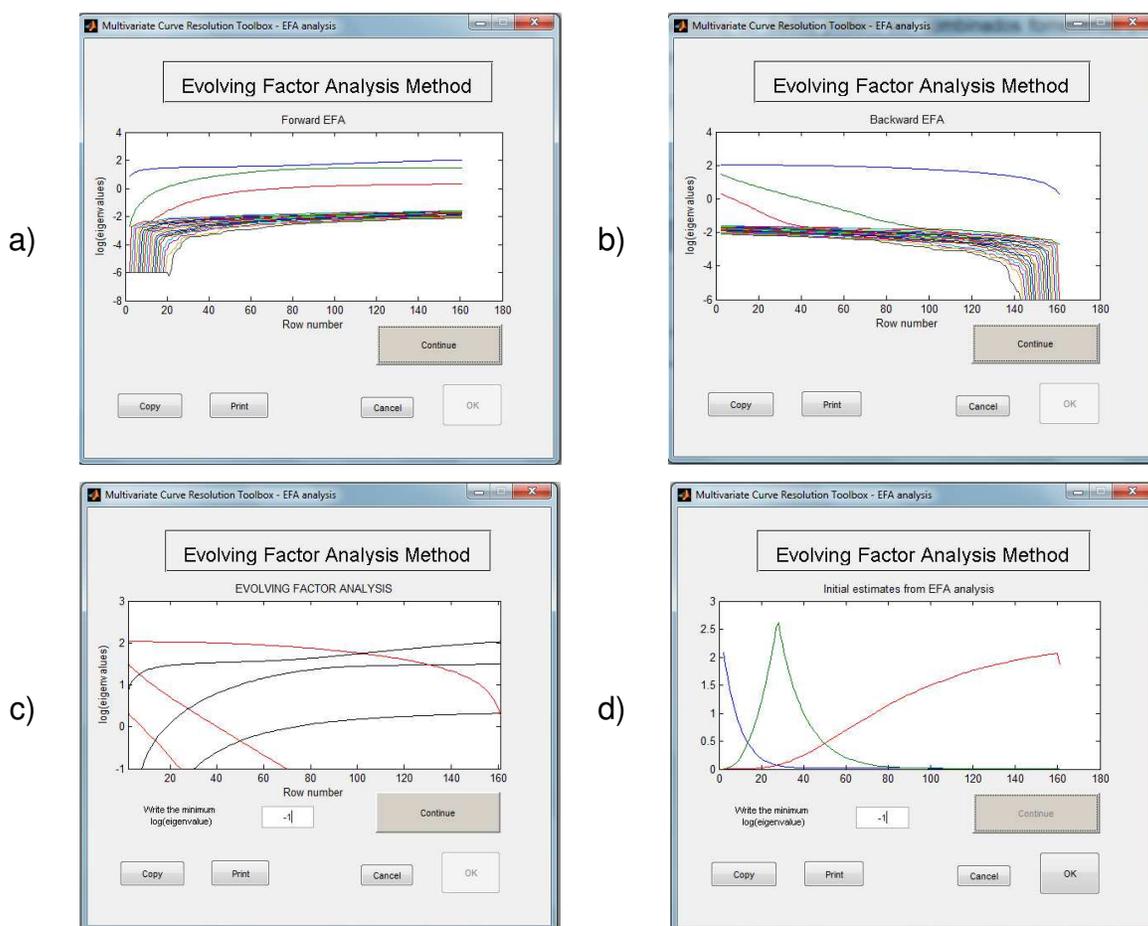
Se o número de fatores na análise é previamente conhecido, basta selecionar o botão manual na janela da Figura A1, e uma nova aba será aberta para se digitar este número (Figura A2b).

A próxima etapa consiste em obter uma estimativa inicial da matriz  $\mathbf{C}$  ou  $\mathbf{S}$ . Novamente, a exploração da matriz de dados  $\mathbf{D}$  pode fornecer informações sobre

a evolução dos componentes ou estimativas iniciais dos perfis de concentração. Esta informação pode então ser utilizada para orientar o processo de resolução e melhorar significativamente o resultado final obtido.

As principais ferramentas quimiométricas de resolução orientadas pertencem à família dos métodos de análise de posto local, como por exemplo, EFA e suas derivações, ou a métodos de seleção de variáveis puras, como por exemplo, SIMPLISMA e suas derivações.

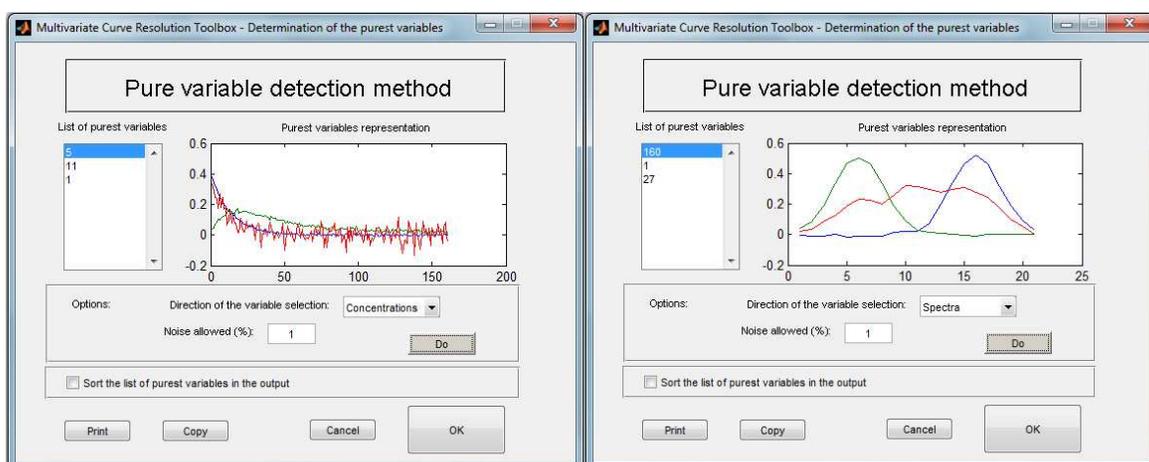
Conforme pode ser visualizado na Figura A1, o método MCR-ALS utiliza essas ferramentas de resolução orientada para obter uma estimativa inicial das matrizes **C** e/ou **S**. A sequência de janelas apresentadas na Figura A3 surgirá ao clicar no botão EFA.



**Figura A3.** Sequência de interfaces que são mostradas no método MCR-ALS para obter estimativas iniciais da matriz **C** por EFA.

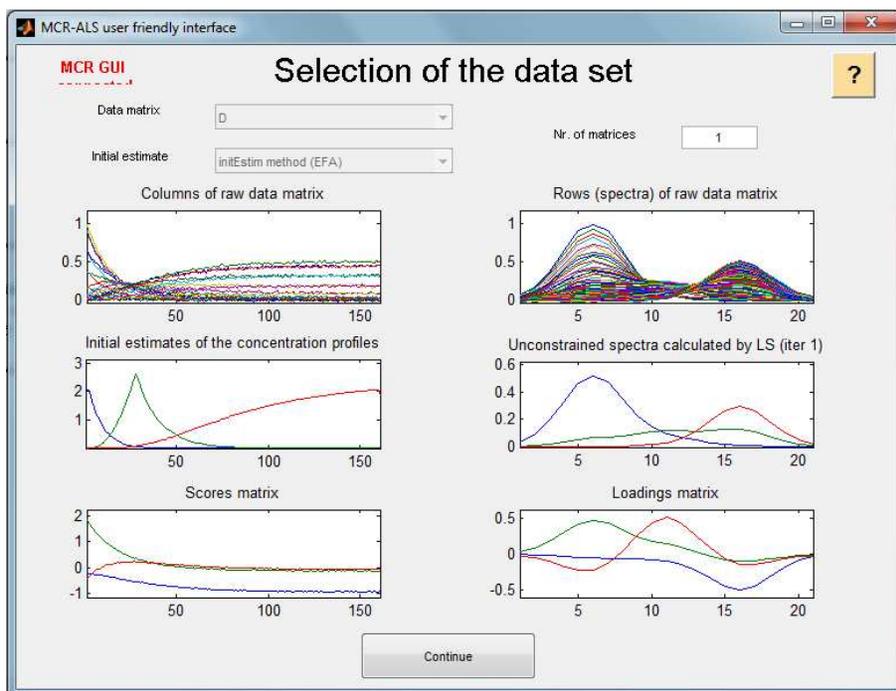
Inicialmente se determinam os valores singulares obtidos nas repetidas análises por SVD na direção de 1 a  $I$  e plotam-se os logaritmos dos valores singulares armazenados em **EFA\_F** (Figura A3a). Em seguida, executa-se o mesmo procedimento só que em submatrizes obtidas na direção oposta e plotam-se os logaritmos dos valores singulares armazenados em **EFA\_T** (Figura A3b). Observe que os valores singulares significativos, de fato, indicam o número de espécies químicas envolvidas na reação,  $A$ ,  $B$  e  $C$ . Os gráficos acima são então combinados, e um valor de corte na escala logarítmica é utilizado para eliminar os valores singulares não significativos (Figuras A3c). Por fim, uma estimativa inicial para a matriz **C** de concentração das espécies químicas é obtida (Figura A3d).

Outra maneira implementada no método MCR-ALS de se obter estimativas iniciais das matrizes **C** ou **S** consiste em utilizar métodos de seleção de variáveis puras (seletivas). Essa ferramenta pode ser facilmente acessada na interface gráfica do pacote computacional MCR-ALS selecionando o botão Pure na Figura A1. Uma nova janela será aberta (Figura A4) onde o usuário poderá selecionar a direção (coluna ou linha) em que se deseja encontrar as variáveis puras, produzindo estimativas iniciais das matrizes **C** e **S**.



**Figura A4.** Interface gráfica para obter estimativas iniciais das matrizes **C** e **S<sup>T</sup>** através de método de seleção de variáveis puras na direção das colunas (esquerda) ou linhas (direita).

Uma vez que a matriz de dados  $D$ , o número de componentes e as estimativas iniciais de  $C$  ou  $S$  foram selecionadas, seis diferentes gráficos que aparecem em uma única interface (Figura A5) informarão todas as condições iniciais disponíveis sobre o sistema em estudo.



**Figura A5.** Janela que contém toda a informação inicial para a execução do algoritmo ALS.

A informação do processo evolucionário está contida nas colunas da matriz  $D$  e é representada no gráfico superior à esquerda. Os espectros experimentais são armazenados nas linhas dessa matriz e visualizados no gráfico superior à direita. O gráfico mediano à esquerda representa as estimativas iniciais da matriz  $C$  obtidas por EFA ou SIMPLISMA, já o da direita, os perfis espectrais, ou seja, a matriz  $S$  obtida pela resolução da equação 18 em uma única iteração e sem restrições. Por fim, os gráficos inferiores representam os escores (esquerda) e pesos (direita) obtidos pela análise por componentes principais (PCA). Observe que estes últimos não apresentam, necessariamente, significado químico, ou seja, mostram concentrações e intensidades espectrais negativas e por isso são chamados de fatores abstratos.

A próxima etapa consiste em selecionar as restrições que serão aplicadas durante a resolução das equações 16 e 17. Para tanto, ao clicar no botão continuar na Figura A5, uma nova janela aparecerá (Figura A6).

The screenshot shows the 'MCR-ALS user friendly interface' window with the 'Selection of ALS constraints' dialog box open. The dialog is organized into several sections:

- No-negativity:** Includes radio buttons for 'Conco', 'Spectra', and 'Conco & Spec'. The 'Conco & Spec' option is selected. It also features dropdown menus for 'Implementation for conco' and 'Implementation for spec', both set to 'fnnis'. Input fields for 'Nr. of species with non-neg conco' and 'Nr. of species with non-neg spec' are both set to 3.
- Unimodality:** Includes radio buttons for 'Conco', 'Spectra', and 'Conco & Spec'. The 'Conco' option is selected. It features a dropdown for 'Implementation of the unimodality constraint' set to 'average'. Input fields for 'Nr. of species with unimodal conco' and 'Nr. of species with unimodal spec' are both set to 3. A 'Constraint tolerance for conco' field is set to 1.1.
- Closure:** Includes a dropdown for 'Nr. of closure constraints to be included?' set to 1. It has radio buttons for 'Conco' (selected) and 'Spectra'. Input fields for 'First Closure constant Equal to' (0.001) and 'Second Closure constant Equal to' are present. There are also fields for 'First variable closure' and 'Second variable closure constants'. A 'Closure condition' dropdown is set to 'equal than'. A 'Closure variable?' checkbox is unchecked. A 'Which species are in 1st closure?' field is set to 'All?' with a '1 1 1' matrix.
- Equality constraints in conco profiles:** Includes a 'Yes?' checkbox (checked) and a 'Select osel matrix:' dropdown set to 'osel'. A 'Constraints are' dropdown is set to 'equal than'.
- Equality constraints in spectra profiles:** Includes a 'Yes?' checkbox (unchecked) and a 'Select ssel matrix:' dropdown. A 'Constraints are' dropdown is set to 'select...'.
- Optimization parameters:** Includes 'Nr. of iterations' set to 50 and 'Convergence criterion' set to 0.1. A 'Graphical output' checkbox is checked.
- Output:** Includes input fields for 'Concentration' (copt), 'Std. dev.' (empty), 'Area opt' (empty), 'Spectra' (sopt), 'Residuals' (empty), and 'Ratio opt' (empty). There are 'Optimize', 'Done', and 'Cancel' buttons.

**Figura A6.** Janela de seleção das restrições aplicadas durante o procedimento ALS para uma única matriz de dados.

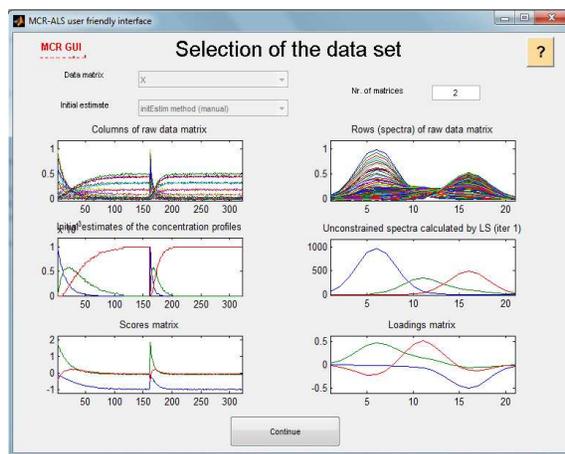
Restrições baseadas em propriedades químicas do sistema tais como não-negatividade, unimodalidade e “closure” podem ser facilmente utilizadas apenas marcando o tipo de restrição, a direção na qual ela deve ser aplicada (concentração e/ou espectral), o tipo de algoritmo a ser usado e quantos e quais componentes devem ser restritos.

As restrições de cunho matemático, associadas com o conceito de posto local, e a possibilidade de fixar valores para os perfis de concentração ou espectrais durante a otimização são aplicadas selecionando-se a caixa de restrição de igualdade. Neste caso, uma matriz deve ser montada com as mesmas dimensões de  $C$  e/ou  $S$  formada pelos valores conhecidos nas posições a serem restritas e valores não definidos, representados pela notação NaN - “Not a Number” do MATLAB. Esta matriz deve estar presente na área de trabalho do

## ANEXOS

MATLAB e seu nome deve ser escrito como entrada nessa caixa. Observe no exemplo apresentado na Figura A6 que esta restrição foi aplicada sobre os perfis de concentração segundo uma matriz de seletividade obtida por EFA, denominada **Csel**.

Voltando a Figura A5, é possível visualizar um pequeno quadro em sua parte superior direita onde se pode preencher o número de matrizes a serem analisadas. No exemplo anterior, apenas uma matriz de dados foi obtida durante o experimento. Agora, imagine que a mesma reação tenha sido monitorada pela mesma técnica em duas diferentes temperaturas. Se as formas dos perfis espectrais não mudam nas temperaturas estudadas, mas apenas os perfis de concentração, pode-se inferir que a ordem espectral é comum às duas matrizes. Neste caso, temos a vantagem de poder trabalhar com uma matriz aumentada (Figura A7).



The figure consists of two screenshots of software interfaces. The top screenshot, titled "Definition of the 3-way data set", shows a dialog box for defining the data set. It includes a dropdown menu for "Define your data set:" set to "Column-wise augmented data matrix (C direction)", a dropdown for "How many submatrices has the C matrix?" set to 2, and a dropdown for "How many submatrices has the S matrix?" set to 1. Below this, there are sections for "C submatrix" and "S submatrix", each with a "Submatrix Nr." dropdown, a "Nr. of rows" or "Nr. of columns" input field, and checkboxes for "Input rows from external variable?" and "All matrices have the same Nr. of rows?". The bottom screenshot, titled "MCR-ALS user friendly interface", shows a "Selection of ALS constraints" dialog box. It has a "Matrix Nr." dropdown set to 1 and "C submatrix" and "S submatrix" dropdowns both set to 1. The interface is divided into several sections: "Identification of species" with radio buttons for "Conc", "Spectra", and "Conc & Spec"; "Unimodality" with radio buttons for "Conc", "Spectra", and "Conc & Spec"; "Closure" with radio buttons for "Conc" and "Spectra"; "Equality constraints in conc profiles" and "Equality constraints in spectra profiles"; "Trilinear" with radio buttons for "C", "S", and "C & S"; and "Optimization parameters" with input fields for "Nr. of iterations" (150) and "Convergence criterion" (0.1). There are also checkboxes for "Graphical output" and "Optimize".

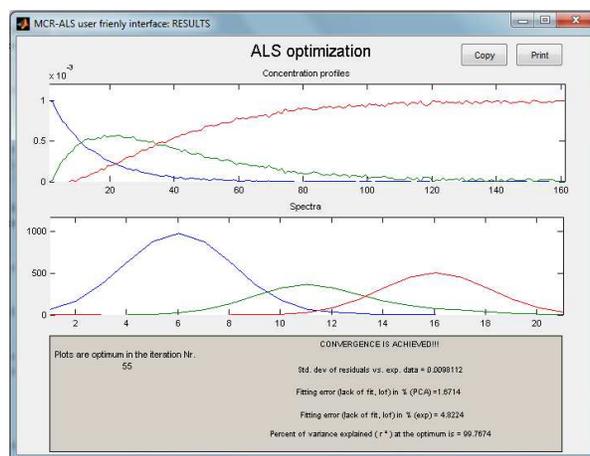
**Figura A7.** Janelas de informação inicial, definição do tipo de matriz aumentada (se por linhas ou colunas) e de seleção das restrições aplicadas durante o procedimento ALS para análise simultânea de diversas matrizes de dados.

A descrição da primeira janela é a mesma que a apresentada na Figura A5, com exceção do número de matrizes analisadas. A segunda janela dá a definição do conjunto de dados, ou seja, se a matriz é aumentada na direção das colunas. A janela de seleção das restrições durante o procedimento ALS para matrizes estendidas é mostrada na terceira interface gráfica. Nela é possível perceber que essa estrutura possibilita aplicar diferentes restrições para cada submatriz individualmente, além de definir novas restrições quando comparado à janela da

Figura A6. As novas restrições são a correspondência entre os componentes em diferentes matrizes analisadas simultaneamente e a trilinearidade.

Uma vez que as restrições são selecionadas na interface gráfica do método MCR-ALS, só resta escolher os parâmetros de otimização (Figuras A6 e A7). Estes são: o número máximo de iterações permitidas e o critério de convergência. Este último é satisfeito quando a diferença relativa no desvio padrão dos resíduos em dois ciclos iterativos consecutivos é menor que um valor previamente selecionado, geralmente escolhido como 0,1%.

Por fim, ao clicar no botão “Optimize”, o processo de otimização é iniciado e os resultados parciais obtidos nas diferentes iterações são mostrados graficamente. Uma vez que o critério de convergência é satisfeito, ou o número máximo de iterações é atingido, ou ainda no caso de divergência, a janela mostrada na Figura A8 é apresentada.



**Figura A8.** Janela de resultados da otimização por ALS.

Nesta figura, os perfis de concentração e espectrais resolvidos são exibidos, bem como as figuras de mérito relacionadas aos resultados da otimização, calculadas de acordo com as equações 20-22.