

## PARAFAC: UMA FERRAMENTA QUIMIOMÉTRICA PARA TRATAMENTO DE DADOS MULTIDIMENSIONAIS. APLICAÇÕES NA DETERMINAÇÃO DIRETA DE FÁRMACOS EM PLASMA HUMANO POR ESPECTROFLUORIMETRIA

**Marcelo M. Sena**

Unidade Universitária de Ciências Exatas e Tecnológicas, Universidade Estadual de Goiás, BR 153, km 98, 75000-000 Anápolis - GO

**Marcello G. Trevisan e Ronei J. Poppi\***

Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13084-971 Campinas - SP

Recebido em em 23/9/04; aceito em 16/2/05; publicado na web em 30/6/05

PARAFAC: A CHEMOMETRIC TOOL FOR MULTI-DIMENSIONAL DATA TREATMENT. APPLICATIONS IN DIRECT DETERMINATION OF DRUGS IN HUMAN PLASMA BY SPECTROFLUORIMETRY. Since the last decade, the combined use of chemometrics and molecular spectroscopic techniques has become a new alternative for direct drug determination, without the need of physical separation. Among the new methodologies developed, the application of PARAFAC in the decomposition of spectrofluorimetric data should be highlighted. The first objective of this article is to describe the theoretical basis of PARAFAC. For this purpose, a discussion about the order of chemometric methods used in multivariate calibration and the development of multi-dimensional methods is presented first. The other objective of this article is to divulge for the Brazilian chemical community the potential of the combination PARAFAC/spectrofluorimetry for the determination of drugs in complex biological matrices. For this purpose, two applications aiming at determining, respectively, doxorubicine and salicylate in human plasma are presented.

Keywords: parallel factor analysis; human blood; molecular fluorescence.

### INTRODUÇÃO

O termo quimiometria foi proposto no final dos anos 70 para descrever as técnicas e operações associadas ao tratamento matemático e à interpretação de dados químicos. Foi dentro da área de química analítica que a quimiometria provavelmente causou mais impacto<sup>1</sup>. Com o avanço da instrumentação e automação dentro dos laboratórios de análise, uma enorme quantidade de dados começou a ser gerada muito rapidamente. A classificação e interpretação desses dados podem ser fatores limitantes nas análises, principalmente sem a utilização de um tratamento adequado. Além disso, novas metodologias de análise, que antes nem poderiam ser pensadas, puderam ser propostas baseadas na utilização da quimiometria. Como exemplo, pode-se citar a determinação de espécies sem resolução de sinal analítico, ou seja, o sinal analítico não precisa ser proveniente apenas da espécie que se deseja determinar. A área da química analítica para qual a quimiometria mais contribuiu foi a espectroscopia molecular. São numerosas as metodologias quantitativas desenvolvidas nas últimas décadas baseadas na combinação de calibração multivariada<sup>2,3</sup> com técnicas tais como, absorção no ultravioleta (UV)/visível, no infravermelho (IV) médio ou próximo, Raman e fluorescência molecular. Uma área importante em que a combinação de quimiometria e espectroscopia molecular vem encontrando aplicação é a determinação direta de fármacos. Embora as Farmacopéias mais utilizadas, tais como a americana ("USP, United States Pharmacopoeia")<sup>4</sup>, não recomendem metodologias que envolvam quimiometria, uma revisão recente da literatura mostra um número crescente de publicações científicas propondo tais métodos<sup>5</sup>. Entretanto, a maioria das metodologias propostas para determinação de fármacos é baseada em técnicas cromatográficas e determinações univariadas.

Estas metodologias estão bem estabelecidas e são reconhecidas pelos organismos reguladores. Apesar disso, elas podem apresentar desvantagens potenciais em certas situações, tais como custo e tempo de análise relativamente altos.

A partir da década passada, o emprego de espectroscopia molecular e calibração multivariada passou a representar uma alternativa para determinação direta de fármacos. Em situações onde a determinação de um analito é dificultada ou impedida devido à presença de interferentes, o uso de calibração multivariada permite a modelagem desses interferentes junto com o analito de interesse. A idéia central consiste em substituir a separação física das espécies pela separação quimiométrica de seus sinais. As vantagens potenciais da determinação direta são rapidez, redução de uso de reagentes, baixo custo e simplificação no preparo da amostra, diminuindo as possibilidades de erro causadas pela manipulação. Nos últimos anos, as principais aplicações desta estratégia estiveram no controle de qualidade de produtos farmacêuticos e em análises clínicas. Tornaram-se possíveis, por ex., a determinação simultânea direta de vários princípios ativos em um mesmo produto farmacêutico<sup>6-12</sup> e a determinação de fármacos em amostras complexas, tais como plasma, soro e urina, sem necessidade de separação prévia dos constituintes<sup>13-18</sup>.

Dentre as novas metodologias desenvolvidas para determinação direta de fármacos, deve-se destacar a aplicação do método PARAFAC ("PARAllel FACtor analysis") no tratamento de dados espectrofluorimétricos. Um dos objetivos deste artigo é divulgar para a comunidade química brasileira o potencial deste tipo de metodologia. Com essa finalidade, serão descritas duas aplicações desenvolvidas no Laboratório de Quimiometria em Química Analítica do Instituto de Química da Unicamp, envolvendo as determinações em plasma humano de doxorubicina e salicilato, respectivamente. O outro objetivo deste artigo é descrever as bases teóricas do PARAFAC. Para isso, uma discussão dos tipos de métodos

\*e-mail: ronei@iqm.unicamp.br

quimiométricos usados em calibração multivariada e um breve histórico do desenvolvimento dos métodos de ordem superior serão previamente apresentados.

## TIPOS DE MÉTODOS QUIMIOMÉTRICOS EM FUNÇÃO DA ORDEM DOS DADOS

Os métodos de calibração usados em química analítica podem ser classificados de acordo com a dimensão dos dados analisados<sup>19</sup>. Métodos de ordem zero são usados para tratar dados univariados, gerados por instrumentos, tais como eletrodos íon-seletivos, pHmetros e colorímetros. Nestes casos, a resposta medida para cada amostra é um valor escalar (tensor de ordem zero). Estes métodos não fornecem resultados aceitáveis na presença de interferentes, pois demandam total seletividade para o analito de interesse. Métodos de calibração de primeira ordem podem ser usados para tratar dados multivariados, gerados por instrumentos, tais como espectrômetros e cromatógrafos, cujas respostas fornecem um vetor (tensor de primeira ordem) de dados para cada amostra. Nestes casos, é possível a calibração na presença de interferentes, desde que estes estejam presentes no conjunto de calibração usado para construir o modelo. Instrumentos que fornecem como resposta uma matriz (tensor de segunda ordem) de dados para cada amostra geram dados de segunda ordem. Como exemplo, podem ser citadas as chamadas técnicas hífenadas, tais como cromatografia gasosa acoplada à espectrometria de massas (GC-MS) e cromatografia líquida de alta eficiência com detecção por arranjo de diodos (HPLC-DAD). Dados de segunda ordem podem ainda ser obtidos por espectrofluorimetria, análise de injeção em fluxo com detecção espectrofotométrica ou análise de imagens, gerando uma grande quantidade de informação. Métodos de calibração de segunda ordem são aplicados a este tipo de dados, possibilitando a obtenção de uma série de vantagens sobre os outros métodos. Uma destas vantagens é a possibilidade de obtenção dos espectros (ou outros sinais, dependendo da técnica empregada) puros de cada componente do sistema, aplicando um número mínimo de restrições ao modelo. Uma outra vantagem importante é a possibilidade de calibração, sob certas condições, na presença de interferentes desconhecidos, que não estejam presentes no conjunto de calibração, a chamada “vantagem de segunda ordem”<sup>19,20</sup>. Por fim, deve-se mencionar que métodos de ordem igual ou superior a três são possíveis, embora não abundantes na literatura química. Superfícies de fluorescência obtidas em função do tempo para cada amostra são um exemplo recente de dados de terceira ordem presentes na literatura<sup>21,22</sup>. De modo geral, dados de ordem igual ou superior a dois são denominados dados de ordem superior.

### Modelos bi- e trilineares

Para o sucesso de uma aplicação, é fundamental a compatibilidade entre a natureza dos dados e o modelo assumido pelo método quimiométrico. Os métodos de primeira ordem são baseados em modelos bilineares, pois os vetores de dados coletados para as várias amostras podem ser agrupados em matrizes de dados. A maioria dos métodos de segunda ordem baseia-se em modelos trilineares, pois dados de segunda ordem coletados para várias amostras podem ser organizados em um arranjo tridimensional (na verdade, nem todos os dados de segunda ordem são trilineares, assim como nem todos os métodos de segunda ordem são trilineares; métodos para tratar desvios de trilinearidade serão mencionados mais adiante neste artigo). As matrizes de dados são colocadas lado a lado, gerando um paralelepípedo, cujas fatias são objetos ou amostras, as colunas, variáveis, e as linhas, outra condição (pH, tempo, tem-

peratura, etc.). Métodos de terceira ordem demandam modelos quadrilineares<sup>21-23</sup> e assim, sucessivamente.

O termo modelo é usado aqui com o mesmo sentido que na estatística clássica. Um modelo é uma aproximação útil para descrever um conjunto de dados reais; por ex., a matriz  $\hat{\mathbf{X}}$  é uma matriz estimada através de um modelo dos dados reais contidos na matriz  $\mathbf{X}$ . A análise de componentes principais (“PCA, Principal Component Analysis”)<sup>2</sup>, um dos métodos quimiométricos mais conhecidos, pode ser tomada como exemplo. O modelo PCA é definido por especificações, tais como base estrutural, restrições e função de ajuste. A base estrutural do modelo PCA é bilinear e dada pela seguinte expressão, a qual considera somente os autovalores mais significativos:

$$\hat{\mathbf{X}} = \mathbf{T} \mathbf{P}^t \quad (1)$$

Os parâmetros do modelo, contidos em  $\mathbf{T}$  (matriz de escores) e  $\mathbf{P}^t$  (matriz transposta de pesos), são estimados sobre certas restrições. No caso da PCA, as componentes principais devem ser ortogonais e, portanto, a restrição  $\mathbf{P}^t \mathbf{P} = \mathbf{I}$  é válida, onde  $\mathbf{I}$  é uma matriz identidade. Finalmente, a função de ajuste serve como critério objetivo para o algoritmo estimar os parâmetros do modelo. A escolha dessa função normalmente é baseada em suposições acerca dos resíduos do modelo. A grande maioria dos métodos quimiométricos, assim como a PCA, baseia-se em funções de ajuste por mínimos quadrados.

## MÉTODOS QUIMIOMÉTRICOS DE SEGUNDA ORDEM

### Histórico

Os métodos multidimensionais de tratamento de dados foram primeiramente desenvolvidos por psicometristas (pesquisadores da área de psicologia que aplicam estatística multivariada a seus dados) na década de 60, com destaque para o pioneiro L. Tucker, que propôs os “métodos de Tucker”<sup>24-26</sup>. Esse desenvolvimento coincidiu com a tendência positivista, vigente na época, de valorizar excessivamente o papel da análise quantitativa nas ciências humanas. Desta forma, grandes quantidades de dados eram obtidos na forma de questionários ou testes aplicados a um grande número de indivíduos sob diferentes condições e analisadas por psicometristas, os quais buscavam extrair componentes puros que deveriam representar influências ou padrões de comportamento.

No início dos anos 70, o pesquisador na área de lingüística R. Harshman desenvolveu um modelo que viria a encontrar aplicação na literatura química anos mais tarde, o PARAFAC<sup>27</sup>. Carrol e Chang propuseram na mesma época e de maneira independente um modelo idêntico, ao qual deram o nome de decomposição canônica (“CANDECOMP, CANonical DECOMPosition”)<sup>28</sup>. A utilização de métodos multidimensionais em química foi relativamente tardia. O artigo de Ho, Christian e Davidson, que propôs o método de análise de fatores por aniquilação do posto (“RAFA, Rank Annihilation Factor Analysis”) para tratamento de dados obtidos por espectrofluorimetria, em 1978, é provavelmente o pioneiro<sup>29</sup>. Durante toda a década de 80, a aplicação destes métodos na literatura química permaneceu relativamente restrita. Em 1981, Appellof e Davidson, baseados no artigo de Carrol e Chang<sup>28</sup>, apresentaram um modelo similar ao PARAFAC, ao qual não deram nome, para tratamento de dados cromatográficos com detecção espectrofluorimétrica<sup>30</sup>. Em uma série de três artigos publicados em 1988, Russell e co-autores usaram o mesmo modelo, ao qual denominaram PCFA (“Principal Component Factor Analysis”), para o tratamento de espectros de fluorescência molecular resolvidos no tempo<sup>31-33</sup>. Um destaque especial deve ser dado para o

método generalizado de aniquilação do posto, GRAM (“Generalized Rank Annihilation Method”) proposto em 1986 por Sanchez e Kowalski<sup>34</sup>. Este método, de solução algébrica, é baseado em um modelo trilinear e numa decomposição dos dados em autovetores, mas possui a limitação de que uma das dimensões dos dados seja igual a dois. Em 1990, os mesmos autores propuseram uma extensão do método, a decomposição direta trilinear (“DTD, Direct Trilinear Decomposition”) <sup>35</sup>, a qual superava a antiga limitação. No entanto, a inexistência de um critério de otimização bem definido (como o ajuste por mínimos quadrados) tem limitado a aplicação deste método a conjuntos de dados livres de ruído<sup>20</sup>. O DTD costuma produzir resultados que não são robustos na presença de ruído instrumental. Em 1989, Wilson e co-autores propuseram uma extensão do GRAM para dados nos quais a contribuição de um analito para o sinal total não corresponde a um sinal de posto um, conforme assumido pelo modelo trilinear do GRAM. Esta alternativa para tratar desvios de trilinearidade foi chamada de método não bilinear de aniquilação do posto (“NBRA, NonBilinear Rank Annihilation”) <sup>36</sup>. Durante os anos 90, alguns métodos que ultimamente vem ganhando destaque foram propostos, a saber, a extensão da resolução multivariada de curvas para dados de segunda ordem (“MCR, Multivariate Curve Resolution”) <sup>37</sup> e o método de regressão dos mínimos quadrados bilineares (“BLLS, BiLinear Least Squares”) <sup>38</sup>.

Em 1989, Geladi apresentou em um artigo tutorial os métodos para análise de dados multidimensionais em química e chamou à atenção para a necessidade de sistematização e generalização<sup>39</sup>. Ele corretamente identificava como fatores limitantes para a difusão destes métodos a não disponibilidade de um algoritmo na forma de um programa razoavelmente amigável e a ausência de generalização para as aplicações encontradas na literatura (nas palavras dele: “aplicações que não vão além do próprio exemplo”). Em 1992, Smilde também chamou a atenção para o potencial de alguns métodos de ordem superior, então ainda pouco usados e conhecidos na literatura química<sup>40</sup>. De 1995 a 1998, ele iria orientar a tese de doutorado de Bro<sup>20</sup>, a qual veio dar uma contribuição fundamental para difundir a aplicação de alguns novos métodos quimiométricos, o PARAFAC<sup>41</sup> e o N-PLS<sup>42</sup>. Esta tese, que gerou mais de uma dezena de artigos científicos, descreve de modo bastante claro e didático os modelos e algoritmos para uma série de métodos multidimensionais e apresenta exemplos de aplicações em diferentes áreas, discutindo suas generalizações. A contribuição talvez mais importante desse trabalho foi o fato de o autor disponibilizar gratuitamente os algoritmos usados, na forma de pacotes (“toolboxes”) <sup>43,44</sup> elaborados em ambiente Matlab (Mathworks), facilitando a difusão do uso desses métodos.

## Modelos de Tucker

Tanto do ponto de vista histórico quanto do didático, a compreensão dos modelos de Tucker precede o desenvolvimento do PARAFAC. Por este motivo, eles serão brevemente descritos a seguir. Durante os anos 60, Tucker desenvolveu os modelos atualmente conhecidos como Tucker1, Tucker2 e Tucker3<sup>24,25</sup>. O modelo Tucker1 corresponde ao simples desdobramento (“unfolding”) do arranjo de dados tridimensional  $\underline{\mathbf{X}}$  (o sublinhado indica uma matriz tridimensional) em uma matriz bidimensional, seguido de uma decomposição bilinear usando PCA. O desdobramento de um arranjo de dados de dimensões  $I \times J \times K$  em uma matriz de dimensões  $I \times JK$  é representado na Figura 1. Este desdobramento também pode ser feito nas outras duas direções ( $J \times KI$  ou  $K \times IJ$ ). O Tucker1, também conhecido como “unfolded”-PCA, ignora a estrutura trilinear dos dados, decompondo-os com um método bilinear. Tal fato leva a modelos que são mais complexos (maior número de graus de liber-

dade), menos robustos, menos interpretáveis e de menor capacidade de previsão que os modelos estritamente trilineares<sup>20</sup>. O modelo Tucker2 é um caso particular do Tucker3, em que uma das dimensões é mantida fixa durante a decomposição.

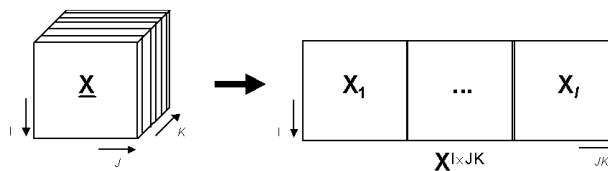


Figura 1. Representação do desdobramento (“unfolding”) de um arranjo de dados tridimensional em uma matriz. Modelo Tucker1

O mais importante dos modelos de Tucker é o Tucker3, cuja base estrutural é dada pela Equação 2

$$\underline{\mathbf{X}} = \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^t + \underline{\mathbf{E}} \quad (2)$$

onde as matrizes  $\mathbf{A}$  ( $I \times D$ ),  $\mathbf{B}$  ( $J \times E$ ) e  $\mathbf{C}$  ( $K \times F$ ) contêm os pesos (“loadings”) do modelo relativos às três dimensões dos dados, respectivamente; a matriz  $\mathbf{G}$  ( $D \times E \times F$ ) é a matriz central ou conectora (“core matrix”); o tensor  $\underline{\mathbf{E}}$  ( $I \times J \times K$ ) contêm os erros do modelo e o símbolo “ $\otimes$ ” representa o produto de Kronecker<sup>45</sup> (Apêndice 1). D, E e F indicam o número de fatores nas três dimensões dos dados, respectivamente. Uma característica particular do modelo Tucker3 é que ele permite que o número de fatores decompostos seja diferente em cada dimensão. Os elementos da matriz  $\mathbf{G}$ , produto do rearranjo do tensor  $\underline{\mathbf{G}}$  ( $D \times E \times F$ ), definem como os pesos individuais interagem nas diferentes dimensões, ou seja, os seus valores absolutos indicam a importância de cada interação entre fatores para o modelo (um valor próximo de zero indica uma interação insignificante). A restrição dos pesos em  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  serem ortogonais é usual. A representação gráfica da decomposição dos dados efetuada pelo método Tucker3 é apresentada na Figura 2.

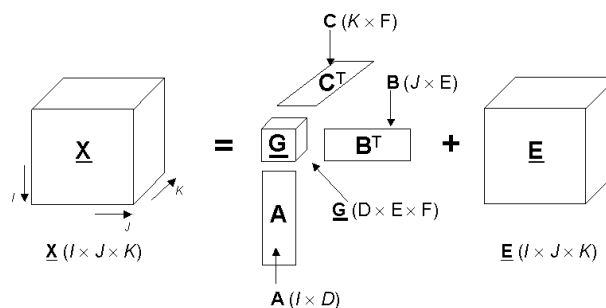


Figura 2. Representação gráfica do modelo Tucker3

## PARAFAC

A análise de fatores paralelos, PARAFAC<sup>20,41</sup>, é um método para decomposição de dados de ordem superior, cuja base estrutural é dada pelas matrizes de pesos  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$ , que contêm os elementos  $a_{if}$ ,  $b_{jf}$  e  $c_{kf}$ , respectivamente, relativos às três dimensões dos dados. O modelo trilinear é ajustado para minimizar a soma dos quadrados dos resíduos  $e_{ijk}$ , de acordo com a Equação 3

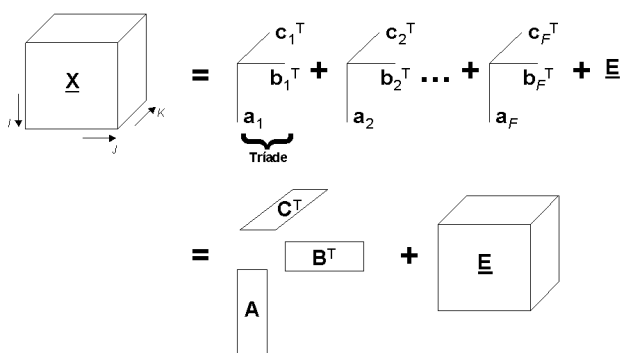
$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (3)$$

onde F é o número de fatores (repare que aqui, ao contrário do Tucker3, este número é igual em todas as dimensões). A Figura 3 mostra a representação gráfica da Equação 3, indicando a decom-

posição de um arranjo de dados tridimensional em  $F$  tríades de vetores pesos. Cada tríade equivale a um fator do modelo PARAFAC. Alternativamente, o modelo PARAFAC pode ser representado pela seguinte equação matricial

$$\underline{\mathbf{X}} = \mathbf{A}(\mathbf{C}|\otimes|\mathbf{B})^t + \mathbf{E} \quad (4)$$

onde as matrizes  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  têm dimensões  $I \times F$ ,  $J \times F$  e  $K \times F$ , respectivamente, e o símbolo “ $|\otimes|$ ” representa o produto de Khatri-Rao<sup>20,46</sup> (Apêndice 2).



**Figura 3.** Representação gráfica do modelo PARAFAC. Decomposição de um arranjo de dados tridimensional em  $F$  tríades de vetores peso

O PARAFAC pode ser considerado uma generalização da PCA para dados multidimensionais, ou ainda, uma versão restrita do método Tucker<sup>3,20,26</sup> com uma matriz central superidentidade ( $g_{def} = 1$  quando  $d = e = f$  e  $g_{def} = 0$  nas outras posições, ou seja, o primeiro fator de uma dimensão só se relaciona com os primeiros fatores das outras e assim, sucessivamente). Em comparação com a PCA e o Tucker3, ele é menos flexível e usa menos graus de liberdade. Por ser um modelo mais restrito, o PARAFAC é adequado para a modelagem de sistemas nos quais os fatores subjacentes estão associados a leis físico-químicas claramente definidas (usa-se o termo “hard modelling” para este tipo de modelagem, em oposição a “soft modelling”, quando a interpretação dos fatores não está restrita a leis físicas tão rígidas, como por ex., na determinação de fontes de poluição em análises ambientais). Qualquer conjunto de dados que puder ser modelado adequadamente com o PARAFAC, também poderá ser modelado com o Tucker3 ou com a PCA, mas usando um maior número de graus de liberdade<sup>41</sup>. Esse excesso de graus de liberdade tende a ser usado para modelar ruído ou variância sistemática de uma maneira redundante. Daí, as limitações da aplicação destes dois métodos a problemas de resolução de curvas. Devido ao menor número de graus de liberdade usados, o PARAFAC pode ser considerado um método mais simples e Bro<sup>41</sup> relaciona as vantagens de seu uso às idéias do filósofo do século XIV Guilherme de Ockham e seu princípio da parcimônia (“Ockham’s Razor”)<sup>47</sup>: “deve-se usar sempre o modelo mais simples possível”. Além disso, ao contrário dos outros métodos citados anteriormente, o PARAFAC fornece solução única, independente de rotações. Matematicamente, isto significa que o modelo estimado não pode sofrer rotação sem perda de ajuste. Esta última propriedade (em inglês ela é denominada “uniqueness”) é de grande importância para sua utilização na modelagem de dados espectroscópicos. Graças a ela, será possível recuperar os espectros (ou o que quer que constitua o perfil das variáveis) dos componentes puros do sistema, contanto que os dados sejam realmente trilineares, o número correto de fatores do modelo seja escolhido e a razão sinal/ruído seja apropriada.

A escolha do número de fatores é uma etapa crucial da análise, não existindo para ela um critério absoluto. A escolha pode ser feita baseada na variância explicada pelo modelo, no conhecimento químico do sistema, em métodos de reamostragem e validação cruzada (“split-half”)<sup>48</sup> ou no recentemente proposto teste de consistência trilinear (“CORCONDIA, CORE CONSistency DIAgnostic”)<sup>49</sup>. O CORCONDIA é baseado na interpretação do PARAFAC como um modelo Tucker3 restrito. O Tucker3 é aplicado aos parâmetros do modelo PARAFAC estimado e, se houver consistência trilinear, espera-se que os elementos da superdiagonal do tensor central ( $\mathbf{G}$ ) tenham valores próximos de 1 e os demais elementos valores próximos de zero. A definição matemática é expressa pela seguinte equação

$$\text{CORCONDIA} = 100 * \left( 1 - \frac{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F (g_{def} - h_{def})^2}{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F h_{def}^2} \right) \quad (5)$$

onde  $g_{def}$  é o elemento da matriz central calculada com o Tucker3 a partir dos pesos do PARAFAC,  $h_{def}$  é o elemento de um tensor binário contendo valores um na superdiagonal e zero nas demais posições e  $F$  é o número de fatores do modelo. Um valor de CORCONDIA superior a 90% indica a adequação do modelo PARAFAC, um valor em torno de 50% indica deficiência de trilinearidade e valores próximos de zero ou negativos indicam inconsistência trilinear. O CORCONDIA é uma ferramenta para detectar sobre-ajuste em um modelo PARAFAC, mas não sub-ajuste. O valor de CORCONDIA para um modelo estimado com um número de componentes menor que o real estará próximo a 100%, pois a estrutura de tal modelo será trilinear. Portanto, o modelo adequado será aquele com maior número de componentes e que preserve a estrutura trilinear. Outro aspecto a ser ressaltado no CORCONDIA é o de não assumir suposições a priori a respeito dos resíduos, pois é a parte sistemática dos dados, e não a aleatória, que é usada na avaliação do modelo. Bro e Kiers observam, ainda, que a validação de um modelo PARAFAC não deve se basear unicamente no CORCONDIA<sup>49</sup>. Outra importante diferença em relação à PCA, é que a solução do PARAFAC não é cumulativa (“not nested”). Isto significa que os parâmetros de um modelo com  $F + 1$  fatores não são iguais aos de um modelo com  $F$  fatores mais um fator adicional. A razão para isso é que, na solução do PARAFAC, os componentes não necessitam ser ortogonais (não correlacionados) entre si. Uma outra característica importante do PARAFAC é a possibilidade do uso de restrições (“constraints”). O ajuste de um modelo restrito sempre é menor que o de um modelo não restrito (menor variância explicada). No entanto, a perda de ajuste é compensada pela obtenção de um modelo mais facilmente interpretável e realista. As restrições mais comuns são não negatividade, ortogonalidade e unimodalidade. A restrição de não negatividade é usualmente adotada na resolução de espectros, já que parâmetros espectrais negativos não possuem significado físico.

O algoritmo comumente usado para a resolução do PARAFAC é o dos mínimos quadrados alternados (“ALS, Alternating Least Squares”), que sucessivamente estima os valores de uma das matrizes de pesos a partir dos valores conhecidos das outras duas<sup>20</sup>. Outros algoritmos foram ainda propostos, mas todos eles levam a modelos de qualidade inferior aos estimados pelo ALS, conforme constatado em um artigo recente<sup>50</sup> de revisão. O ALS pode ser iniciado a partir de valores aleatórios ou usando algum tipo de decomposição dos dados. O algoritmo converge iterativamente até atingir um critério de convergência ou um número de iterações previamente definidos. Os passos do algoritmo para um modelo



PARAFAC são resumidos a seguir<sup>41</sup>:

- (0) decida o número de fatores,  $F$ ;
- (1) inicie  $\mathbf{B}$  e  $\mathbf{C}$ ;
- (2) estime  $\mathbf{A}$  por um ajuste de mínimos quadrados a partir de  $\mathbf{X}$ ,  $\mathbf{B}$  e  $\mathbf{C} \Rightarrow \mathbf{A} = \mathbf{XZ}(\mathbf{ZZ}^t)^{-1}$ , onde  $\mathbf{Z} = (\mathbf{C} \otimes \mathbf{I})\mathbf{B}^t$ ;
- (3) estime  $\mathbf{B}$  de maneira análoga;
- (4) estime  $\mathbf{C}$  de maneira análoga;
- (5) retorne ao passo (2) até a convergência (pequena mudança no ajuste).

Os principais objetivos das aplicações de PARAFAC em química<sup>20</sup> são a resolução de curvas e a calibração multivariada, embora aplicações em análise exploratória<sup>51,52</sup> e em análise de variância<sup>20,53</sup> também sejam encontradas na literatura. O uso do PARAFAC para calibração segue uma estratégia análoga à da regressão em componentes principais. As amostras de concentrações conhecidas e desconhecidas são decompostas simultaneamente. Os pesos especificamente obtidos para o componente de interesse na amostra desconhecida, também chamados de escores, são relacionados com as concentrações conhecidas do conjunto de calibração através de uma regressão linear. Um último aspecto a ser comentado diz respeito ao pré-processamento dos dados. O pré-processamento de dados multidimensionais é mais difícil que o de dados bidimensionais<sup>20,54</sup>. Por ex., ao centrar na média os dados em uma das dimensões, o escalamento feito em qualquer outra dimensão é afetado. Dessa forma, não se costuma efetuar nenhum pré-processamento quando se aplica o PARAFAC na resolução de espectros.

## PARAFAC E DADOS ESPECTROFLUORIMÉTRICOS

A espectrofluorimetria<sup>55</sup> tem sido muito usada como técnica quantitativa em áreas, tais como química, medicina, meio ambiente e ciência de alimentos. Entretanto, a análise de misturas multicomponentes que apresentam espécies com sinais sobrepostos é bastante limitada quando as medidas são feitas em um único comprimento de onda de emissão/excitação. A aplicação de espectrofluorimetria em análises clínicas é dificultada pela complexidade de matrizes, tais como sangue e urina, que apresentam uma grande variedade natural de compostos fluorescentes, os quais frequentemente possuem espectros que se sobrepõem ao sinal da espécie de interesse. Esta situação exige o uso de tediosos processos de separação para que a determinação dessa espécie seja possível.

Quando a fluorescência de uma amostra é medida em diversos comprimentos de onda de emissão para diversos comprimentos de onda de excitação, obtém-se uma matriz de excitação-emissão. Quando um conjunto de amostras é medido nessas mesmas condições, obtém-se um arranjo de dados multidimensional, o qual deve apresentar comportamento trilinear, a despeito da presença de ruído, espalhamentos (Rayleigh e Raman) e outros fenômenos. Este comportamento torna os métodos quimiométricos de segunda ordem, tais como o PARAFAC, bastante adequados para tratamento de dados espectrofluorimétricos, pois o modelo matemático assumido por eles é condizente com o modelo físico-químico que gera os dados. Recordando a base estrutural do modelo PARAFAC, na Equação 3:  $x_{ijk}$  representará a intensidade de fluorescência medida para a  $i$ -ésima amostra nos comprimentos de onda de excitação  $j$  e de emissão  $k$ ;  $a_{if}$  será a concentração do  $f$ -ésimo fluoróforo na amostra  $i$ ;  $b_{jf}$  será o coeficiente de absorção molar do  $f$ -ésimo fluoróforo no comprimento de onda de excitação  $j$ ;  $c_{kf}$  será o coeficiente de emissão relativa do  $f$ -ésimo fluoróforo no comprimento de onda de emissão  $k$ ; e  $e_{ijk}$  deverá representar os resíduos, contendo a variação não capturada pelo modelo. Pelo fato de gerar facilmente da-

dos de estrutura trilinear, a espectrofluorimetria tem sido objeto do maior número de aplicações do PARAFAC encontradas na literatura, tanto em calibração quanto na resolução de espectros. Dentre essas aplicações, podem ser citadas a determinação de hidrocarbonetos policíclicos aromáticos e pesticidas<sup>56-59</sup>, a estimativa do conteúdo de dioxinas em óleo de peixe<sup>60</sup>, o estudo de parâmetros de qualidade e do processo de produção de açúcar<sup>61,62</sup>, a resolução dos espectros das diferentes formas do fármaco anti-inflamatório piroxicam<sup>63</sup> e dos produtos de degradação da clorofila<sup>64</sup>. Um artigo de revisão recente discute uma série de aspectos práticos sobre este tipo de análise, dentre eles, a escolha do número correto de fatores, o uso de valores ausentes ("missing values") na correção da trilinearidade das regiões espectrais onde ocorre espalhamento e a detecção de amostras anômalas ("outliers")<sup>65</sup>.

Nos últimos anos, a combinação de PARAFAC e espectrofluorimetria abriu a possibilidade de determinações diretas de fármacos em amostras biológicas complexas, sem a necessidade de uma etapa de extração de proteínas, comum nas análises tradicionais. As proteínas presentes no sangue costumam interagir com os fármacos, exercendo um forte efeito supressor ("quenching") sobre o sinal de fluorescência. A partir de 2002, surgiram os primeiros artigos na literatura, relatando o uso da combinação PARAFAC/espectrofluorimetria na determinação de fármacos, tais como, ácido salicílico e naproxen<sup>14,15</sup>, piroxicam<sup>16</sup>, antibióticos fluorquinolonas<sup>17</sup>, doxorrubicina<sup>66</sup>, carbamazepina<sup>67</sup> e ciprofloxacina<sup>18</sup>, em matrizes tais como plasma, soro ou urina. Em particular, a determinação direta de um fármaco sem a necessidade de uma etapa de diluição da matriz foi possível em apenas uma dessas aplicações<sup>66</sup>.

Com o objetivo de demonstrar o potencial desse tipo de metodologia, duas aplicações envolvendo a determinação direta de fármacos em plasma humano são descritas a seguir. Nelas, duas estratégias diferentes de análise foram empregadas. Na primeira aplicação, foi feita a determinação de doxorrubicina através de uma curva de calibração. Na segunda, foi feita a determinação de salicilato usando o método da adição padrão. Em ambas as aplicações, os espectros foram obtidos em um espectrofluorímetro Perkin Elmer LS55; os dados foram tratados no programa *MATLAB*<sup>7M</sup>, versão 6.1 (*The MathWorks*, Natick: EUA), usando o pacote "The N-way Toolbox for MATLAB", versão 2.00<sup>43,44</sup>; as amostras de plasma humano de voluntários saudáveis foram obtidas junto ao Hemocentro da Unicamp; todas as amostras analisadas foram preparadas pela contaminação artificial do plasma com a respectiva droga, em concentrações conhecidas.

## DETERMINAÇÃO DE DOXORRUBICINA USANDO UMA CURVA DE CALIBRAÇÃO<sup>66,68</sup>

A doxorrubicina (DXR) é um antibiótico citotóxico antracíclico que apresenta atividade em uma larga faixa de neoplasmas humanos, incluindo uma grande variedade de tumores sólidos<sup>69</sup>. Ela é uma molécula anfótera, com um núcleo naftacenediona, ligado através do carbono C7 a uma cadeia hidrofílica aminoglicosídica. O interesse na determinação de antraciclina em fluidos biológicos é devido a seus efeitos tóxicos, comumente observados após tratamento prolongado<sup>70</sup>. Nos últimos anos, várias metodologias foram desenvolvidas para determinação de DXR nesse tipo de matriz, as quais se basearam em técnicas como HPLC<sup>71,72</sup>, eletroforese capilar<sup>73</sup> e espectroscopia Raman<sup>74</sup>. A maioria dessas metodologias necessita de uma etapa prévia de extração, de modo que o desenvolvimento de um método para determinação de DXR em fluidos biológicos, que seja não invasivo, rápido e de baixo custo, pode ser vantajoso. Como a DXR apresenta fluorescência natural, a

espectrofluorimetria pode ser considerada uma técnica de alta sensibilidade para sua determinação. Entretanto, conforme discutido na seção anterior, a presença de interferentes no plasma limita o uso desta técnica. Desta forma, a obtenção de dados de fluorescência de segunda ordem e o uso do PARAFAC são apresentados como alternativa para esta determinação.

Considerando a faixa de concentração de DXR de interesse clínico, prepararam-se dez amostras entre 0,75 e 11,25  $\mu\text{g mL}^{-1}$ . O plasma de um indivíduo diferente foi usado para cada nível de concentração. As superfícies espectrais foram obtidas nas faixas de excitação de 390 a 500 nm (passo 2 nm) e de emissão de 510 a 650 nm (passo 0,5 nm). Foi montado um arranjo de dados cúbico com dimensões 9 x 56 x 282 (amostras x comprimentos de onda de excitação x comprimentos de onda de emissão).

Na faixa de excitação entre 490 e 500 nm e de emissão entre 510 e 520 nm, observou-se a presença de espalhamento Rayleigh. A região espectral onde ocorre este fenômeno não apresenta um caráter trilinear, ao contrário do restante da superfície de dados e, portanto, não é compatível com o modelo PARAFAC<sup>20,65</sup>. Deste modo, os valores registrados nesta região foram substituídos por valores ausentes, ou seja, sem informação espectral. O algoritmo usado neste trabalho<sup>43,44</sup> substitui inicialmente os valores ausentes por valores aleatórios. Após cada interação, um novo modelo é desenvolvido e esses valores são trocados por novas estimativas, até que ocorra convergência. É importante salientar que os elementos dessa região não podem ser substituídos por zeros, pois isto implicaria na perda de trilinearidade, levando a uma certa instabilidade no modelo.

O melhor modelo PARAFAC foi obtido sem restrições e com dois fatores, explicando 98,75% da variância total dos dados. A escolha do número de fatores baseou-se em dois critérios principais, o CORCONDIA<sup>49</sup> (Equação 5) e um procedimento de reamostragem<sup>48</sup>. Um modelo com dois fatores apresentou 100% de CORCONDIA, indicando perfeita trilinearidade; por outro lado, um modelo com três fatores apresentou 0,83% de CORCONDIA, demonstrando sua inconsistência. Um procedimento de reamostragem ("split-half"), no qual o arranjo de dados original foi dividido em cem sub-arranjos, corroborou a escolha de dois fatores, pois um número maior de fatores levou a modelos divergentes entre si. Os pesos relativos às dimensões de excitação e de emissão do modelo PARAFAC escolhido são apresentados na Figura 4. O primeiro fator foi atribuído à DXR, por comparação com os seus espectros de excitação e emissão, obtidos experimentalmente em água. O segundo fator foi associado a um interferente presente no plasma. Este interferente pode ser atribuído à bilirrubina, à riboflavina (vitamina B<sub>2</sub>) ou a uma mistura de ambas, com base na comparação com a literatura<sup>75</sup>.

Os pesos do primeiro fator na dimensão das concentrações foram usados para estabelecer uma relação linear com a concentração de DXR, gerando um modelo de regressão univariado. Um procedimento de validação cruzada ("leave-one-out") das amostras foi realizado para verificar a viabilidade do modelo de calibração. Neste procedimento, dez modelos foram desenvolvidos com uma amostra de previsão diferente retirada do conjunto de cada vez. Curvas de calibração com coeficientes de regressão superiores a 0,95 foram obtidas. Uma delas é mostrada na Figura 5. A raiz quadrada dos erros médios quadrados de validação foi igual a 0,060  $\mu\text{g mL}^{-1}$  e o erro relativo médio de previsão foi de 4,4%. Estes resultados indicaram que foi possível quantificar DXR em plasma humano, através dos pesos de um modelo PARAFAC. Este método possui potencial para ser aprimorado e então ser empregado em aplicações reais, tais como investigações farmacocinéticas em pacientes com câncer e

submetidos a tratamento com DXR ou outro tipo de antraciclina.

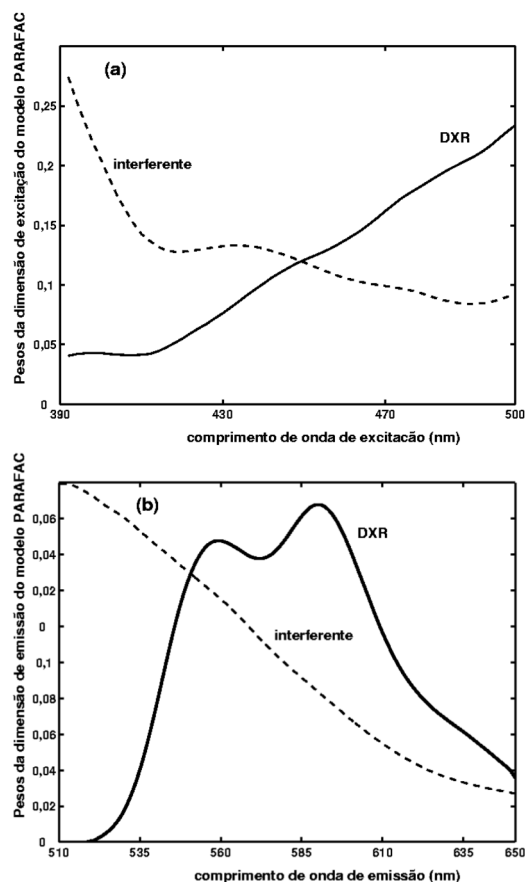


Figura 4. Espectros de fluorescência deconvoluídos, obtidos através dos pesos do modelo PARAFAC construído para a determinação de DXR. (a) excitação e (b) emissão

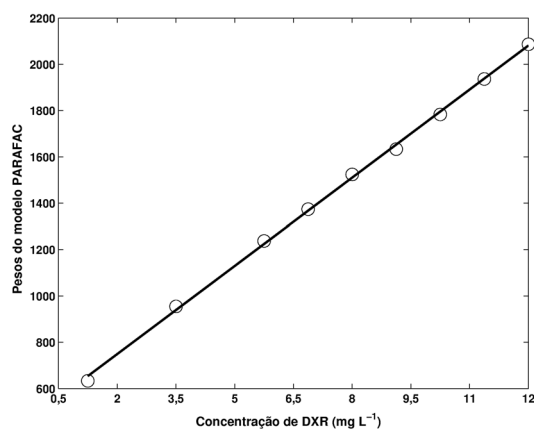


Figura 5. Curva de calibração para determinação de DXR em plasma humano, obtida a partir dos pesos de um modelo PARAFAC

#### DETERMINAÇÃO DE SALICILATO USANDO O MÉTODO DA ADIÇÃO PADRÃO<sup>76,77</sup>

O ácido salicílico (AS)<sup>78</sup> é o produto da hidrólise e o principal metabólito da aspirina. No sangue humano (pH  $\approx$  7,4), ele se apresenta em sua forma desprotonada, o íon salicilato. O AS é ainda usado em pomadas, por seu efeito analgésico e anti-séptico. O interesse na determinação de AS em plasma humano está ligado a

casos de ingestão de grande quantidade de aspirina, normalmente por crianças, e a intoxicações pelo uso de pomadas<sup>79</sup>. A meia vida biológica do AS, entre 3 e 19 h, é bastante superior à meia vida do AAS, entre 15 e 20 min. Voluntários saudáveis, que ingeriram uma dose de 500 mg de aspirina apresentaram uma concentração máxima de 30  $\mu\text{g mL}^{-1}$  nas primeiras 3 a 4 h<sup>80</sup>. A determinação de AS em amostras biológicas tem sido feita por volumetria (método de Trinder)<sup>81</sup>, por HPLC<sup>82</sup>, por eletroforese capilar<sup>83</sup> e usando um biossensor amperométrico<sup>84</sup>. A maioria desses métodos exige etapas de extração e precipitação de proteínas. Outras determinações foram baseadas em espectros de fluorescência obtidos em um comprimento de onda de excitação fixo e métodos quimiométricos de primeira ordem (PLS)<sup>13,85</sup>. Porém, estas determinações também exigiram etapas de extração com solvente. O objetivo desta aplicação foi determinar diretamente salicilato em plasma humano, usando um método que simplifique a manipulação das amostras. Essa determinação em plasma puro, na presença de interferentes e de um forte efeito de matriz, foi possibilitada pela utilização de uma nova estratégia, a qual incluiu, além do PARAFAC, o uso do método da adição padrão.

### Método da adição padrão e vantagem de segunda ordem

O método da adição padrão, bastante conhecido em química analítica<sup>86</sup>, é aplicado a dados univariados como um meio de superar o efeito de matriz, que afeta a maneira pela qual o instrumento responde ao analito. Em comparação com um processo de calibração externa, esse método apresenta a desvantagem de consumir maior tempo de análise, em função da necessidade de várias adições do padrão. No entanto, ele surge como alternativa em situações em que a calibração externa é inviável. O método da adição padrão requer duas condições para o sucesso da calibração: (1) o instrumento deve responder linearmente ao aumento da concentração do analito; (2) quando a concentração do analito for zero, a resposta instrumental também deve ser igual a zero. A estimativa da concentração do analito na amostra é obtida através de um gráfico da resposta instrumental em função da quantidade de padrão adicionada, usando o valor da interseção da reta ajustada aos dados com o eixo das abscissas. Em 1979, Saxberg e Kowalski propuseram uma extensão desse método para dados multivariados, o método da adição padrão generalizada ("GSAM, Generalized Standard Addition Method")<sup>87</sup>. O GSAM requer que tanto o analito quanto os interferentes sejam adicionados à amostra sequencialmente. Isto relaxa a restrição de que o método analítico seja inteiramente seletivo para o analito de interesse. Entretanto, resultados adequados não são obtidos se houver alguma fonte de sinal instrumental não calibrada. Portanto, na ausência de todas as espécies incluídas no modelo de calibração, a resposta instrumental deverá ser zero em todos os canais. Em 1995, Booksh e co-autores propuseram outra extensão do método da adição padrão, desta vez para dados multidimensionais, o método da adição padrão de segunda ordem ("SOSAM, Second-Order Standard Addition Method")<sup>88</sup>. Este método usa a decomposição direta trilinear (DTD)<sup>35</sup> como meio de decompor os dados e pode ser resumido em três etapas: (1) o DTD é aplicado na decomposição de um arranjo de dados de terceira ordem, formado pelo agrupamento dos dados de segunda ordem da amostra e de cada adição sucessiva. O número de fatores usados na decomposição deverá corresponder ao número de analitos mais o de interferentes; (2) a matriz de pesos da dimensão correspondente à composição da amostra deverá conter em suas colunas a informação relativa às concentrações do analito e dos interferentes; a identificação de qual coluna corresponde ao analito de interesse pode ser feita pela comparação dos pesos da dimensão dos espec-

tros (ou de outro tipo de sinal) com o espectro do analito puro; (3) os valores da coluna identificada, correspondentes às concentrações do analito na amostra e após as adições sucessivas, serão usados em uma regressão, da mesma maneira que em uma adição padrão univariada.

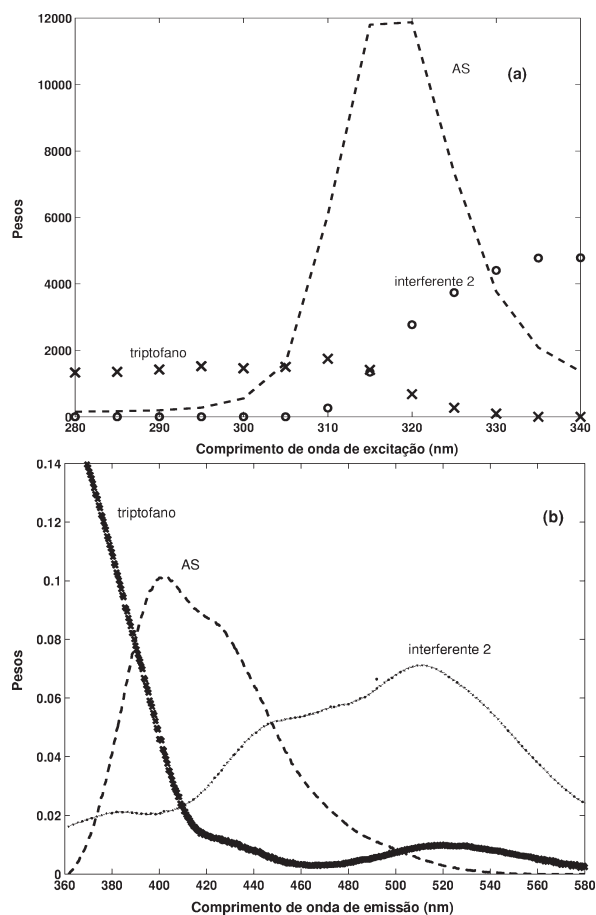
Um dos principais aspectos a ser ressaltado no SOSAM é a utilização da "vantagem de segunda ordem", definida como a habilidade de realizar uma determinação na presença de interferentes desconhecidos ou não incluídos no modelo de calibração<sup>19,20</sup>. Para que a vantagem de segunda ordem ocorra, duas condições devem ser satisfeitas: (1) os dados devem ser, no mínimo, trilineares ou aproximadamente trilineares; (2) o método usado deve ser de segunda ordem e decompor as amostras a serem previstas simultaneamente com as amostras de calibração. O uso de um método de segunda ordem é condição necessária, mas não suficiente, para que haja vantagem de segunda ordem. O N-PLS<sup>42</sup>, por ex., é um método de segunda ordem, mas não propicia vantagem de segunda ordem, pois realiza primeiro a decomposição do conjunto de calibração para depois projetar as amostras a serem previstas. Dentre os métodos que permitem o uso da vantagem de segunda ordem, estão o PARAFAC, o DTD e o GRAM<sup>34</sup>, os quais satisfazem a segunda condição mencionada acima. A vantagem de segunda ordem pode ser obtida ainda pela combinação dos métodos MCR<sup>37</sup> ou BLS<sup>38</sup> com um procedimento conhecido como bilinearização residual ("RBL, Residual BiLinearization")<sup>18</sup>.

Apesar de seu potencial, o número de artigos publicados usando o SOSAM é escasso. Herrero *et al.* usaram-no em dados espectro-eletroanalíticos<sup>89</sup>. Wu e co-autores usaram uma variação do SOSAM, na qual a decomposição dos dados é feita por um algoritmo desenvolvido por eles próprios, para determinar benzo-clorados por HPLC-DAD<sup>90</sup>. Mais recentemente, Comas *et al.* combinaram o método da adição padrão com o GRAM na determinação de sulfonatos aromáticos em água<sup>91</sup> e hidrocarbonetos policíclicos aromáticos em sedimentos marinhos<sup>92</sup>, ambos por HPLC-DAD. Como o GRAM possui a limitação de que uma das dimensões dos dados seja igual a dois, estas duas últimas aplicações foram limitadas a apenas uma adição em cada amostra. Na presente aplicação, o PARAFAC será empregado na decomposição dos dados, ao invés do DTD, originalmente proposto no SOSAM. Na comparação entre esses dois métodos, o PARAFAC apresenta a vantagem de ser mais robusto à presença de ruído nos dados, o que muitas vezes impede a aplicação do DTD a dados reais<sup>20</sup>.

### Estudos preliminares: resolução dos espectros e tentativa de determinar AS através de uma curva de calibração

A idéia inicial desta aplicação era determinar AS através de uma curva de calibração preparada com amostras de plasmas de indivíduos diferentes, seguindo a estratégia adotada na determinação de DXR. Em um estudo preliminar, foram preparadas trinta soluções de AS em plasma, na faixa de 1,5 a 30  $\mu\text{g mL}^{-1}$ , e obtidas suas superfícies de fluorescência, nas faixas de excitação de 280 a 340 nm (passo 5 nm) e de emissão de 360 a 580 nm (passo 0,5 nm). O PARAFAC foi aplicado, sob restrição de não negatividade nas dimensões espectrais, a um arranjo de dados tridimensional formado por 442 comprimentos de onda de emissão, 13 comprimentos de onda de excitação e 30 amostras. A decomposição desses dados produziu um modelo que, com três fatores, explicou 99,2% da variância total. A consistência trilinear desse modelo foi constatada pela obtenção de um valor de CORCONDIA de 95,3%. Os pesos desse modelo, relativos às dimensões espectrais de excitação e de emissão, são mostrados na Figura 6. Com base nessa figura, os fatores modelados foram associados ao AS e a dois possíveis

interferentes presentes no plasma (“background”). A atribuição do AS baseou-se na semelhança dos pesos do primeiro fator com um espectro de AS obtido em solução aquosa tamponada em pH 7,4 (pH aproximado do sangue humano). A tentativa de identificação dos interferentes modelados foi feita com base no artigo de Wolfbeis e Leiner, que mapearam a fluorescência total do soro humano nas regiões do UV e do visível<sup>75</sup>. O interferente 1 foi associado ao triptofano, enquanto a caracterização do interferente 2 foi mais difícil, sendo sugeridas as seguintes possibilidades: riboflavina, livre ou na forma de suas coenzimas, e nicotinamida adenina dinucleotídeo (NADH) ou seu fosfato.



**Figura 6.** Espectros de fluorescência deconvoluídos, obtidos através dos pesos do modelo PARAFAC construído para determinação de AS. (a) excitação e (b) emissão

Os pesos do modelo PARAFAC relativos à dimensão das amostras poderiam ser usados para calibração, como na determinação de DXR. Contudo, os pesos deste modelo não apresentaram um comportamento linear em função da concentração de AS, impossibilitando a construção de uma curva de calibração. Além disso, verificou-se que replicatas de mesma concentração de AS em diferentes plasmas mostravam uma diferença de até 30% nos valores de seus pesos, enquanto amostras de mesma concentração no mesmo plasma mostravam diferenças insignificantes. Portanto, concluiu-se não ser possível construir um modelo para calibração de AS em diferentes plasmas, pois o efeito da supressão (“quenching”) do seu sinal variou de acordo com a composição do plasma de cada indivíduo. Esta situação foi diferente da observada na determinação de DXR, indicando que a mudança espectral causada pela interação com as proteínas do plasma é mais fraca no caso da DXR

que no caso do AS. Como um forte efeito de matriz foi observado e, obviamente, não é viável construir uma curva de calibração específica para cada indivíduo, optou-se por outra estratégia.

### Determinação de AS em plasma usando o método da adição padrão

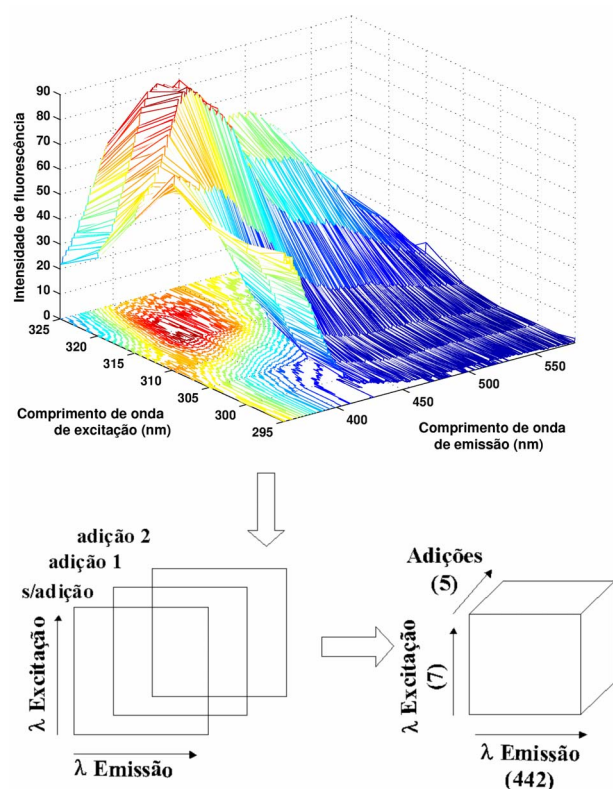
Optou-se por usar o método da adição padrão, indicado para situações como esta em que o efeito de matriz inviabiliza a construção de uma curva de calibração robusta. Um dos aspectos a ser destacado nesta metodologia é o fato de que a curva de adição padrão foi construída usando os pesos de um modelo PARAFAC, ao invés do sinal instrumental puro. Esses pesos equivaleram ao sinal “filtrado”, livre do sinal dos interferentes. As determinações foram feitas em uma situação de volume total variável em função da adição do padrão<sup>93</sup>.

Nesta nova análise, dez amostras foram preparadas da mesma forma que na análise preliminar, na faixa de 1,5 a 24,0  $\mu\text{g mL}^{-1}$ . O plasma de um indivíduo diferente foi usado para cada nível de concentração. Em cada medida, 2,5 mL de amostra foram adicionados à cubeta. Em seguida, foram feitas e medidas quatro adições sucessivas de 50  $\mu\text{L}$  de uma solução 120  $\mu\text{g mL}^{-1}$  de AS. Todas as análises foram feitas em triplicata. A faixa de comprimento de onda de excitação foi restrita de 295 a 325 nm, não acarretando alteração significativa nos modelos. Os modelos PARAFAC foram construídos a partir de arranjos de dados formados por 442 comprimentos de onda de emissão, 7 comprimentos de onda de excitação e 5 medidas (a amostra original mais quatro adições), conforme representado na Figura 7. Foi construído um modelo específico para cada triplicata de cada amostra, totalizando trinta modelos. Em todos os casos, eles foram construídos com três fatores, sem restrições, e explicaram mais que 99,90% da variância total. Observou-se que modelos sem restrições produziram resultados com erros menores em relação aos que foram construídos com restrição de não negatividade nas duas dimensões espectrais. Os valores de CORCONDIA obtidos estiveram entre 25 e 80%, indicando, em vários casos, deficiência de trilinearidade. Modelos com dois fatores apresentaram quase 100% de trilinearidade, mas forneceram erros de previsão muito superiores aos modelos com três fatores. Além disso, eles apresentaram variância residual sistemática e significativa, indicando a existência de mais um fator a ser extraído. Em todos os casos, os pesos obtidos em relação aos espectros de emissão e excitação apresentaram boa similaridade com os pesos mostrados na Figura 6. Os resultados das determinações das dez amostras (estimativas da média e do desvio padrão das triplicatas) são mostrados na Tabela 1, juntamente com os erros em relação aos valores esperados. Na grande maioria dos casos, as retas de adição padrão apresentaram coeficientes de correlação ( $r$ ) superiores a 0,990; em alguns poucos casos, esse valor esteve entre 0,970 e 0,990. A amostra no nível inferior de concentração, 1,5  $\mu\text{g mL}^{-1}$ , apresentou os piores resultados, com um erro de previsão em torno de 20%. Este resultado pode indicar um possível limite de quantificação do método, levando em conta que nessa concentração, o sinal do AS encontra-se completamente encoberto pelo sinal da fluorescência natural (“background”) do plasma. As previsões das demais amostras apresentaram erros de, no máximo, 6,3%, sendo que a grande maioria apresentou erros inferiores a 5%.

Esta aplicação representou um exemplo de uso da vantagem de segunda ordem. Os resultados mostraram ser possível a determinação direta de AS em plasma pela combinação de espectrofluorimetria, PARAFAC e adição padrão. Foram obtidos bons resultados para a faixa de concentração de 3 a 24  $\mu\text{g mL}^{-1}$ . As vantagens potenciais dessa determinação, menor custo e maior rapi-



dez, podem ser ainda mais evidenciadas considerando-se a possibilidade de automação do método proposto. A escala da análise pode ser diminuída, assim como os erros causados pela manipulação, e o ganho de tempo pode ser enormemente aumentado. A contribuição deste trabalho foi, mais que a determinação específica de AS em plasma, a proposição de uma alternativa nova para determinação de substâncias fluorescentes em matrizes complexas. Além da possível aplicação em análises clínicas, esta alternativa pode ser usada em outros tipos de matrizes, tais como leite, alimentos, plantas, etc.



**Figura 7.** Representação da montagem dos arranjos de dados tratados na determinação de AS pelo método da adição padrão. Superfície de excitação-emissão para a amostra de plasma  $3,0 \mu\text{g mL}^{-1}$  em AS

**Tabela 1.** Determinação de AS em plasma usando PARAFAC e adição padrão

Quantidade adicionada ( $\mu\text{g mL}^{-1}$ )	Quantidade prevista ( $\mu\text{g mL}^{-1}$ ) <sup>a</sup>	Erro de Previsão (%)
1,5	$1,82 \pm 0,16$	21,3
3,0	$3,19 \pm 0,28$	6,3
4,5	$4,65 \pm 0,33$	3,3
6,0	$6,27 \pm 0,26$	4,5
7,5	$7,76 \pm 0,37$	3,5
9,0	$8,94 \pm 0,24$	-0,7
12,0	$12,49 \pm 0,48$	4,1
15,0	$15,66 \pm 0,53$	4,4
21,0	$20,73 \pm 0,42$	-1,3
24,0	$24,37 \pm 0,46$	1,5

<sup>a</sup> Valores médios e desvios padrão de três determinações

## CONCLUSÃO E PERSPECTIVAS

Este artigo veio contribuir na divulgação do potencial do PARAFAC como método para tratamento de dados analíticos multidimensionais, com ênfase na sua aplicação em dados espectrofluorimétricos visando a determinação de fármacos em plasma humano. Embora os métodos quimiométricos de primeira ordem, como o PLS, já estejam razoavelmente popularizados entre a comunidade de química analítica, os métodos de ordem superior ainda são relativamente pouco conhecidos e muito do seu potencial ainda está por ser explorado. As contínuas automação e miniaturização de análises abrem também novas perspectivas, como, por ex., a obtenção de dados multidimensionais de absorção no UV pela análise em fluxo com gradiente de pH e detecção por arranjo de diodos. Outros sistemas de detecção, como sensores de fluorescência, e outras situações, como a obtenção de dados em função do tempo, permitem imaginar aplicações que envolvam quatro, cinco ou até mais dimensões dos dados, uma área ainda pouco estudada. Além disso, métodos que procurem se adequar a situações de desvio de trilinearidade dos dados, tais como o PARAFAC<sup>20,94</sup> e o PARALIND (“PARAllel profiles with LINEar Dependence”) ou análise de fatores paralelos com dependência linear<sup>95</sup> (originalmente denominado PARATUCK2<sup>20</sup>) já foram propostos e são uma alternativa para situações em que métodos como o PARAFAC não funcionem bem, como no caso de dados com deficiência de posto (“rank deficiency”)<sup>20</sup>. A deficiência ou sobreposição de posto nos dados ocorre quando há dependência linear entre os componentes, como no caso de experimentos cinéticos em que a formação de alguns produtos e, portanto, sua concentração, depende da quantidade de seus precursores, ou em experimentos envolvendo gradiente de pH, nos quais ambas as formas, ácida e básica, estão relacionadas ao mesmo analito. Nestes casos, o posto total dos dados não será igual à soma dos postos das espécies contribuintes<sup>96</sup>.

O uso da vantagem de segunda ordem na calibração de sistemas que contenham interferentes desconhecidos também abre a possibilidade de se elaborar novas estratégias de análise. Nesse sentido, o PARAFAC, quando combinado com o método da adição padrão, apresenta uma grande vantagem sobre o GRAM, seu potencial competidor. Ele permite múltiplas adições de padrão, ao contrário do GRAM, o qual está limitado a uma única adição e tem sua aplicação restrita em amostras complexas, tais como alimentos, amostras biológicas e de interesse ambiental. Outro aspecto a ser comentado é a necessidade do estabelecimento de figuras de mérito como requisito para reconhecimento deste tipo de métodos por órgãos oficiais, tais como as Farmacopéias. A determinação de figuras de mérito para métodos de ordem superior é uma área ainda pouco explorada e que apresenta muitas perspectivas de desenvolvimento<sup>17,97,98</sup>.

Finalmente, apesar de se considerar que ainda existe um longo caminho a ser percorrido até que os métodos baseados em quimiometria sejam aceitos por órgãos reguladores das áreas clínica e farmacêutica, e levando em conta todo o rigor necessário para a o controle de qualidade nessas áreas, não se pode deixar de reconhecer o enorme potencial e os muitos caminhos abertos para as pesquisas direcionadas a esses problemas.

## AGRADECIMENTOS

Ao CNPq pelo apoio financeiro, ao HEMOCENTRO da UNICAMP pelas amostras de plasma humano e ao Dr. R. Bro pela disponibilização do *The N-way Toolbox for MATLAB*.

## REFERÊNCIAS

- Otto, M.; Wegscheider, W.; Zambonin, P. G. Em *Analytical Chemistry*; Kellner, R.; Mermet, J. M.; Otto, M.; Widmer, H. M., eds.; Wiley-VCH: Berlin, 1998, p. 709-808.
- Martens, H.; Naes, T.; *Multivariate Calibration*, Wiley: Chichester, 1989.
- Brereton, R. G.; *Analyst* **2000**, *125*, 2125.
- The United States Pharmacopoeia*, 25<sup>a</sup> rev., U.S.P. Convention: Rockville, 2002.
- Gilpin, R. K.; Pachla, L. A.; *Anal. Chem.* **2003**, *75*, 2907.
- Bautista, R. D.; Jimenez, A. I.; Jimenez, F.; Arias, J. J.; *Fresenius J. Anal. Chem.* **1997**, *357*, 449.
- Bouhsain, Z.; Garrigues, S.; de la Guardia, M.; *Fresenius J. Anal. Chem.* **1997**, *357*, 973.
- Wiberg, K.; Hagman, A.; Buren, P.; Jacobsson, S. P.; *Analyst* **2001**, *126*, 1142.
- Ferraro, M. C. F.; Castellano, P. M.; Kaufman, T. S.; *J. Pharm. Biomed. Anal.* **2002**, *30*, 1121.
- Sena, M. M.; Fernandes, J. C. B.; Rover Jr, L.; Poppi, R. J.; Kubota, L. T.; *Anal. Chim. Acta* **2000**, *409*, 159.
- Sena, M. M.; Poppi, R. J.; *J. Pharm. Biomed. Anal.* **2004**, *34*, 27.
- Sena, M. M.; Chaudhry, Z. F.; Collins, C. H.; Poppi, R. J.; *J. Pharm. Biomed. Anal.* **2004**, *36*, 743.
- Navalón, A.; Blanc, R.; del Olmo, M.; Vilchez, J. L.; *Talanta* **1999**, *48*, 469.
- Damiani, P. C.; Borraccetti, M. D.; Olivieri, A. C.; *Anal. Chim. Acta* **2002**, *471*, 87.
- Arancibia, J. A.; Olivieri, A. C.; Escandar, G. M.; *Anal. Bioanal. Chem.* **2002**, *374*, 45.
- Arancibia, J. A.; Escandar, G. M.; *Talanta* **2003**, *60*, 1113.
- Muñoz de la Peña, A.; Espinosa Mansilla, A.; González Gómez, D.; Olivieri, A. C.; Goicoechea, H. C.; *Anal. Chem.* **2003**, *75*, 2640.
- Damiani, P. C.; Nepote, A. J.; Bearzotti, M.; Olivieri, A. C.; *Anal. Chem.* **2004**, *76*, 2798.
- Booksh, K. S.; Kowalski, B. R.; *Anal. Chem.* **1994**, *66*, 782A.
- Bro, R.; *Tese de Doutorado*, Universidade de Amsterdam, Holanda, 1998.
- Nikolajsen, R. P. H.; Booksh, K. S.; Hansen, A. M.; Bro, R.; *Anal. Chim. Acta* **2003**, *475*, 137.
- Olivieri, A. C.; Arancibia, J. A.; Muñoz de la Peña, A.; Durán-Merás, I.; Espinosa Mansilla, A.; *Anal. Chem.* **2004**, *76*, 5657.
- Durell, S. R.; Lee, C. H.; Ross, R. T.; Gross, E. L.; *Arch. Biochem. Biophys.* **1990**, *278*, 148.
- Tucker, L. R.; *Psychometrika* **1966**, *31*, 279.
- Kroonenberg, P. M.; *Three Mode Principal Component Analysis: Theory and Applications*, DSWO: Leiden, 1983.
- Kiers, H. A. L.; *Psychometrika* **1991**, *56*, 449.
- Harshman, R. A.; *UCLA Working Papers in Phonetics* **1970**, *16*, 1.
- Carroll, J. D.; Chang, J. J.; *Psychometrika* **1970**, *35*, 283.
- Ho, C. N.; Christian, G. D.; Davidson, E. R.; *Anal. Chem.* **1978**, *50*, 1108.
- Appellof, C. J.; Davidson, E. R.; *Anal. Chem.* **1981**, *53*, 2053.
- Russell, M. D.; Gouterman, M.; *Spectrochim. Acta, Part A* **1988**, *44*, 857.
- Russell, M. D.; Gouterman, M.; *Spectrochim. Acta, Part A* **1988**, *44*, 863.
- Russell, M. D.; Gouterman, M.; Vanzee, J. A.; *Spectrochim. Acta, Part A* **1988**, *44*, 873.
- Sanchez, E.; Kowalski, B. R.; *Anal. Chem.* **1986**, *58*, 496.
- Sanchez, E.; Kowalski, B. R.; *J. Chemom.* **1990**, *4*, 29.
- Wilson, B. E.; Lindberg, W.; Kowalski, B. R.; *J. Am. Chem. Soc.* **1989**, *111*, 3797.
- Tauler, R.; *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133.
- Linder, M.; Sundberg, R.; *Chemom. Intell. Lab. Syst.* **1998**, *42*, 159.
- Geladi, P.; *Chemom. Intell. Lab. Syst.* **1989**, *7*, 11.
- Smilde, A. K.; *Chemom. Intell. Lab. Syst.* **1992**, *15*, 143.
- Bro, R.; *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149.
- Bro, R.; *J. Chemom.* **1996**, *10*, 47.
- Andersson, C. A.; Bro, R.; *Chemom. Intell. Lab. Syst.* **2000**, *52*, 1.
- <http://www.models.kvl.dk/source>, acessada em Agosto 2004.
- Graham, A.; *Kronecker Products and Matrix Calculus with Applications*, Wiley: New York, 1981.
- Khatri, C. G.; Rao, C. R.; *Generalized Inverse of Matrices and its Applications*, Wiley: New York, 1968, p. 12.
- Hoffmann, R.; Minkin, V. I.; Carpenter, B. K.; *Hyle* **1997**, *3*, 3.
- Reis, M. M.; Biloti, D. N.; Ferreira, M. M. C.; Pessine, F. B. T.; Teixeira, G. M.; *Appl. Spectrosc.* **2001**, *55*, 847.
- Bro, R.; Kiers, H. A. L.; *J. Chemom.* **2003**, *17*, 274.
- Faber, N. M.; Bro, R.; Hopke, P. K.; *Chemom. Intell. Lab. Syst.* **2003**, *65*, 119.
- Munck, L.; Norgaard, L.; Engelsen, S. B.; Bro, R.; Andersson, C. A.; *Chemom. Intell. Lab. Syst.* **1998**, *44*, 31.
- Pereira-Filho, E. R.; Sena, M. M.; Arruda, M. A. S.; Poppi, R. J.; *Anal. Chim. Acta* **2003**, *495*, 177.
- Bro, R.; Jakobsen, M.; *J. Chemom.* **2002**, *16*, 294.
- Bro, R.; Smilde, A. K.; *J. Chemom.* **2003**, *17*, 16.
- Lakowicz, J. R.; *Principles of Fluorescence Spectroscopy*, 3<sup>rd</sup> ed., Plenum: New York, 2002.
- Booksh, K. S.; Muroski, A. R.; Myrick, M. L.; *Anal. Chem.* **1996**, *68*, 3539.
- Beltran, J. L.; Ferrer, R.; Guiteras, J.; *Anal. Chim. Acta* **1998**, *373*, 311.
- Jiji, R. D.; Cooper, G. A.; Booksh, K. S.; *Anal. Chim. Acta* **1999**, *397*, 61.
- Jiji, R. D.; Andersson, G. G.; Booksh, K. S.; *J. Chemom.* **2000**, *14*, 171.
- Pedersen, D. K.; Munck, L.; Engelsen, S. B.; *J. Chemom.* **2002**, *16*, 451.
- Bro, R.; *Chemom. Intell. Lab. Syst.* **1999**, *46*, 133.
- Baunsgaard, D.; Andersson, C. A.; Arndal, A.; Munck, L.; *Food Chem.* **2000**, *70*, 113.
- Martins, J. A.; Sena, M. M.; Poppi, R. J.; Pessine, F. B. T.; *Appl. Spectrosc.* **1999**, *53*, 510.
- Tan, Y. X.; Jiang, J. H.; Wu, H. L.; Cui, H.; Yu, R. Q.; *Anal. Chim. Acta* **2000**, *412*, 195.
- Andersen, C. M.; Bro, R.; *J. Chemom.* **2003**, *17*, 200.
- Trevisan, M. G.; Poppi, R. J.; *Anal. Chim. Acta* **2003**, *493*, 69.
- Escandar, G. M.; Gomez, D. G.; Mansilla, A. E.; de la Pena, A. M.; Goicoechea, H. C.; *Anal. Chim. Acta* **2004**, *506*, 161.
- Trevisan, M. G.; *Dissertação de Mestrado*, Universidade Estadual de Campinas, Brasil, 2003.
- Arcamone, F.; Cassinelli, G.; Fantini, G.; Grein, A.; Orezzi, P.; Pol, C.; Spalla, C.; *Biotechnol. Bioeng.* **1969**, *11*, 1101.
- Hortobagyi, G. N.; *Drugs* **1997**, *54* (Suppl. 4), 1.
- van Asperen, J.; Van Tellingen, O.; Beijnen, J. H.; *J. Chromatogr., B: Biomed. Sci. Appl.* **1998**, *712*, 129.
- Lachâtre, F.; Marquet, P.; Ragot, S.; Gaulier, J. M.; Cardot, P.; Dupuy, J. L.; *J. Chromatogr., B: Biomed. Sci. Appl.* **2000**, *738*, 281.
- Perez Ruiz, T.; Martinez Lozano, C.; Sanz, A.; Bravo, E.; *Electrophoresis* **2001**, *22*, 134.
- Loren, A.; Eliasson, C.; Josefson, M.; Murty, K. V. G. K.; Käll, M.; Abrahamsson, J.; Abrahamsson, K.; *J. Raman Spectrosc.* **2001**, *32*, 971.
- Wolfbeis, O. S.; Leiner, M.; *Anal. Chim. Acta* **1985**, *167*, 203.
- Sena M. M.; *Tese de Doutorado*, Universidade Estadual de Campinas, Brasil, 2004.
- Sena M. M.; Trevisan, M. G.; Poppi, R. J.; *Talanta*, no prelo.
- Abounassif, M. A.; Mian, M. S.; Mian, N. A. A. Em *Analytical Profiles of Drug Substances and Excipients 23*; Brittain, H. G., ed.; 1994, p. 421.
- Snodgrass, W. R.; *Ped. Clin. N. Am.* **1986**, *32*, 381.
- Kees, F.; Jehnich, D.; Grobecker, H.; *J. Chromatogr., B: Biomed. Sci. Appl.* **1996**, *677*, 172.
- Trinder, P.; *Biochem. J.* **1954**, *57*, 301.
- Pirola, R.; Bareggi, S. R.; DeBenedittis, G.; *J. Chromatogr., B: Biomed. Sci. Appl.* **1998**, *705*, 309.
- Goto, Y.; Makino, K.; Kataoka, Y.; Shuto, H.; Oishi, R.; *J. Chromatogr., B: Biomed. Sci. Appl.* **1998**, *706*, 329.
- Rover Jr, L.; Neto, G. O.; Fernandes, J. R.; Kubota, L. T.; *Talanta* **2000**, *51*, 547.
- Muñoz de la Peña, A.; Moreno, M. D.; Durán-Merás, I.; Salinas, F.; *Talanta* **1996**, *43*, 1349.
- Harris, D. C.; *Análise Química Quantitativa*, 5<sup>a</sup> ed., LTC: Rio de Janeiro, 2001, p. 87-89.
- Saxberg, B. E. H.; Kowalski, B. R.; *Anal. Chem.* **1979**, *51*, 1031.
- Booksh, K.; Henshaw, J. M.; Burgess, L. W.; Kowalski, B. R.; *J. Chemom.* **1995**, *9*, 263.
- Herrero, A.; Zamponi, S.; Marassi, R.; Conti, P.; Ortiz, M. C.; Sarabia, L. A.; *Chemom. Intell. Lab. Syst.* **2002**, *61*, 63.
- Wu, H. L.; Yu, R. Q.; Shibukawa, M.; Oguma, K.; *Anal. Sci.* **2000**, *16*, 217.
- Comas, E.; Gimeno, R. A.; Ferré, J.; Marcé, R. M.; Borrull, F.; Rius, F. X.; *J. Chromatogr., A* **2003**, *988*, 277.
- Gimeno, R. A.; Comas, E.; Marcé, R. M.; Ferré, J.; Rius, F. X.; Borrull, F.; *Anal. Chim. Acta* **2003**, *498*, 47.
- Bader, M.; *J. Chem. Educ.* **1980**, *57*, 703.
- Bro, R.; Andersson, C. A.; Kiers, H. A. L.; *J. Chemom.* **1999**, *13*, 295.
- Bro, R.; Harshman, R.; Sidiropoulos, N. D.; *J. Chemom.*, no prelo.
- Reis, M. M.; Gurden, S. P.; Smilde, A. K.; Ferreira, M. M. C.; *Anal. Chim. Acta* **2000**, *422*, 21.
- Olivieri, A. C.; Faber, N. M.; *Chemom. Intell. Lab. Syst.* **2004**, *70*, 75.
- Ni, Y.; Huang, C.; Kokot, S.; *Chemom. Intell. Lab. Syst.* **2004**, *71*, 177.

## APÊNDICE 1 – PRODUTO DE KRONECKER<sup>45</sup>

O produto de Kronecker, também conhecido como produto tensorial e representado pelo símbolo “ $\otimes$ ”, foi proposto pelo alemão L. Kronecker, no século XIX. Caracteriza-se como um operador matricial binário, transformando duas matrizes de dimensões arbitrárias em uma matriz de dimensão maior, com uma estrutura especial de bloco. Dadas as matrizes  $\mathbf{A}_{n \times m}$ , de dimensões  $n \times m$ , e  $\mathbf{B}_{p \times q}$ , de dimensões  $p \times q$ :

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{bmatrix}_{n \times m} \quad \mathbf{B} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,q} \\ \vdots & \ddots & \vdots \\ b_{p,1} & \cdots & b_{p,q} \end{bmatrix}_{p \times q}$$

O produto de Kronecker, definido por  $\mathbf{A} \otimes \mathbf{B}$ , é uma matriz de dimensões  $np \times mq$ , com a estrutura de bloco dada por:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,m}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n,1}\mathbf{B} & \cdots & a_{n,m}\mathbf{B} \end{bmatrix}_{np \times mq} \quad (6)$$

## APÊNDICE 2 – PRODUTO DE KHATRI-RAO<sup>46</sup>

O produto de Khatri-Rao foi proposto pelos indianos C. Khatri e C. Rao. O produto de Khatri-Rao também é um operador matricial binário, representado pelo símbolo “ $\text{|\otimes|}$ ”, e que pode ser escrito em função do produto de Kronecker. A vantagem de se utilizar o operador de Khatri-Rao na definição do modelo PARAFAC, em relação ao operador de Kronecker, é a simplificação matemática.

Dadas duas matrizes  $\mathbf{A}_{i \times f}$  e  $\mathbf{B}_{j \times r}$  com o mesmo número de colunas  $f = F$ , o produto de Khatri-Rao pode ser definido por:

$$\mathbf{A} \text{|\otimes|} \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \cdots \quad \mathbf{a}_F \otimes \mathbf{b}_F] \quad (7)$$

onde  $\mathbf{a}_i$  e  $\mathbf{b}_i$  são os vetores coluna das matrizes  $\mathbf{A}$  e  $\mathbf{B}$ , respectivamente.