Article

The Violacein Biosynthesis Monitored by Multi-Wavelength Fluorescence Spectroscopy and by the PARAFAC Method

Clecio Dantas, Pedro L. O. Volpe, Nelson Durán and Márcia M. C. Ferreira*

Institute of Chemistry, University of Campinas (UNICAMP), CP 6154, 13083-970 Campinas-SP, Brazil

A obtenção de informações sobre uma rota biossintética é um procedimento complexo e trabalhoso. Neste sentido, o presente trabalho apresenta uma nova abordagem para a análise inicial da biossíntese de produtos naturais fluorescentes usando a biossíntese da violaceína como exemplo. Para tanto, uma cultura da *Chromobacterium violaceum* foi inoculada em um biorreator de onde alíquotas eram retiradas a cada 2 h para posterior análise por espectroscopia de fluorescência 2D. As matrizes de emissão-excitação obtidas demonstraram o comportamento dinâmico dos sinais de fluoróforos que são consumidos e produzidos pela bactéria. Estes sinais foram resolvidos pelo método PARAFAC (análise paralela de fatores) perfazendo um total de seis espécies químicas. O triptofano e a violaceína foram identificados por comparação espectral. A identificação dos outros fluoróforos mostrou-se a etapa crítica devido à falta de banco de dados de produtos naturais fluorescentes para comparação. Por fim, esta metodologia apresenta um grande potencial para fornecer informações da biossíntese de produtos naturais.

Obtaining information about a biosynthetic pathway is a complex and laborious procedure. In this sense, this work presents a new approach for the initial analysis of the biosynthesis of fluorescent natural products using as example the violacein biosynthesis. For this, a culture of *Chromobacterium violaceum* was grown in a bioreactor from which aliquots were collected every 2 h for subsequent analysis by multi-wavelength fluorescence spectroscopy. The excitation-emission matrices demonstrated the dynamic behavior of the fluorophores signal that are consumed and produced by the bacterium. These signals were resolved by PARAFAC (parallel factor analysis) method totalizing six pure components. Tryptophan and violacein were identified by comparison to spectra available in the literature. The identification of other fluorophores was critical step due to the lack of a database of fluorescent natural products to compare spectra. Finally, this methodology has great potential to achieve a deeper insight into the biosynthesis of natural products.

Keywords: biotransformation, violacein, fluorescence spectroscopy, curve resolution

Introduction

Natural products have had historical success as biologically active structures. Among these, the violet pigment, named violacein, produced mainly by bacteria of the genus *Chromobacterium* has attracted increased interest owing to its important biological activities and pharmacological and industrial potentials.[1,2]

Violacein (Figure 1) is a secondary metabolite with a molar mass of 343.3 amu and is constituted of 5-hydroxyindole, 2-pyrrolidone and 2-oxindole moieties that show strong absorption in the visible region due to resonance.[3]

Studies about violacein biosynthesis began in 1934 with Tobie,[4] who observed that oxygenation of a
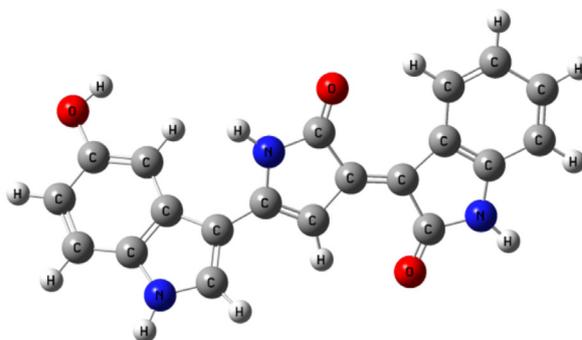
Figure 1. Chemical structure of violacein.

Chromobacterium violaceum culture greatly reduced the time required for maximum pigment production. Twenty five years later, DeMoss and Evans[5,6] discovered that to synthesize violacein, the bacteria needed molecular

*e-mail: marcia@iqm.unicamp.br

oxygen and that L-tryptophan was incorporated into violacein, except for the carboxylic carbon, eliminated by a decarboxylation process during the biosynthesis. From 1987 to 1990, Hoshino *et al.*[7-9] demonstrated that (*i*) the carbon skeleton of the pyrrolidone moiety was built up by the condensation of the side chains of two L-tryptophan molecules accompanied by the 1,2-shift of the indole ring on one molecule of tryptophan;[7] (*ii*) all carbon, nitrogen and hydrogen atoms in the pyrrolidone moiety were provided exclusively by L-tryptophan and that the oxygen atoms come from molecular oxygen;[8] and (*iii*) the intermediacy of 5-hydroxy-L-tryptophan in the violacein biosynthesis since the hydroxylation of tryptophan was the first step of violacein biosynthesis.[9]

After that, the violacein biosynthesis has only been evaluated by radiolabel incorporation experiments. In 1991, Pemberton *et al.*[10] began a study at the genetic level by isolating a 14.5 kb fragment containing the genes encoding violacein biosynthesis from *C. violaceum*. From 1993 to 2000, in a series of works, Hoshino *et al.*[11-14] discovered that (*i*) chromopyrrolic acid (CPA) was produced independently from the violacein biosynthesis;[11] (*ii*) proviolacein, prodeoxyviolacein and pseudoviolacein were identified as plausible biosynthetic intermediates of violacein and that oxygenation the 2-position of the indole ring occurs in a final step in this biosynthesis;[12] (*iii*) oxygenases were involved and cofactor NADPH was needed for the biosynthesis of violacein;[13] and (*iv*) that the 1,2-shift of the indole ring occurred through an intramolecular process during the formation of the left part (5-hydroxyindole side) of the violacein skeleton.[14]

In 2000, August *et al.*[15] showed that the violacein biosynthetic gene cluster was comprised of four genes, VioABCD. The products of the VioA, VioC and VioD were nucleotide-dependent monooxygenases. Disruption of VioA or VioB completely abrogated the biosynthesis of violacein intermediates, while disruption of VioC or VioD genes resulted in the production of violacein precursors. Collecting all information obtained so far, August *et al.*[15] proposed a hypothetical route for violacein biosynthesis from L-tryptophan. This pathway consisted of: (*i*) modification of one L-tryptophan by VioD, forming 5-hydroxy-L-tryptophan, and oxidative deamination of the other by VioA, producing indole-3-pyruvic acid (IPA); (*ii*) decarboxylative fusion of these two tryptophan-derived units and a 1,2-shift of the 5-hydroxyindole ring by VioB, resulting in prodeoxyviolacein; and (*iii*) oxygenation of the latter by VioC yielding violacein.

In 2005, Howard-Jones and Walsh[16] reported a similarity between violacein biosynthesis and indolecarbazole biosynthesis. Both pathways include a decarboxylative fusion of two tryptophan-derived units. Additionally, the authors found a level of identity of 34% between VioB and RebD (a heme protein required for formation of CPA and involved in the biosynthesis of the indocarbazoles rebeccamycin and staurosporine). Based on the similarities between these two pathways, Sánchez *et al.*,[17] in 2006, hypothesized the construction of hybrid pathways for production of oxygenated indolocarbazole, by coexpression of the genes. The hybrid pathways did not yield oxygenated indolocarbazoles, however their results provided new information on violacein biosynthesis. First, a pair of genes (VioAB), responsible for the earliest steps in violacein biosynthesis, was equivalent to the homologous pair in the indolocarbazole pathway (RebOD), directing the formation of CPA. Second, in addition to VioABCD, a fifth gene (VioE) was essential for violacein biosynthesis. Third, CPA was not an intermediate product of this biosynthesis. However, its formation was not independent of violacein biosynthesis as it was produced by VioAB in the absence of VioE. Fourth, both VioC and VioD oxygenases appeared to act on the latter steps of violacein biosynthesis. The last two results were contradictory to those reported by Hoshino *et al.*,[9,11,14] and, consequently, to the mechanism proposed by August *et al.*[15]

In 2006, Balibar and Walsh[18] reconstituted *in vitro* the entire violacein pathway. Their results demonstrated that: (*i*) L-tryptophan, and not 5-hydroxy-L-tryptophan, was the sole precursor of violacein; (*ii*) VioB was responsible for the oxidative coupling of two molecules of IPA imine generated by VioA for formation of the pyrrole core; (*iii*) the fifth gene, VioE, was responsible for the 1,2-shift of the indole ring that resulted in the formation of prodeoxyviolacein rather than CPA; (*iv*) formation of violacein then proceeded by the sequential action of VioD and VioC. These results confirmed the *in vivo* results proposed by Sánchez *et al.*[17]

In 2007, Shinoda *et al.*[19] reported the identification of a true intermediate of violacein biosynthesis, named protoviolaceinic acid, which was produced by incubating L-tryptophan with VioABDE in the presence of NADPH. The production of this compound confirmed that VioC works to oxygenate the 2-position of the right side indole ring. Additionally, their results indicated that the oxygenation reaction to form the central pyrrolidone core proceeded in a non-enzymatic fashion. An overview of the violacein biosynthesis proposed by these authors is shown in Figure 2.

It begins with the VioA oxidation of L-tryptophan to indole-3-pyruvic acid (IPA) imine. Then, the oxidative coupling of two molecules of IPA imine by VioB gives an uncharacterized intermediate proposed to be an IPA
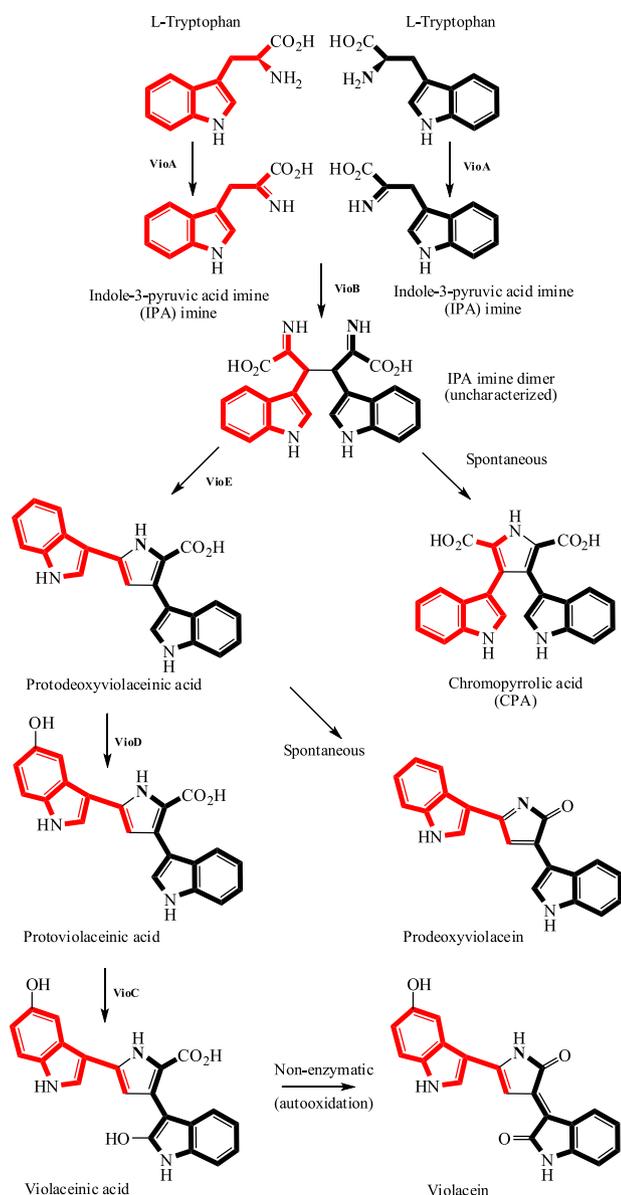
Figure 2. Overall pathway to violacein biosynthesis.

imine dimer. The skeleton of this compound is catalytically arranged by VioE through an intramolecular rearrangement of the indole ring producing protodeoxyviolaceinic acid. This latter can be converted spontaneously by autoxidation into prodeoxyviolacein or it can undergo hydroxylation catalyzed by VioD at the 5-position of the left side indole ring giving protoviolaceinic acid. The protoviolaceinic acid is hydroxylated by VioC at the 2-position of the right side indole ring to form violaceinic acid. The subsequent conversion to violacein involves a non-enzymatic process of oxidative decarboxylation.

The mechanistic investigation of this complex biosynthetic pathway at the chemical, biochemical and genetic levels involved many steps. Radiolabel incorporation experiments coupled with techniques for separation and

structural elucidation, molecular tools and recombinant methods were used. However, a special property of this system was not explored. In the violacein biosynthesis, the precursor (L-tryptophan), the product (violacein) and, probably, the intermediaries are fluorescent compounds. Hence, it is expected that information about this biosynthesis can be obtained from fluorescence spectroscopy when it is applied to detect compounds that are consumed and produced during cultivation.

Thus, a methodology for obtaining information about the biosynthetic pathway of fluorescent compounds is proposed in this work. It consists in employing multi-wavelength fluorescence spectroscopy and PARAFAC method to detect and identify (through spectral resolution) fluorophores that are consumed and produced by a bacterium and, finally, to establish a relationship with the natural product biosynthesis.

The combination of the multi-wavelength fluorescence spectroscopy and the PARAFAC method has already been applied to monitor bioprocesses in which the time evolution profiles obtained for the fluorophores are related to the target bioprocess variables.[20-23] However, this methodology was not employed to directly link the spectral resolution information to a biosynthetic pathway. Therefore, this is a new and complementary approach for the initial analysis of the biosynthesis of fluorescent natural products.

Experimental

Organisms, culture media and cultivation conditions

Starting from a culture of *C. violaceum* CCT 3496, a colony was isolated, after culture by the streak plate method. This pure colony was inoculated into a 250 mL Erlenmeyer flask containing 50 mL of sterilized culture medium (0.50% D-glucose, 0.50% bacteriologic peptone, 0.25% yeast extract and 0.03% L-tryptophan) and grown for 24 h at 33 °C on an orbital shaker at 200 rpm. Afterwards, 10 mL of this bacterial culture was transferred to a 1500 mL BioFlo bioreactor (New Brunswick Scientific) containing 1000 mL of sterilized culture medium and the parameters temperature, stirring and air flow were adjusted (33 °C, 200 rpm and 1.0 L min⁻¹, respectively) and kept constant throughout the cultivation (for 36 h).

Sample collection

Every 2 h, aliquots of 10 mL were withdrawn from the bioreactor with the aid of a glass syringe, stored in 15 mL Falcon tubes and frozen (−20 °C). At a later stage, each aliquot was thawed in a thermostatic bath at 30 °C for

5 min and then centrifuged at 7000 rpm for 10 min. The supernatant was eliminated and 5 mL absolute ethanol was added to extract intracellular compounds, which were then centrifuged at 7000 rpm for 10 min for cell removal and the new supernatant was collected. The crude ethanolic extract was passed through a 0.45 μm pore-size filter. Altogether, 18 samples corresponding to the crude ethanolic extract of *C. violaceum* at different stages of biosynthesis were collected.

Multi-wavelength fluorescence spectroscopy

Fluorescence measurements were performed using a Varian Cary Eclipse fluorescence spectrophotometer in the scan mode. Aliquots of 0.5 mL of the samples were diluted in 2.5 mL of absolute ethanol in the quartz cuvette to reduce effects of quenching, inner filter and energy transfer processes. Initially, the excitation-emission matrices (EEMs) were collected for each sample with spectral ranges of 250-620 nm for excitation and 270-800 nm for emission. The scan rate for each EEM was 1200 nm min^{-1}. The increments for the excitation and emission wavelengths were 5 and 2 nm, respectively. The excitation and emission slits were adjusted to 5 nm. Filters were set on "Auto" for both excitation and emission monochromators. This means that the filter wheel is automatically moved to the appropriate position according to the selected excitation/emission wavelength. The photomultiplier tube (PMT) voltage was set to 600 V. The measurements were carried out in 90° geometry.

It is known that the violacein molecule shows an emission band at 675 nm when excited at 575 nm.[2] However, with the measurement conditions described above, it was not possible to detect fluorescence signals in the spectral region of 550-620 nm for excitation and 600-800 nm for emission (Figure 3a). To explore this problem, new instrumental settings were tested. Filters were set on "Open" for both excitation and emission monochromators, which means that no filter was used and the photomultiplier tube (PMT) voltage was increased to 800 V. Usually, excitation and emission filters are used to reduce the stray radiation effect. Additionally, the output signal for a given amount of light will be enhanced by increasing the PMT voltage.

These new instrumental settings allowed visualization of fluorescence signals in the low energy region as shown in Figure 3b. Hence, it was decided to investigate the entire spectral range by dividing it into smaller regions and using specific instrumental settings for each. Altogether, five regions were defined (Figure 3a) and their instrumental settings are summarized in Table 1.

Following this procedure, six EEMs, covering the entire spectral region and the regions B, C, D, E and F were collected for each sample.

Data preprocessing

Rayleigh scatter peaks are largely unrelated to the chemical properties of the sample. They occur when a molecule has been excited to a virtual energy state by a photon with insufficient energy to completely excite the molecule.[24] They are situated diagonally in EEMs (Figure 3a) and do not present linear behavior which may complicate the fluorescence data modeling. Therefore, Rayleigh scatter peaks should be treated prior to modeling.

Many approaches have been developed for that.[25-28] The procedure employed here consisted in removing and replacing affected areas by interpolated values according to the data in the remaining parts of EEM. Interpolation of scattered areas was performed by applying the method developed by Bahram *et al.*,[29] which is available from another work.[30]
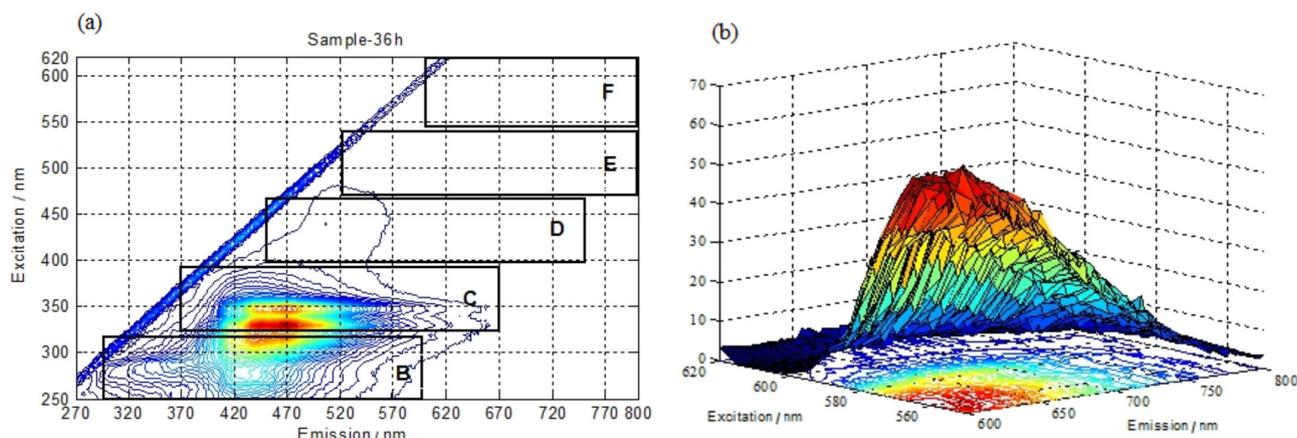
Figure 3. (a) 2D contour plot of an EEM and spectral range of each region; (b) 3D surface and contour plots of an EEM in the spectral region of 550-620 nm for excitation and 600-800 nm for emission.

Table 1. Measurement conditions for each spectral region

Region	Excitation range / nm	Emission range / nm	Excitation filter	Emission filter	PMT voltage / V	Excitation slit / nm	Emission slit / 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	open	open	800	5	5
E	475-545	525-800	open	open	800	5	5
F	550-620	600-800	open	open	800	5	5

PMT: photomultiplier tube.

Data modeling

Multi-wavelength fluorescence spectroscopy involves successive acquisition of emission spectra at multiple excitation wavelengths generating a two-dimensional matrix (EEM) *per* sample. Using the culture time, it is possible to produce a 3D array $\underline{\mathbf{X}}$ with dimensions $I \times J \times K$ by stacking EEMs one on top of another (Figure 4), in which I is the number of samples collected during the reaction, J is the number of emission wavelengths and K is the number of excitation wavelengths.

This array $\underline{\mathbf{X}}$ can be modeled by PARAFAC,[31] a trilinear decomposition method for higher order data whose structural basis is given by three loading matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ (Figure 4). These matrices are obtained using an alternating least squares algorithm and contain information related to relative concentration, emission spectrum and excitation spectrum of each analyte. The core $\underline{\mathbf{H}}$ is a binary array ($F \times F \times F$) with ones in the super-diagonal and zeros in all other elements and indicates that only loading vectors with the same column number interact, i.e., only the fth column of $\mathbf{A}$ interacts with the fth columns of the $\mathbf{B}$ and $\mathbf{C}$ matrices.

The fluorescence excitation-emission matrices are known to approximately follow a trilinear model.[32] Therefore, the application of the PARAFAC method in fluorescence excitation-emission data can resolve overlapping signals into pure concentration and spectral profiles.

The trilinear model is found to minimize the sum of squares of the residuals, e_{ijk}, according to equation 1,

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

where, x_{ijk} is the intensity of the ith sample at the jth emission wavelength and at the kth excitation wavelength. The a_{if} element may be interpreted as the relative concentration of analyte f in sample i. The j-vector $\mathbf{b}_f$ with elements b_{jf} ($j = 1,\ldots, J$) is the estimated emission spectrum of this analyte and likewise $\mathbf{c}_f$ is the estimated excitation spectrum.

The most important step when applying the PARAFAC method is to estimate the appropriate number of components or factors (F) in the data set (see equation 1). This number can be determined based on several criteria as, for example, visual appearance, explained variance (R^2), core consistency diagnostic (CORCONDIA), among others.[32-34] In this work, the number of components was assessed on the basis of three criteria: explained variance, CORCONDIA and lack of fit (LOF) of the models.

The explained variance (R^2) is calculated by taking into account the sum of the squares of the residuals, e_{ijk}, and the sum of the squares of the elements of the array $\underline{\mathbf{X}}$, x_{ijk}, according to equation 2.

$$\mathrm{R}^2(\%) = \left(1 - \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} e_{ijk}^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}^2} \right) 100 \tag{2}$$
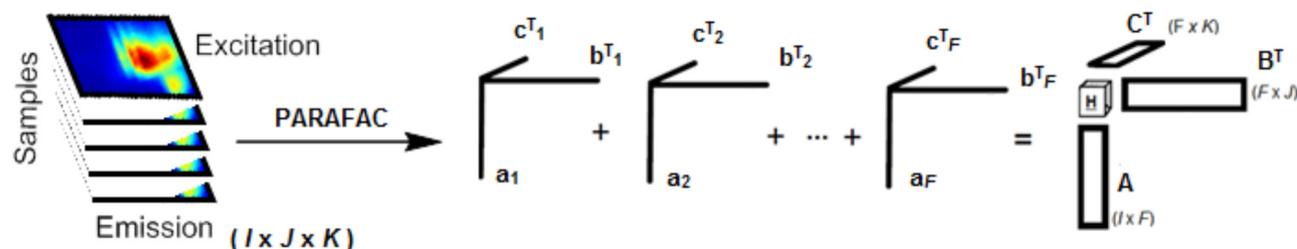
Figure 4. Three-way data array obtained by stacking EEMs for each sample and graphical representation of PARAFAC decomposition.

CORCONDIA is defined as

$$CORCONDIA(\%) = \left(1 - \frac{\sum\limits_{d=1}^{F}\sum\limits_{e=1}^{F}\sum\limits_{f=1}^{F}\left(g_{def} - h_{def}\right)^2}{\sum\limits_{d=1}^{F}\sum\limits_{e=1}^{F}\sum\limits_{f=1}^{F} h_{def}^2}\right)100 \quad (3)$$

where, g_{def} is the calculated element of the core using the PARAFAC model; h_{def} the element of a binary array and F is the number of factors in the model. In an ideal PARAFAC model, g_{def} is equal to h_{def} and, in this case, CORCONDIA will be equal to 100%. A model with a CORCONDIA value above 90% can be understood as "very trilinear", whereas a model with a CORCONDIA in the neighborhood of 50% would mean a problematic model with signs of both trilinear and not trilinear variations. A CORCONDIA close to zero or even negative implies an invalid model. More detailed information can be seen in elsewhere.[33]

In complex biological matrices, the determination of the number of factors by CORCONDIA is sometimes elusive and cannot be automated.[35] In this sense, the appropriate number of factors should be evaluated by other criteria. Here, the LOF criterion is introduced and calculated according to equation 4.

$$LOF(\%) = \left(\frac{\sqrt{\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}\sum\limits_{k=1}^{K} e_{ijk}^2}}{\sqrt{\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}\sum\limits_{k=1}^{K} x_{ijk}^2}}\right)100 \quad (4)$$

The PARAFAC algorithm used in this work is from the N-way Toolbox,[36] downloaded from elsewhere[37] and used with the software MatLab 7.0 (Mathworks Inc, USA).

Results and Discussion

Growth phases of bacteria

The violacein biosynthesis was monitored in the bioreactor for 36 h producing a total of 18 samples. The visual changes that occurred in the culture medium during this period are presented in Figure 5.

The first hours after inoculation corresponded to a lag phase that is the period of adaptation of bacteria to the environment and is characterized by intense metabolic activity and the absence of cell division. Following the lag phase is the exponential or log phase, in which the population grows in a logarithmic fashion. This phase can be easily visualized by the turbidity of the culture medium. Finally, it is possible to see the stationary phase where the growth rate slows down as a result of nutrient depletion and
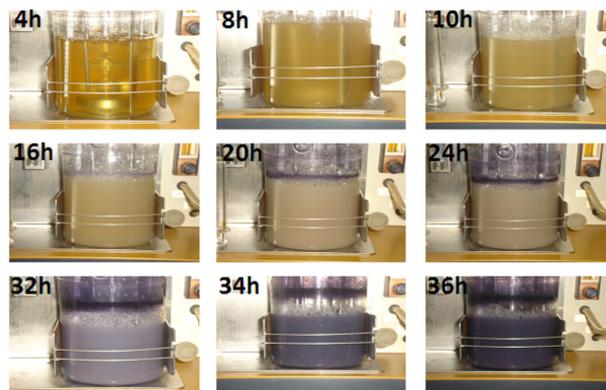
Figure 5. Visual changes occurring during the violacein biosynthesis.

accumulation of toxic products. It is in the stationary phase that several secondary metabolites (among them, violacein) are synthesized.[38]

Fluorescence data

The fluorescence data set consisted of 18 excitation-emission matrices (EEMs) of ethanolic extracts of *C. violaceum* collected at different stages of violacein biosynthesis. The fluorescence landscapes were initially measured at 266 emission wavelengths (from 270 to 800 nm with 2 nm intervals) and 75 excitation wavelengths (from 250 to 620 nm with 5 nm intervals) providing a detailed map of the fluorescence properties of the sample. The dynamic behavior of fluorophores involved in the violacein biosynthesis is shown in Figure 6 in the form of contour plots of some preprocessed EEMs.

In Figure 6a, the first fluorophore shows a broad band of emission between 330 and 370 nm when excited from 270-290 nm. This compound seems to be consumed during the fermentation since the intensity of the emission band decreases with the reaction time. The second fluorophore shows two excitation maxima (270-285 and 295-320 nm) and one emission maximum between 400-450 nm. The third fluorophore shows three excitation maxima (315-320, 325-340 and 345-355 nm) and one emission maximum from 440-480 nm. These last two compounds are possible intermediates in the violacein biosynthesis since they are being produced with the progress of the fermentation.

As previously described in the Experimental section, with the initial instrumental settings it was not possible to detect the fluorescence signal from violacein. However, the violet color of the culture medium (Figure 5) indicates that this compound is being biosynthesized, a fact that was confirmed by UV-Vis absorption spectroscopy (data not shown). This apparent contradiction is probably caused by the dependence of sensitivity of fluorescence in both the fluorophore and the instrument.
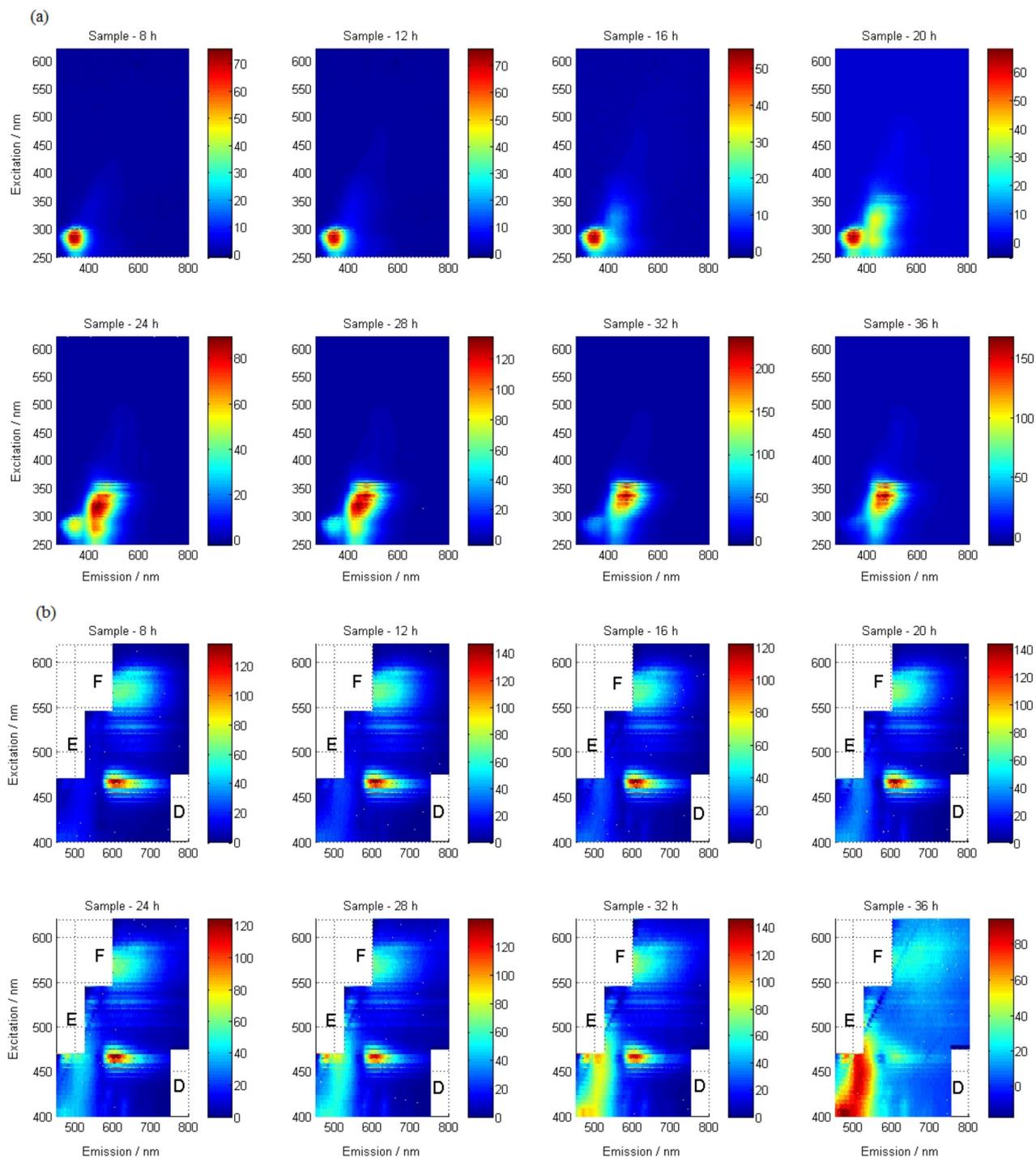
Figure 6. Kinetic evolution of EEMs obtained during the violacein biosynthesis (a) entire spectral range and (b) low energy regions (D, E and F).

To assess this problem, the entire spectral range was divided into five smaller regions, defined in Table 1, and specific measurement conditions were used for each of them. This procedure produced 18 fluorescence landscapes for each region.

The dynamic behavior of fluorophores in the low energy regions (D, E and F regions) is shown in Figure 6b, in which new EEMs were created by juxtaposition of preprocessed

D, E and F EEMs to give an overview of the fluorescence signals. The missing values were set to NaN.

The fluorophore with maximum emission between 590-620 nm shows three excitation maxima (450-470, 525-535 and 550-570 nm). In this case, the absorption peak of lowest energy is due to fluorophore excitation to an accessible vibrational level of the first electronic excited state of the same multiplicity. The absorption peak of

highest energy is probably due to fluorophore excitation to an accessible vibrational level of the second electronic excited state of the same multiplicity. The only emission band observed indicates the occurrence of the internal conversion process, wherein excited molecules generally relax toward the lowest vibrational level within the first electronic excited state. Also it is possible to observe a fluorophore that is produced during the fermentation and presents broad excitation (400-460 nm) and emission (450-540 nm) bands.

PARAFAC modeling

Pure excitation and emission profiles for each of the six regions were extracted from the preprocessed EEMs by applying PARAFAC. The appropriate number of components in each data set was determined based on the explained variance, CORCONDIA and LOF values.

PARAFAC models with non-negativity constraints in the three modes were calculated from one to six components and the relative values of explained variance, CORCONDIA and LOF are listed in Table 2.

Analyzing the results in Table 2 for the entire spectral range and for region B, it can be inferred that the explained variance and LOF values indicate that a model with three components is suitable since the improvement obtained in these parameters when one more component is included in the model is very small (0.12 and 0.98 for entire spectral range and 0.13 and 1.34 for region B, respectively). On the other hand, the CORCONDIA values obtained from PARAFAC calculations upon varying the number of components for these two regions show that the optimal number of components is not very clear (a three component model has CORCONDIA values of 99.58 and 99.01% for the entire spectral range and for region B, respectively, while a four component has CORCONDIA values of 70.68 and 72.69%). However, the graphical visualization of spectral loading matrices obtained for PARAFAC models with four components for the entire spectral range and for region B shows that, except for the fourth species, the shape and position of the recovered profiles do not change. This feature is consistent with the trilinearity property. In this way, it is

Table 2. Percentage of explained variance, CORCONDIA and LOF, calculated for trilinear models using the PARAFAC method with 1 to 6 components

No. components	1	2	3	4	5	6
			Entire spectral range			
Explained variance / %	82.30	96.03	99.56	99.68	99.75	99.83
CORCONDIA / %	100	99.93	99.58	70.65	11.58	5.69
LOF / %	42.07	19.93	6.61	5.63	5.02	4.08
			Region B			
Explained variance / %	80.85	98.35	99.70	99.83	99.87	99.93
CORCONDIA / %	100	99.97	99.01	72.69	25.83	13.31
LOF / %	43.76	12.85	5.49	4.15	3.62	2.70
			Region C			
Explained variance / %	98.91	99.71	99.89	99.95	99.96	99.97
CORCONDIA / %	100	99.93	22.65	11.25	−0.23	−0.05
LOF / %	10.45	5.38	3.34	2.20	1.90	1.65
			Region D			
Explained variance / %	75.03	96.51	98.16	98.83	99.21	99.63
CORCONDIA / %	100	99.77	51.39	33.75	12.41	4.07
LOF / %	49.97	18.69	13.55	10.82	8.87	6.09
			Region E			
Explained variance / %	86.65	97.13	97.96	98.50	98.87	99.14
CORCONDIA / %	100	99.98	35.39	42.35	7.64	6.14
LOF / %	36.53	16.94	14.26	12.23	10.62	9.26
			Region F			
Explained variance / %	96.18	98.60	99.44	99.63	99.67	99.71
CORCONDIA / %	100	97.72	55.56	27.05	11.73	−1.21
LOF / %	19.55	11.82	7.49	6.09	5.57	5.36

R^2: explained variance; CORCONDIA: core consistency diagnostic; LOF: lack of fit.

concluded that the fourth component is not a chemical but a physical component. Therefore, the number of factors was chosen to be three for both regions. This observation reinforces that the determination of the number of factors by CORCONDIA in complex biological matrices is sometimes elusive and cannot be automated.

For regions C and E, the explained variance, LOF values and CORCONDIA values indicate that two is an appropriate number of factors to model the data. Although these same criteria point to three factor models for regions D and F, it was assumed that two is a more reasonable value. This choice was based on drastic variation of the CORCONDIA values for models with 3 or more factors when N-way Toolbox and PLS-Toolbox[39] were run several times (an indication of ill-posed models).

The concentration, emission and excitation profiles obtained by the PARAFAC models for each region, with the suitable number of components, are shown in Figure 7.

In all, six chemical species probably involved in the biosynthesis of violacein could be detected by this methodology. In Figure 7a, the first excitation profile has a narrow maximum centered at 280 nm and a sharp emission maximum at 342 nm. The respective concentration profile shows a maximum concentration at 6 h (lag phase) and then decreases with reaction time. The emission loadings are similar to the pure emission spectrum of tryptophan.[40] This fact corroborates the mechanism presented in Figure 2, in which tryptophan is the precursor in the violacein

biosynthesis. The second fluorophore has an emission peak at 423 nm when excited at 280 and 308 nm (see Figure 6a, sample-20 h). This compound is probably a tryptophan derivative since the excitation at 280 nm is due to the indole nucleus. The double absorption can be understood as electronic transitions to different excited states or as electronic transitions between two different vibrational levels in the same excited state. The concentration profile indicates that this fluorophore reaches its maximum concentration after approximately 30 h and then begins to be consumed. Unfortunately, it was not possible to identify this compound by comparison to spectra available in the literature. However, the concentration profile suggests that it is an intermediate of the violacein biosynthesis. The third fluorophore shows three excitation peaks at 315, 330 and 347 nm and one emission peak at 470 nm (see Figure 6a, sample-36 h). The concentration profile indicates that this fluorophore starts to be produced after approximately 18 h and reaches its maximum concentration at 32 h.

The emission profiles obtained by the resolution in the regions B and C are the same visualized in Figure 7a. The excitation profiles are also the same. However, as the excitation range is smaller, the excitation profiles are truncated. The juxtaposition of the excitation profiles obtained for the regions B and C recovers the excitation profile of the entire spectral range.

The fluorophore represented by the continuous red line in Figures 7b shows similar concentration profiles to that of
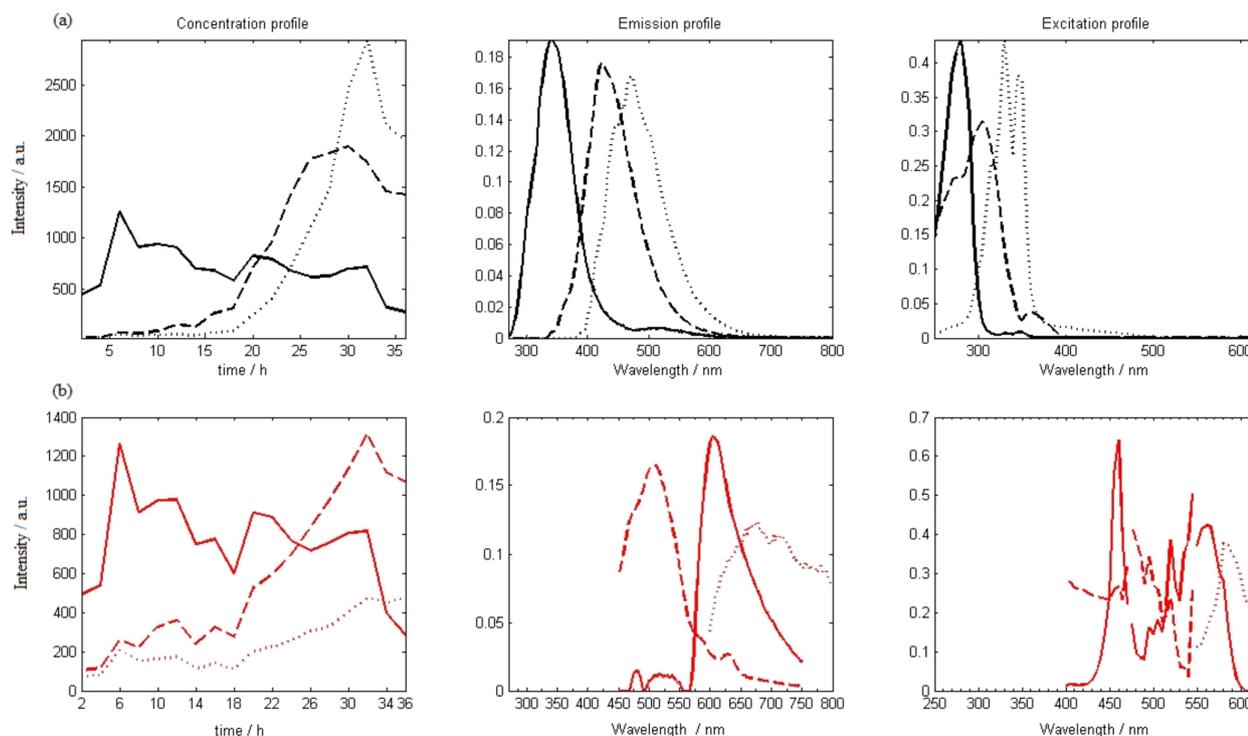
Figure 7. Estimates of emission and excitation spectra obtained by PARAFAC: (a) entire spectral range and (b) regions D, E and F juxtaposed.

tryptophan. This compound seems to have three excitation peaks at 460, 525 and 560 nm and one emission peak close to 605 nm (see Figure 6b, sample-8 h).

Asamizu *et al.*[41] reported the electronic absorption spectra of StaD, an enzyme involved in saturosporine biosynthesis that is a homolog to VioB and responsible for the coupling reaction between two molecules of indole-3-pyruvic acid and NH_4^+ to yield CPA. It has an absorption maximum at 430 nm and smaller absorption bands at 530 and 565 nm and presents an excitation spectra similar to that of the fluorophore above described. This fact probably indicates that other indolocarbazoles are being produced during the cultivation.

The compound represented by the dashed red line in Figures 7b shows a broad excitation band at 400-470 nm and peak emission around 512 nm. The concentration profile indicates that this fluorophore is produced during violacein biosynthesis. Finally, in Figure 7b, the excitation profile represented by the dotted red line with narrow maximum centered at 575 nm and an emission maximum at 675 nm is similar to the spectrum of violacein.[3] As expected, the concentration profile of this compound increases with reaction time.

Conclusion

The use of the multi-wavelength fluorescence spectroscopy and PARAFAC was introduced as a new approach for obtaining mechanistic information about the biosynthesis of fluorescent natural products. Violacein biosynthesis is used as an example in which six fluorescent compounds were detected and their concentration and spectral profiles were resolved by this methodology. Tryptophan (identified by comparison to emission and excitation spectra) is consumed during the bioprocess, corroborating the mechanism presented in Figure 2, in which this compound is the precursor molecule in violacein biosynthesis. Additionally, the final product of this pathway, the violacein molecule, also was identified by means of its concentration and spectral profiles.

Unfortunately, the identification of other fluorophores was not possible by spectra comparison due the lack of a database of fluorescent natural products. This critical step highlights the necessity of the improvement of databases to aid the type of analysis proposed herein, where a complex mechanism could be elucidated by simpler methodology than the traditional methods. Thus, it is believed that the dissemination of this methodology is the first step in its consolidation. Future efforts are being conducted to try identifying these compounds using techniques of separation and structural elucidation.

Another important conclusion is to show that the use of different instrumental settings is important to enable the visualization of fluorescence signals in the lower energy spectral region due the wavelength dependent sensitivity of the spectrofluorometer PMT.

Finally, this methodology has great potential to achieve a deeper insight into the biosynthesis of natural products.

Acknowledgments

References

1. Durán, N; Menck, C. F. M; *Crit. Rev. Microbiol.* **2001**, *27*, 201.

2. Durán, N.; Justo, G. Z.; Ferreira, C. V.; Melo, P. S.; Cordi, L.; Martins, D.; *Biotechnol. Appl. Biochem.* **2007**, *48*, 127.

3. Rettori, D.; Durán, N.; *World J. Microbiol. Biotechnol.* **1998**, *14*, 685.

4. Tobie, W. C.; *J. Bacteriol.* **1934**, *29*, 223.

5. DeMoss, R. D.; Evans, N. R.; *J. Bacteriol.* **1959**, *78*, 583.

6. DeMoss, R. D.; Evans, N. R.; *J. Bacteriol.* **1960**, *79*, 729.

7. Hoshino, T.; Kondo, T;. Uchiyama, T.; Ogasawara, N.; *Agric. Biol. Chem.* **1987**, *51*, 965.

8. Hoshino, T.; Takano, T.; Hori, S.; Ogasawara, N.; *Agric. Biol. Chem.* **1987**, *51*, 2733.

9. Hoshino, T.; Ogasawara, N.; *Agric. Biol. Chem.* **1990**, *54*, 2339.

10. Pemberton, J. M.; Vicent, K. M.; Penfold, R. J.; *Curr. Microbiol.* **1991**, *22*, 355.

11. Hoshino, T.; Kojima, Y.; Kayashi, T.; Uchiyama, T.; Kaneko, K.; *Biosci. Biotechol., Biochem.* **1993**, *57*, 775.

12. Hoshino, T.; Hayashi, T.; Odajima, T.; *J. Chem. Soc., Perkin. Trans.* **1995**, *1*, 1565.

13. Hoshino, T.; Yamamoto, M.; *Biosci. Biotechnol., Biochem.* **1997**, *61*, 2134.

14. Ruhul Momen, A. Z. M.; Hoshino, T.; *Biosci. Biotechnol., Biochem.* **2000**, *64*, 539.

15. August, P. R.; Grossman, T. H.; Minor, C.; Draper, M. P.; MacNeil, I. A.; Pemberton, J. M.; Call, K. M.; Holt, D.; Osburne, M. S.; *J. Mol. Microbiol. Biotechnol.* **2000**, *2*, 513.

16. Howard-Jones, A. R.; Walsh, C. T.; *Biochemistry* **2005**, *44*, 15652.

17. Sánchez, C.; Braña, A. F.; Méndez, C.; Salas, J. A.; *ChemBioChem* **2006**, *7*, 1231.

18. Balibar, C. J.; Walsh, C. T.; *Biochemistry* **2006**, *45*, 15444.

19. Shinoda, K.; Hasegawa, T.; Sato, H.; Shinozaki, M.; Kuramoto, H.; Takamiya, Y.; Sato, T.; Nikaidou, N.; Watanabe, T.; Hoshino, T.; *Chem. Commun.* **2007**, *40*, 4140.

20. Haack, M. B.; Eliasson, A.; Olsson, L. J.; *Biotechnol.* **2004**, *114*, 199.

21. Surribas, A.; Amigo, J. M.; Coello, J.; Montesinos, J. L.; Valero, F.; Maspoch, S.; *Anal. Bioanal. Chem.* **2006**, *385*, 1281.

22. Haack, M. B.; Lantz, A. E.; Mortensen, P. P.; Olsson, L.; *Biotechnol. Bioeng.* **2007**, *96*, 904.

23. Amigo, J. M.; Surribas, A.; Coello, J.; Montesinos, J. L.; Maspoch, S.; Valero, F.; *Chemom. Intell. Lab. Syst.* **2008**, *92*, 44.

24. Larson, T.; Wedborg, M.; Turner, D.; *Anal. Chim. Acta* **2007**, *583*, 357.

25. Thygesen, L. G.; Rinnam, A.; Barsberg, S.; Moller, J. K. S.; *Chemom. Intell. Lab. Syst.* **2004**, *71*, 97.

26. Rinnam, A.; Andersen, C. M.; *Chemom. Intell. Lab. Syst.* **2005**, *76*, 91.

27. Rinnam, A.; Booksh, K.; Bro, R.; *Anal. Chim. Acta* **2005**, *537*, 349.

28. Bouveresse, D. J. R.; Benabid, H.; Rutledge, D. N.; *Anal. Chim. Acta* **2007**, *589*, 216.

29. Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A.; *J. Chemom.* **2006**, *20*, 99.

30. http://www.models.kvl.dk/EEM_correction accessed in October 2012.

31. Bro, R.; *Chem. Intel. Lab. Syst.* **1997**, *38*, 149.

32. Andersen, C. M.; Bro, R.; *J. Chemom.* **2003**, *17*, 200.

33. Bro, R.; Kiers, H. A. L.; *J. Chemom.* **2003**, *17*, 274.

34. Bro, R.; *Chem. Intel. Lab. Syst.* **1999**, *46*, 133.

35. Bro, R.; Viereck, N.; Toft, M.; Toft, H.; Hansen, P. I.; Engelsen, S. B.; *TrAC, Trends Anal. Chem.* **2010**, *29*, 281.

36. Andersson, C. A.; Bro, R.; *Chem. Intel. Lab. Syst.* **2000**, *52*, 1.

37. http://www.model s.life.ku.dk/nwaytoolbox accessed in October 2012.

38. Tortora, G. J.; Funke, B. R.; Case, C. L.; *Microbiology: an Introduction*, 10th ed.; Benjamin Cummings: San Francisco, USA, 2009.

39. PLS Toolbox Version 3.5, Eigenvector Research Inc., Manson, WA, 2004.

40. Du, H.; Fuh, R. A.; Li, J.; Corkan, A.; Lindsey, J. S.; *Photochem. Photobiol.* **1998**, *68*, 141.

41. Asamizu, S.; Kato, Y.; Igarashi, Y.; Tomotsu, F.; Onaka, H.; *Tetrahedron Lett.* **2006**, *47*, 473.