Manuscript Details

Manuscript number	CHEMOLAB_2017_453
Title	Investigation of the photodegradation profile of tamoxifen using spectroscopic and chromatographic analysis and Multivariate Curve Resolution
Article type	Research paper

Abstract

Photodegradation of tamoxifen (TAM) is investigated by chemometric analysis of the multiset data obtained by LC-DAD/MS and UV spectrophotometry. A hydroalcoholic solution of TAM was submitted to photodegradation by means of a dedicate irradiation cabinet able to simulate natural irradiation sources. The irradiance conditions were stressed by increasing the irradiation power to produce a rapid photodegradation of TAM. Drug photodegradation was monitored through UV spectrophotometry and the obtained photoproducts were investigated in detail by DAD/MS liquid chromatography. Data collected from combination of both instrumental techniques were fused and processed jointly using the Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) method. A total number of five compounds were identified during the drug photodegradation and their kinetic evolution was described. The process included the isomerization of TAM to its (E)-form and the subsequent cyclization of these both compounds to give two phenanthrene derivatives. A photooxygenation process can also occur, giving a benzophenone derivative photoproduct as a result thereof. The multivariate resolution method proposed in this work allowed the resolution of this complex multicomponent system by the direct analysis of the experimental data.

Keywords	Tamoxifen; Drug photostability; UV Spectroscopy; LC-DAD-MS; Fused data; MCR-ALS
Taxonomy	Photodegradation, Data Fusion, Multivariate Models
Manuscript category	General Section

Corresponding Author	Roma Tauler
Corresponding Author's Institution	IDAEA-CSIC
Order of Authors	Marc Marín-García, Giuseppina Ioele, Helena Franquet-Griell, Silvia Lacorte, Gaetano Ragno, Roma Tauler
Suggested reviewers	Maria Soledad Larrechi, Renato Carneiro, Bruno Debus, Hector Goicoechea, Daniel W. Cook

Submission Files Included in this PDF

File Name [File Type]

cover_letter_CHEMOLAB_RT.docx [Cover Letter]

Highlights_RT.docx [Highlights]

TAM_manuscript_final_RT.docx [Manuscript File]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:
Data will be made available on request

Jordi Girona, 18-26
08034, Barcelona, Spain
E-mail: Roma.Tauler@idaea.csic.es
Phone: +34934006100

Romà Tauler, Dr.
Department of Environmental Chemistry
IDAEA-CSIC
Barcelona, September 21st, 2017

Dear Editor,

Attached you will find the manuscript entitled **"Investigation of the photodegradation profile of tamoxifen using spectroscopic and chromatographic analysis and Multivariate Curve Resolution."** from Marc Marín-García, Giuseppina Ioele, Helena Franquet-Griell, Sílvia Lacorte, Gaetano Ragno and Romà Tauler for publication in *Chemometrics and Intelligent Laboratory Systems.*

In this manuscript, we propose a new strategy for the chemometric analysis of the multiset data obtained by UV spectrophotometry and LC-DAD/MS from the photodegradation study of tamoxifen (TAM) drug. This approach consists of applying the Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) method to a multiset data structure. The feasibility of the proposed approach is demonstrated by its application to experimental data sets obtained in the photodegradation study of the cytostatic drug tamoxifen. Drug photodegradation was monitored through UV spectrophotometry and the obtained photoproducts were investigated in detail by DAD/MS liquid chromatography. Data collected from combination of both instrumental techniques were fused and processed jointly using the Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) method. The multivariate resolution method proposed in this work allowed the resolution of this complex multicomponent system by the direct analysis of the experimental data collected as well as the identification of the photoproducts formed and the proposal of a reaction pathway.

Finally, we suggest the following reviewers due to their expertise in chromatographic and chemometric fields:

- Dr. Daniel W. Cook
  Virginia Commonwealth University (VCU), Department of Chemistry, Richmond, VA, US
  E-mail: dwcook@vcu.edu
- Dr. Hector C. Goicoechea
  Universidad Nacional del Litoral (UNL), Analytical Development Laboratory and Chemometrics, Santa Fe de la Vera Cruz, Santa Fe, Argentina
  E-mail: hgoico@fbcb.unl.edu.ar
- Dr. Marisol Larrechi
  Universitat Rovira i Virgili (URV), Departament de Química Analítica i Química

Orgànica, Tarragona, Catalunya, Spain
E-mail: mariasoledad.larrechi@urv.cat

- Dr. Bruno Debus
  University of California (UCD), UC Davis Air Quality Research Center, Davis, CA, US
  E-mail: b.debus87@gmail.com
- Dr. Renato L. Carneiro
  Universidade Federal de São Carlos (UFSCar), Department of Chemistry, São Carlos, Estado de Sao Paulo, Brazil
  E-mail: renato.lajarim@ufscar.br

Yours sincerely,

Dr. Romà Tauler,
IDAEA-CSIC

Highlights

- Photodegradation of tamoxifen is investigated by chemometric analysis.
- Analyzed multiset data was obtained by LC-DAD/MS and UV spectrophotometry.
- A data fusion strategy and processing is proposed using the MCR-ALS method.
- Five photoproducts were identified and their kinetic evolution was described.
- A photodegradation reaction pathway is suggested.

INVESTIGATION OF THE PHOTODEGRADATION PROFILE OF TAMOXIFEN USING SPECTROSCOPIC AND CHROMATOGRAPHIC ANALYSIS AND MULTIVARIATE CURVE RESOLUTION

Marc Marín-García[1], Giuseppina Ioele[2], Helena Franquet-Griell[1], Sílvia Lacorte[1], Gaetano Ragno[2], Romà Tauler[1*]

[1] Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Catalonia, Spain.

[2] Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, 87036 Rende (CS), Italy.

**Corresponding author email Roma.Tauler@idaea.csic.es*

Abstract/Summary

Photodegradation of tamoxifen (TAM) is investigated by chemometric analysis of the multiset data obtained by LC-DAD/MS and UV spectrophotometry. A hydroalcoholic solution of TAM was submitted to photodegradation by means of a dedicate irradiation cabinet able to simulate natural irradiation sources. The irradiance conditions were stressed by increasing the irradiation power to produce a rapid photodegradation of TAM. Drug photodegradation was monitored through UV spectrophotometry and the obtained photoproducts were investigated in detail by DAD/MS liquid chromatography. Data collected from combination of both instrumental techniques were fused and processed jointly using the Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) method. A total number of five compounds were identified during the drug photodegradation and their kinetic evolution was described. The process included the isomerization of TAM to its (*E*)-form and the subsequent cyclization of these both compounds to give two phenanthrene derivatives. A photooxygenation process can also occur, giving a benzophenone derivative photoproduct as a result thereof. The multivariate resolution method proposed in this work allowed the resolution of this complex multicomponent system by the direct analysis of the experimental data.

Keywords: Tamoxifen; Drug photostability; UV Spectroscopy; LC-DAD-MS; Fused data; MCR-ALS.

1. Introduction

In the last years, the increasing production and release of synthetic organic chemical compounds into the environment, especially pharmaceutical drugs, has made that these substances have accumulated significantly on biological ecosystems as extensive and persistent pollutants, especially on surface water systems [1-6]. The current purification systems for environmental cleaning remove a considerable amount of these chemicals, but these technologies are not able to eliminate completely these compounds yet [7]. In addition, the possible negative effects for the nature are not only caused by the original compounds released into the environment, but also by their degradation products or metabolites [8]. Therefore, it is important to know how these organic pollutants can be transformed in the environment. Photodegradation due to UV-light (solar) irradiation represents one of the most important natural transformation processes of the organic products [9].

Tamoxifen (TMX or TAM), (2-{4-[(1*Z*)-1,2-diphenylbut-1-en-1-yl]phenoxy}ethyl)dimethylamine, is a non-steroidal antiestrogen drug used to prevent and treat breast cancer in men and women [10-12]. The parent compound is synthesized as a mixture of (*E*)- and (*Z*)-isomers, although the (*E*)-isomer has no clinical use and only the (*Z*) one acts as an estrogen antagonist [13, 14]. TAM has been considered one of the most studied cytostatic drugs [3] and many reviews about its presence in aquatic environment have been recently published [15, 16]. Despite its low water solubility (<1 mg/L) [17], the occurrence of this drug in environmental matrices has increased considerably in the last years [1, 6, 18]. The high octanol/water partition coefficient of the drug (log K_{ow} 7.88) indicates a great absorption affinity in fatty tissues, soil, and sediments. This behavior can cause bioaccumulation and increase the toxicity of this drug [6, 16].

Photosensitivity of TAM is well known [7, 19, 20], but only few papers describe its photodegradation profile in detail [21-23]. An elimination study of TAM using advanced oxidation processes (AOPs) as degradation method was recently published [24]. Photoisomerization, photocyclization and photooxygenation appear to be involved in the photodegradation process of this drug [21, 23, 24].

In the last few years, multicomponent systems involving equilibria and kinetic chemical processes have been investigated by combining several analytic and chemometric methods with satisfactory results [25-32]. In this work, a new analytical approach, based on the Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) technique, to monitor and investigate the photodegradation of TAM in water solution is proposed. This

approach combines UV spectrophotometric and LC-DAD-MS data with chemometric procedures. The degradation process was investigated in water after prolonged simulated sunlight irradiation. The main photoproducts, isolated by chromatographic techniques, have been identified by UV and mass spectrometry spectroscopic tools.

2. Materials and procedures

2.1 Chemicals

(*Z*)-Tamoxifen ((*Z*)-TAM) (≥99%) was purchased from Sigma-Aldrich (St Louis, USA). Methanol (MeOH) and water, both HPLC-grade (≥99.9%), were supplied by Merck-Millipore (Darmstad, Germany). To avoid light contact, all used materials were covered with aluminum foil.

2.2 Instruments

Photodegradation experiments were performed in a light cabinet SUNTEST® CPS (Atlas Material Testing Solutions, IL, USA) equipped with a xenon arc lamp of 1500 W. This system closely simulated sunlight exposure and was equipped with filters to select several spectral regions.

UV-VIS spectra were recorded using a UV-Visible molecular absorption spectrophotometer HP-Agilent 8453 (Agilent Technologies, CA, USA) with a diode array detector (DAD).

Chromatographic equipment consisted of an ultra-performance liquid chromatograph connected to a UV-Visible diode array Waters® ACQUITY® PDA Detector (Waters Corporation, MA, USA) and coupled to a benchtop triple quadrupole Waters® ACQUITY® TQ Detector (Waters Corporation, MA, USA) (UPLC-DAD-MS/MS). The analytes were separated on a 2.1 x 150 mm ID, particle size 5 μm, ZORBAX Eclipse XDB-C18 column (Agilent Technologies, CA, USA).

2.3 Sample preparation

A 100 mg/L stock solution of (*Z*)-TAM was prepared using a 50:50 methanol/water mixture as solvent, due to the poor solubility of the drug in water. This solution was diluted

to 12 mg/L (12 ppm) before every degradation experiment, using always the same solvent mixture.

2.4 Photodegradation experiments

In agreement with the rules of the International Conference on Harmonization, ICH guidelines Q1B [33], a coated quartz filter segment was interposed between the xenon arc lamp and the sample which filters the spectral range 300-800 nm. The light stressing tests were conducted at two different irradiation power conditions: approximately at 400 and 765 W/m^2, corresponding to 24 and 46 kJ/min·m^2, respectively. The inner temperature in the light cabinet was always maintained constant at 35 ºC. The samples were exposed to light in quartz cells perfectly stoppered, to avoid any evaporation of the solvent.

UV-VIS spectra were recorded at the following conditions: wavelength range 200-400 nm, scan rate 1 nm/s, time response 1 s, and spectral band 1 nm. The UV-VIS spectra were recorded just after sample preparation (t = 0), every 2 min until to 60 min, then every 5 min up to 150 min, and finally at 160 min when 400 W/m^2 irradiation power was selected. When the higher value was chosen, the experiment was stopped after 120 min, collecting spectra every 2 min.

During both experiments, 11 sample aliquots were collected from the photoreactor vessel and analyzed by LC-DAD-MS at the following exposure times: 0, 2, 6, 8, 15, 20, 40, 60, 80, 140 and 190 minutes when 400 W/m^2 irradiation power was selected and at 0, 2, 4, 8, 15, 25, 40, 50, 80, 100, 200 minutes for the 765 W/m^2 irradiation power value.

2.5 LC-DAD-MS analysis

Chromatographic analysis was performed with a 70:30 methanol/water mixture as mobile phase in isocratic elution conditions using the following instrumental parameters: injection volume 10 µL, flow rate 400 µL/min, cone voltage 55 V, spray voltage 3.5 kV, extractor voltage 3 V, desolvation gas flow 600 µL/min, source temperature 150 ºC, solvent temperature 400 ºC, acquisition rate 0.05 s (UV-DAD) and 0.21 s (MS), and time of analysis 30 min. A positive electrospray ionization source (ESI+) was used as radiation source. Acquisitions were analyzed in full scan mode. UV-DAD detection was performed in the scan range 200-400 nm. MS detector worked with the mass range 50-500 Da.

2.6 Software

UV Spectrophotometer ChemStation software (Agilent Technologies, CA, USA) and MassLynx® 4.1, from the Mass Spectrometer (Waters Corporation, MA, USA), were used for control, data acquisition and initial data preprocessing. DataBridge was the file converter provided with MassLynx® to convert LC-DAD-MS raw files (.raw) into Common Data Format files (.cdf). A Bioinformatics Toolbox MATLAB® routine (The Mathworks, Inc., MA, USA) was employed to read .cdf files format into MATLAB®. UV Spectrophotometric data were directly exported to .csv files and imported to MATLAB®. All chemometric analyses were performed under MATLAB® computer environment. MCR-ALS (2.0 GUI version [34]) method (from "http://www.mcrals.info/") was implemented as MATLAB® function.

3. Data structure

Table 1 near here

Figure 1 near here

The data sets obtained from the photodegradation experiments were arranged in different data matrices listed in **Table 1**. **Figure 1** details the matrix structure of the different data sets obtained from the different TAM exposure experiments at the UV radiation power of 400 W/m^2.

3.1 Single instrumental detection (UV or MS) data matrix arrangement

A) Data matrices $\mathbf{D}_{d400sp}$ of size (50,131) and $\mathbf{D}_{d765sp}$ of size (61,131) refer to the UV spectrophotometric analysis of TAM photodegradation at the two different irradiation power conditions: 400 and 765 W/m^2, respectively. All collected spectra were acquired at 131 wavelengths, between 210 and 340 nm. When the lower irradiation power was selected, 50 UV-VIS spectra were acquired between 0 min and 160 min of irradiation time (see **Figure 1**) while for the higher irradiation power, 61 UV-VIS spectra were acquired between 0 min and 120 min. These two data matrices have the information concerning the UV-VIS spectral evolution and concentration changes of the chemical species involved in the respective photodegradation process [28].

B) A comparison study of the photodegradation of TAM at the two irradiation power conditions was performed through the simultaneous analysis of $\mathbf{D}_{d400sp}$ and $\mathbf{D}_{d765sp}$ data matrices, which was achieved by building a column-wise augmented matrix $\mathbf{D}_{d400sp,d765sp}$ of size (61+50=111,131) (see **Figure 1**). The analysis of this augmented data matrix allows the comparison of the two experiments. In this new data arrangement the columns in each one of the two merged data matrices were the same: 131 wavelengths (210-340 nm) [28].

C) A deeper investigation of the species produced during photodegradation was achieved by chromatographic analysis on the TAM samples collected at the different irradiation times above reported. Therefore, a total number of 22 sample aliquots of the TAM solution along the photodegradation experiments (11x2) were analyzed by LC-DAD and LC-MS, giving the corresponding data matrices sets (see **Figure 1**):

- The 11 aliquots collected under the lower irradiation power and analyzed by LC-DAD gave 11 data matrices $\mathbf{D}_{400DADn}$, where n=1,...,11 sample aliquots, and, the same 11 aliquots analyzed by LC-MS, also gave 11 data matrices $\mathbf{D}_{400MSn}$ (see **Figure 1**). The rows of these data matrices contained the spectra (UV-DAD or MS) at the different elution times (394 time values between 15.7 and 22.5 min of elution time, for the aliquot at 0 min of irradiation (n=1), and 985 time values, between 13 and 30 min of elution time, for the other aliquots (n=2,...,11)). The columns of these matrices had, respectively, the chromatographic data profiles recorded at the 131 wavelengths ($\mathbf{D}_{400DADn}$ in the range 210-340 nm) and at the 351 m/z values ($\mathbf{D}_{400MSn}$ in the range 50-400 m/z) [30, 32].
- Similarly, for the other experiment carried out at the higher irradiation power, 11 data matrices $\mathbf{D}_{765DADn}$ of size (1042,131) and 11 data matrices $\mathbf{D}_{765MSn}$ of size (1042,351) were obtained. Once again, in these data matrices, the rows contained the spectrometric responses (UV-DAD or MS) at the different elution times (1042 time values in all cases, in the range 12-30 min) and the columns had the chromatographic data profiles recorded at different wavelengths ($\mathbf{D}_{765DADn}$ at the 131 wavelengths in the range 210-340 nm) or at the different m/z values ($\mathbf{D}_{765MSn}$ at the 351 m/z values in the range 50-400 m/z), respectively.

D) A more involved analysis of the photodegradation process was achieved using the complete information from the two chromatographic approaches, LC-DAD and LC-

MS, by means of the joint analysis of the 11 $\mathbf{D}_{DADn}$ and 11 $\mathbf{D}_{MSn}$ data matrices, using first each one of the two LC detection methods separately. This analysis could be achieved as shown in **Figure 1** by building the two 'tall' column-wise augmented matrices for each one of the two detection systems at both irradiation power conditions: $\mathbf{D}_{400DAD,aug}$ (of size (394+10x985,131) or (10244,131)) at 400 W/m^2 and UV-DAD detection, $\mathbf{D}_{765DAD,aug}$ (of size (11x1042,131) or (11462,131)) at 765 W/m^2 and UV-DAD detection, and $\mathbf{D}_{400MS,aug}$ (of size (394+10x985,351) or (10244,351)) at 400 W/m^2 and MS detection, and $\mathbf{D}_{765MS,aug}$ (of size (11x1042,351) or (11462,351)) at 765 W/m^2 and MS detection. These four column-wise augmented data matrices contained the complete set of chromatographic data arrays from the 11 sample aliquots (at the different photodegradation reaction times). This matrix augmentation procedure was possible because all the aliquots were analyzed by the same instrumental technique, either LC-DAD or LC-MS. In such a way, to assemble these data matrices in a column-wise way, the number and meaning of the columns in each of the data sets must be necessarily the same, either the 131 wavelengths (210-340 nm) or the 351 m/z values (50-400 m/z) [28, 30]. These four column-wise augmented matrices, $\mathbf{D}_{400DAD,aug}$, $\mathbf{D}_{765DAD,aug}$, $\mathbf{D}_{400MS,aug}$, and $\mathbf{D}_{765MS,aug}$, were first analyzed separately.

3.2 Multiple instrumental detection (UV and MS) data matrix arrangements: data fusion (UV, LC-DAD, and LC-MS data)

E) In the particular case of UV-DAD detection system, an additional data matrix arrangement was possible by considering that the photodegradation experiments monitored by UV spectrophotometry and LC-DAD runs produced individual data matrices having in common the same vector column space (wavelengths) spanned by the UV-VIS spectra of the common components [35]. In this case, the UV spectrophotometric data matrices obtained during the initial monitoring of the photodegradation reaction at both studied irradiation power conditions, $\mathbf{D}_{d400sp}$ (50,131) and $\mathbf{D}_{d765sp}$ (61,131), could be set on top of their respective column-wise augmented LC-DAD data matrices: $\mathbf{D}_{400DAD,aug}$ (10244,131) and $\mathbf{D}_{765DAD,aug}$ (11462,131) (see **Figure 1**). These two new column-wise augmented matrices will be named $\mathbf{D}_{d400sp,400DAD,aug}$ (of size (50+394+10x985,131) or (10294,131)) and $\mathbf{D}_{d765sp,765DAD,aug}$ (of size (61+11x1042,131) or (11523,131)).

F) Additionally, it was still possible to consider the possibility of the joint chemometric analysis of the 11 aliquots using simultaneously the two detection systems, UV-DAD and MS, in each chromatographic run. This would imply performing a different type of matrix augmentation. Since the same samples were analyzed by both detection systems, the data matrix augmentation could be performed also row-wisely (see **Figure 1**) [36]. In this case, a new superaugmented data matrix was obtained for each irradiation power condition: $\mathbf{D}_{400DAD,400MS,supaug}$ of size (10244,482) and $\mathbf{D}_{765DAD,765MS,supaug}$ of size (11462,482) for 400 and 765 W/m^2, respectively, with 482 columns each one (131 wavelengths + 351 m/z values). In these superaugmented data matrices, the rows represented the UV-VIS and MS spectra at different elution times.

4. Data pretreatment and analysis

4.1 Data synchronization and preprocessing

In all chromatographic experiments, the elution time window between 13 and 30 min was selected and further processed by MCR. In the chromatographic elution before 13 min no chemical species of interest were detected; only very low intensity signals, mostly due to the solvent and instrumental noise (background contributions) were present. After 30 min, no further elution of sample components was observed. Noise contributions were removed from the selected elution time window by applying a Savitzky-Golay smoothing data filter [37]. The final chromatographic data had then a better signal to noise quality.

Due to the different frequency of the two chromatographic detectors (UV-DAD and MS) in spectra acquisition, as above described, a data pretreatment was additionally needed to check for the correspondence in time of the elution profiles in both detectors and for the further simultaneous chemometric analysis (data fusion). Since the spectra acquisition speed of the DAD system was faster than the spectra acquisition of the MS detection system, a higher number of UV than MS spectra were acquired. To match the two detection systems, a linear interpolation (and smoothing) was used to synchronize UV-DAD and MS detector signals at the same time frequencies. Moreover, since the two detection systems were in tandem, a brief time delay on peak signal recording occurred, due to the transfer tubing between UV-DAD (first) and MS (second) detectors. This time delay was estimated to be approximately 0.2020 min (12 s) and required the time axis shifting of MS data to previously acquired UV-DAD time scale.

4.2 MCR-ALS: data analysis method

MCR-ALS is a mixture analysis chemometric method which decomposes in a bilinear way an experimental data matrix, **D**, having the analytical mixed responses of multiple components, into their pure contributions (profiles) in its two data modes [38]:

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

In this data decomposition, **C** is a matrix with the concentration profiles of the mixture constituents, $\mathbf{S}^{\mathbf{T}}$ matrix has their pure spectra profiles, and **E** has the residual variance not explained by the MCR model, $\mathbf{CS}^{\mathbf{T}}$. This equation is an extension of the Beer-Lambert's law to multi-wavelength and multi-sample analysis and summarizes the set of linear equations defining the concentration and spectral contributions of each component in the chemical mixture system [25]. Next section describes the application of the MCR-ALS bilinear method to the analysis of the different data matrices generated during the photodegradation study (see **Figure 2**):

Figure 2 near here

1) The spectrophotometric monitoring of the photodegradation experiments provided a data matrix **D** (either $\mathbf{D}_{d400sp}$ or $\mathbf{D}_{d765sp}$ in **Section 3.1 A)**) where rows are the spectra collected at different reaction times and columns are kinetic traces at different wavelengths [26, 28]. The MCR bilinear model applied to them is defined, as explained before, by **Equation 1** and shown graphically in **Figure 2a**. **C** and $\mathbf{S}^{\mathbf{T}}$ matrices give the concentration and UV-VIS spectra profiles of the components (chemical species) formed during the UV-light irradiation experiments, respectively. When the simultaneous analysis of $\mathbf{D}_{d400sp}$ and $\mathbf{D}_{d765sp}$ data matrices is done to compare the photodegradation of TAM drug solutions at the two studied irradiation power conditions, the resulting **C** matrix is a column-wise augmented matrix ($\mathbf{C}_{\mathbf{aug}}$) which contains the concentration profiles of the chemical species formed during each one of the UV-light irradiation experiments, and $\mathbf{S}^{\mathbf{T}}$ is a matrix that has their related UV-VIS spectra, common in both photodegradation experiments.

2) The chromatographic analysis of the set of aliquots extracted from the degradation experiments gives, for every sample, a data matrix ($\mathbf{D}_{400DADn}$ or $\mathbf{D}_{765DADn}$ in LC-DAD

and $\mathbf{D}_{400MSn}$ or $\mathbf{D}_{765MSn}$ in LC-MS, where n=1,...,11 sample aliquots) that can also be analyzed by MCR-ALS using a bilinear model similar to that shown in **Equation 1** and **Figure 2a**. One of these data matrices (see **Section 3.1 C)**) has one single chromatographic run and it contains in its rows the UV-VIS or MS spectra recorded at the different elution times and in its columns the chromatograms related to different wavelengths or different m/z values [30, 39]. **C** and $\mathbf{S}^{\mathbf{T}}$ matrices give the concentration (elution) and UV-VIS or MS spectra profiles of the components (species) present in the aliquot at the considered reaction (degradation) time.

3) The set of data matrices at the different chromatographic runs (n=1,...,11) obtained by the LC-DAD and LC-MS analysis of the 11 reaction sample aliquots can be merged, for both irradiation power conditions, in column-wise augmented data matrices $\mathbf{D}_{\mathbf{aug}}$, as it was previously described in **Section 3.1 D)** ($\mathbf{D}_{400DAD,\mathbf{aug}}$, $\mathbf{D}_{765DAD,\mathbf{aug}}$, $\mathbf{D}_{400MS,\mathbf{aug}}$, and $\mathbf{D}_{765MS,\mathbf{aug}}$) and shown in **Figure 1** for 400 W/m^2. The application of MCR-ALS to these column-wise augmented data matrices is described by **Equation 2** below and it is shown graphically in **Figure 2b** for the case of the LC-DAD detection:

$$\mathbf{D}_{\mathbf{aug}} = [\mathbf{D}_1; \mathbf{D}_2; ...; \mathbf{D}_n] = [\mathbf{C}_1; \mathbf{C}_2; ...; \mathbf{C}_n]\mathbf{S}^{\mathbf{T}} + [\mathbf{E}_1; \mathbf{E}_2; ...; \mathbf{E}_n] = \mathbf{C}_{\mathbf{aug}}\mathbf{S}^{\mathbf{T}} + \mathbf{E}_{\mathbf{aug}}, \quad \textbf{Equation 2}$$

where $\mathbf{C}_{\mathbf{aug}} = [\mathbf{C}_1; \mathbf{C}_2; ...; \mathbf{C}_n]$, with n=1,...,11 chromatographic runs, is a column-wise augmented concentration matrix formed by the $\mathbf{C}_n$ submatrices containing the pure elution profiles of the individual species involved in each chromatographic run, and $\mathbf{S}^{\mathbf{T}}$ is a matrix that has their related UV-VIS or MS spectra, common to all chromatographic runs of all sample aliquots simultaneously analyzed [28, 30, 31]. From the resolved elution profiles, their peak areas can be calculated as a function of the reaction time, and the corresponding kinetic profiles of the components involved in the photodegradation process can be derived. In the case of MS detection, these components can be identified from their resolved mass spectra.

4) The set of chromatographic runs using the LC-DAD system can be also simultaneously analyzed with the matrix obtained in the UV-Visible spectroscopic monitoring of the degradation experiments, as it was explained previously in **Section 3.2 E)** and shown in **Figure 1** for 400 W/m^2 ($\mathbf{D}_{d400sp,400DAD,\mathbf{aug}}$ and $\mathbf{D}_{d765sp,765DAD,\mathbf{aug}}$ augmented data matrices) [32]. The MCR-ALS analysis of these two augmented data matrices can be

also analyzed by the bilinear model in **Equation 2**. In this case, moreover, the concentration submatrix corresponding to the UV-Visible spectroscopic monitoring experiment provides the kinetic reaction profiles of the resolved components (see **Figure 2c**). This analysis will provide the correspondence between the reaction species, their chromatographic elution responses, and their corresponding kinetic profiles and UV-VIS spectra.

5) Finally, it is indeed also possible (as already shown in **Section 3.2 F)** and **Figure 1** for the superaugmented data matrices $\mathbf{D}_{400DAD,400MS,supaug}$ and $\mathbf{D}_{765DAD,765MS,supaug}$) the simultaneous chemometric analysis of the different aliquots using both detection systems (UV-DAD and MS). This implies performing the MCR-ALS analysis on the superaugmented data matrix ($\mathbf{D}_{supaug}$). This analysis of the 'fused' UV-DAD and MS data by the bilinear model is described in **Equation 3** below and in **Figure 2d**, where $\mathbf{C}_{aug}$ is the column-wise augmented concentration matrix formed by the $\mathbf{C}_n$ submatrices containing the pure elution profiles of the individual species involved in each chromatographic run, common for both detection systems, and $\mathbf{S^T}_{aug}$ is the row-wise augmented matrix with their related UV-VIS ($\mathbf{S}_{DAD}^{T}$) and MS spectra ($\mathbf{S}_{MS}^{T}$), common to all chromatograms analyzed [32]:

$$\mathbf{D}_{supaug} = \left[\left[\mathbf{D}_{1,DAD}; \mathbf{D}_{2,DAD};\ldots;\mathbf{D}_{n,DAD}\right],\left[\mathbf{D}_{1,MS}; \mathbf{D}_{2,MS};\ldots;\mathbf{D}_{n,MS}\right]\right] = \left[\mathbf{C}_1; \mathbf{C}_2;\ldots;\mathbf{C}_n\right] \left[\mathbf{S}_{DAD}^{T},\mathbf{S}_{MS}^{T}\right] + \left[\left[\mathbf{E}_{1,DAD}; \mathbf{E}_{2,DAD};\ldots;\mathbf{E}_{n,DAD}\right],\left[\mathbf{E}_{1,MS}; \mathbf{E}_{2,MS};\ldots;\mathbf{E}_{n,MS}\right]\right] = \mathbf{C}_{aug}\mathbf{S}_{aug}^{T} + \mathbf{E}_{supaug} \quad \textbf{Equation 3}$$

This analysis requires the time synchronization between UV-DAD and MS detectors (see **Section 4.1**). In this simultaneous analysis, besides, from the resolved elution profiles in $\mathbf{C}_{aug}$ matrix, their peak areas can be calculated as a function of the reaction time, and the corresponding kinetic profiles of the components involved in the photodegradation process can be derived.

MCR-ALS algorithm uses an iterative Alternating Least Squares (ALS) optimization procedure for the decomposition of the data matrix, under a suitable set of constraints. This optimization runs until the model $\mathbf{CS^T}$ minimizes as much as possible the error in the reproduction of the original data set, **D** [32]. MCR-ALS requires the initial postulation of the number of components which can be derived from the results of the Singular Value Decomposition (SVD) [40] of the original data matrix, **D**. In the case of photodegradation

studies, however, where the reaction pathway is often unknown and different photoproducts can be generated, the determination of the number of species involved in the process by SVD is sometimes difficult due to linear dependence in the concentration profiles of the reaction products [25, 27]. Initial estimates, either of $\mathbf{S^T}$ or **C** matrices, needed to start the iterative ALS optimization can be obtained from purest experimental spectra or purest elution profiles respectively, using a similar procedure than for the SIMPLISMA method [41, 42].

MCR-ALS constraints are used to provide meaningful shapes to the profiles in **C** and $\mathbf{S^T}$ matrices and to suppress or minimize as much as possible the ambiguity in the final solutions [32]. Constraints are chemical or mathematical properties that the pure component profiles should accomplish. They can be applied in a different way to the concentration and spectral directions (**C** and $\mathbf{S^T}$ matrices), to the profiles of the different components and to the different submatrices in a multiset structure [31]. The constraints used in this photodegradation study were: *non-negativity* (NN), which avoids the presence of negative values in the concentration and spectral profiles; *unimodality* (U), which allows only the presence of a single maximum per concentration/elution profile; *closure* (C), which forces concentration profiles within the closed system to add up to a certain constant value to fulfill the mass balance condition; and *selectivity* or *local rank information*, with which some of the components may be forced to be absent at some time/elution or spectral ranges [38]. In multiset data structures, the constraint named *correspondence of species* acts as a selectivity/local rank constraint setting the absence of some compounds in full concentration submatrices of the augmented $\mathbf{C_{aug}}$ matrix [36]. More details about the different steps of the MCR-ALS procedure can be found in the literature [26, 38].

The parameters used to indicate the fit quality of the MCR-ALS results are the percentage of lack of fit (% lof) and the explained variance (R^2), which are defined as follows:

$$\%\, lof = 100 \cdot \sqrt{\frac{\sum_{ij}\left(d_{ij} - d_{ij}^{*}\right)^2}{\sum_{ij} d_{ij}^2}}$$

$$R^2 = 100 \cdot \frac{\sum_{ij} d_{ij}^2 - \sum_{ij}\left(d_{ij} - d_{ij}^{*}\right)^2}{\sum_{ij} d_{ij}^2}$$

where d_{ij} is the matrix element in the row i and column j from the original data matrix (**D**) and d_{ij}^{*} is the same element obtained with the MCR-ALS model [30, 43].

5. Results and discussion

5.1 Analysis of TAM photodegradation UV spectrophotometric data

Figure 3 near here

TAM photodegradation process was firstly monitored by UV-Visible spectroscopy. **Figure 3** (on the left) shows the evolution of the acquired spectra along time for the two experiments performed with the SUNTEST® at 400 and 765 W/m^2. The plot of these two data matrices ($\mathbf{D}_{d400sp}$ and $\mathbf{D}_{d765sp}$) provided an initial view of the evolution of the photodegradation process. It can be observed that during the first part of the experiments the spectra variation was larger than afterwards, during the remaining process. This fact suggests that major chemical transformations occur at the beginning of the photodegradation process which evolves more slowly thereafter.

MCR-ALS analysis of the UV-Visible spectrophotometric data from these two experiments ($\mathbf{D}_{d400sp}$ and $\mathbf{D}_{d765sp}$ in **Table 1**) provided a first estimation of the kinetic profiles (concentration profiles, **C** matrix in **Equation 1** and **Figure 2a**) for the species formed during the photodegradation process, and also of the pure UV-VIS spectra associated with them ($\mathbf{S}^{\mathbf{T}}$ matrix in **Equation 1** and **Figure 2a**). Initial SVD analysis showed the possible presence of at least four species during the photodegradation process. Initial estimates of their spectra were obtained from the purest experimental spectra. Constraints used during the ALS optimization were non-negativity of pure spectra and of kinetic profiles, and selectivity at the starting point of the photodegradation experiment (initial (*Z*)-TAM is the only component at the starting time conditions of the photodegradation experiment). Closure constraint was also applied to kinetic profiles to ensure a mass balance equation. Explained variances and lack of fit values (in %) of the MCR-ALS analyses at the two irradiation power conditions are given below in **Table 2**. In both cases, excellent data fits were obtained.

Table 2 near here

Figure 3 also shows the concentration profiles (kinetic profiles, **C** matrix) and the pure UV-VIS spectra ($\mathbf{S^T}$ matrix) resolved by MCR-ALS in these two photodegradation experiments. During the experiment under 400 W/m^2, tamoxifen ((*Z*)-TAM), is represented by a blue line. After 2 min of UV-light irradiation, its concentration decreased to half its initial concentration (12 ppm) and continued being reduced until disappearing afterwards. Simultaneously, a first transformation product (TP1, green line in **Figure 3**) was rapidly formed and its kinetic profile decreased and finally disappeared. A new photoproduct (TP2, red line in **Figure 3**) appeared at the beginning of the experiment reaching its maximum concentration level at 90 min at 400 W/m^2 and decreased thereafter. Concurrently with the decrease of TP1, a new photoproduct (TP3, cyan line in **Figure 3**) appears, with a very similar spectrum to TP2, with increasing concentration until the end of the experiment. When the irradiation power was increased to 765 W/m^2, the photodegradation process occurred more quickly, but essentially, it followed the same pathway, as it can also be seen in **Figure 3**.

Despite this useful information about the degradation kinetics of TAM obtained by UV-Visible spectroscopy, only a rough description of the process was achieved, because of the likely problems associated with the resolution of kinetic processes [32, 44]. Chemical species with similar kinetics and photoproducts with very similar UV-VIS spectra would not be easily distinguished using this approach. Moreover, photoproducts not giving UV-Visible absorption would not be detected. The complementary use of the LC-DAD-MS powerful analytical methodology can provide a deeper insight on the distinct species formed during the photodegradation process and of their reaction pathway.

5.2 Analysis of TAM photodegradation LC-DAD and LC-MS data

11 sample aliquots of the TAM solution were collected and analyzed by LC-DAD-MS at various times along the photodegradation reaction, for each experiment at the two studied irradiation power conditions. To improve the resolution of the coeluted chromatographic peaks, MCR-ALS was applied to the data sets obtained using UV-DAD and MS full scan detection methods ($\mathbf{D}_{400DADn}$, $\mathbf{D}_{765DADn}$, $\mathbf{D}_{400MSn}$ and $\mathbf{D}_{765MSn}$ data matrices in **Table 1**) as above explained. Results from LC-MS were used for the identification and confirmation of the formed photoproducts.

Table 2 summarizes MCR-ALS results obtained in the individual and simultaneous analysis of all the analyzed aliquots ($\mathbf{D}_{400DADn}$ and $\mathbf{D}_{400MSn}$ with n=1,...,11 sample aliquots), using the irradiation power of 400 W/m^2, both for LC-DAD and LC-MS. Constraints used

in these MCR-ALS analyses were non-negativity of **C** and $\mathbf{S}^{\mathbf{T}}$ profiles, and unimodality and selectivity of **C** profile. These two last constraints were especially useful to resolve the species profiles of some of the photoproducts and to disregard background contributions. The number of components considered in each case varied from 2 (at reaction time 0 min) to 7 (at the latest reaction time values). The presence of strong baseline and background contributions were modelled with the presence of two extra MCR-ALS components which allowed a better data fit and improved the resolution of the four reaction photoproducts and of initial (*Z*)-TAM. Therefore, five of the seven components resolved by MCR-ALS, including initial TAM, were finally assigned to the investigated photodegradation products (see below). Lack of fit values were always good, between 1-3% in the case of LC-DAD data, and between 2-5% in the case of LC-MS data.

Figure 4 near here

Results of the simultaneous MCR-ALS analysis of the complete set of chromatographic runs, at the irradiation power of 400 W/m^2 and using only one of the two detection systems ($\mathbf{D}_{400DAD,\mathbf{aug}}$ and $\mathbf{D}_{400MS,\mathbf{aug}}$ augmented data matrices, see **Equation 2**), are given in **Table 2**. In this case, MCR-ALS analysis gives the $\mathbf{C}_{\mathbf{aug}}$ matrix, which has the elution profiles of all resolved species formed during the photodegradation process separately for each chromatographic run, and the $\mathbf{S}^{\mathbf{T}}$ matrix which has their corresponding pure UV-VIS or MS spectra, which are the same for the same species in the different chromatographic runs (see **Figure 2b**). Constraints used in this MCR-ALS analysis were: non-negativity of **C** and $\mathbf{S}^{\mathbf{T}}$ and unimodality of **C**. Correspondence between species in the different runs [36] was also constrained to set their presence/absence in the different chromatographic runs of the augmented $\mathbf{C}_{\mathbf{aug}}$ concentration matrix.

Figure 4 shows $\mathbf{C}_{\mathbf{aug}}$ and $\mathbf{S}^{\mathbf{T}}$ matrices for the resolution of $\mathbf{D}_{400DAD,\mathbf{aug}}$, where the blue component is assigned to the initial tamoxifen ((*Z*)-TAM), and the green component to the first transformation product immediately formed after initial light exposure (TP1). Due to the similarity between the UV-VIS spectra of these two components, it can be assumed that the green component should have a very similar structure to tamoxifen, indicating therefore that it could be its isomer ((*E*)-TAM). Its MS spectrum can confirm this hypothesis and also previous results in the literature [23]. TP2 (yellow) and TP3 (magenta) components in **Figure 4** should correspond to the red and cyan components in **Figure 3**, obtained in the analysis of the UV-Visible spectroscopic monitoring experiments. These two species have been now correctly differentiate each other due to their slightly shifted

chromatographic elution profiles, although their UV-VIS spectra were very similar. The resolution of their MS spectra will confirm them and give more information about the identity of these two photoproducts. A new last species (TP4) can be assigned to the red component, which appears after 6 min of light irradiation and disappears by the end of the photodegradation process. This component, however, was not detected in the UV-Visible spectroscopic monitoring experiments.

Figure 5 near here

From the changes of the peak areas of the resolved elution profiles of the different components with time (see **Figure 2b**), a rough estimation of the kinetic profiles of the species detected by LC-DAD and resolved by MCR-ALS can be derived, as is shown in **Figure 5** for the 400 W/m^2 experiment. The concentration of the initial (*Z*)-TAM species decreases gradually until it disappears before the end of the photodegradation experiment. Green species (TP1), which was formed instantly at the beginning of the UV-light irradiation, reaches its maximum at 8 min reaction time, and disappears after 140 minutes of light irradiation. Yellow and magenta species (TP2 and TP3, respectively), evolve from the beginning and are still present at the end of the photodegradation experiment. TP2 seems to appear first and reaches a higher concentration than TP3. The last red component (TP4) (not shown in **Figure 5**) seems to be formed from a secondary reaction because it appeared during the photodegradation process and disappeared before it ended.

Resolution of $\mathbf{D}_{400MS,\mathbf{aug}}$ LC-MS augmented matrix was more challenging. The presence of isomers among the obtained photoproducts ((*Z*)-TAM and TP1, and TP2 and TP3), as confirmed in the next section, implies that these species will have the same molecular mass (same molecular ion $[M+H]^+$) and, probably, similar fragmentation. Unimodality and selectivity constraints, as discussed above, are especially useful to resolve the profiles of isomeric photoproducts and to disregard background contributions. In fact, their application allowed the proper resolution of these compounds in the UV-DAD detector, due to their slightly shifted chromatographic elution and to minor differences in their UV-VIS spectra (see **Figure 4**). However, in the case of using only the MS detector, these species could not be completely resolved.

5.3 Simultaneous analysis of TAM photodegradation UV spectrophotometric and LC-DAD data

Simultaneous MCR-ALS analysis of the UV spectroscopic monitoring experiments (matrices $\mathbf{D}_{d400sp}$ and $\mathbf{D}_{d765sp}$) and of the LC-DAD chromatographic data sets grouped in the corresponding augmented matrices ($\mathbf{D}_{400DAD,aug}$ and $\mathbf{D}_{765DAD,aug}$) was performed by creating the new column-wise augmented data matrices $\mathbf{D}_{d400sp,400DAD,aug}$ and $\mathbf{D}_{d765sp,765DAD,aug}$, as explained in **Section 3.2 E)** and shown in **Figure 1** for 400 W/m^2. In this case (see **Equation 2**), the new $\mathbf{C}_{aug}$ augmented matrix has the kinetic traces of the species formed during the photodegradation process and their elution profiles in each chromatographic run (see **Figure 2c**). However, the $\mathbf{S}^T$ matrix has the common pure UV-VIS spectra of the species in both, in the photodegradation experiments and chromatographic runs. A clear relationship between the reaction species, their kinetic profiles, their corresponding chromatographic elution profiles, and their UV-VIS spectra is obtained. **Table 2** also gives a summary of the results of this MCR-ALS analysis for the lower irradiation power condition experiment. Constraints used in this MCR-ALS analysis were the same as those previously applied in the resolution of LC-DAD data matrices. A MCR-ALS model with a total number of four species plus (*Z*)-TAM was used (two extra MCR-ALS components were assigned to baseline and background contributions). In this case, the $\mathbf{C}_{aug}$ and $\mathbf{S}^T$ matrices were practically the same as those already shown in **Figure 4**.

According to the results obtained until now, the pure compound (*Z*)-TAM (blue) and its initial transformation product (TP1, green) have a slightly different absorbance spectrum: (*Z*)-TAM exhibits a strong UV absorption band at 277 nm with a tail at wavelengths over 310 nm and another absorption band at 236 nm. TP1, instead, does not present this band at 236 nm and only exhibits a UV band at 277 nm. This is in agreement with the UV-VIS spectra found in the literature [23]. Therefore, the first transformation photoproduct could be (*E*)-TAM, the isomer of (*Z*)-TAM initial compound, which is immediately formed when UV-light was irradiated. On the other hand, TP2 and TP3 photoproducts exhibit rather similar spectra too, although it was possible to differentiate two characteristic bands in their respective absorbance spectra. The component designed as TP3 (magenta) has absorption bands at 236, 255, 280, and over 300 nm, whereas the compound designed as TP2 (yellow) has only the absorption bands at 255 and 300 nm. This last species has also an unusual valley between 220 and 240 nm. In the literature [23], UV-VIS spectra with these slight differences correspond to Phenanthrene I and II, respectively. TP4 (red) shows a strong UV absorption band near 280 nm. No information

about this UV-VIS spectrum for TP4 was found in the literature. As it will be shown below, the MS spectrum obtained for this species allows for its possible identification.

Based on the time evolution of the kinetic profiles resolved in the kinetic and chromatographic experiments (see **Figure 5**), a model with two parallel reactions can be then proposed. (*Z*)-TAM is firstly very fast transformed to give its isomer, (*E*)-TAM, which stands probably in equilibrium, and thereafter both isomers photodegraded to give their derivatives Phenanthrene I (TP2) and II (TP3), respectively. On the other hand, TP4 seems to be formed from a secondary reaction. The full reaction pathway could be:

$$(Z)\text{-TAM (blue)} \xrightarrow{h\nu} (E)\text{-TAM (TP1, green)} \xrightarrow{h\nu} \text{Phenanthrene II (TP2, yellow)}$$

$$(Z)\text{-TAM (blue)} \xrightarrow{h\nu} \text{Phenanthrene I (TP3, magenta)}$$

$$(Z)\text{- or } (E)\text{-TAM} \xrightarrow{h\nu,\ O_2} \text{?? (TP4, red)}$$

According to this proposal, the photodegradation pathway for the photodegradation of TAM in solution implies that (*Z*)-TAM, when exposed to UV-light, gives a reaction intermediate (excited state) which undergoes an isomerization reaction or a photocyclization reaction, giving (*E*)-TAM and Phenanthrene I, respectively. Moreover, when the isomer (*E*)-TAM is excited again by UV-light, another reaction intermediate is reached and another photocyclization takes place, giving Phenanthrene II [23]. Finally, according to the literature, it is postulated that when oxygen is present, these reaction intermediates could interact with oxygen molecules and produce the unknown photoproduct (TP4) [21].

5.4 Simultaneous analysis of TAM photodegradation LC-DAD-MS data and final photoproduct identification

In order to identify and confirm the proposed photoproduct species described above, LC-DAD and LC-MS data matrices ($\mathbf{D}_{DAD,\mathbf{aug}}$ and $\mathbf{D}_{MS,\mathbf{aug}}$) were fused (merged) and analyzed simultaneously, also at both irradiation power conditions, using the row- and column-wise superaugmented data matrices ($\mathbf{D}_{\text{supaug}}$) (see **Equation 3**). Resolved elution profiles of all species formed during the photodegradation process were obtained separately for each chromatographic run ($\mathbf{C}_{\text{aug}}$), and the pure UV-VIS and MS spectra of all these species were arranged in the new row-wise augmented matrix of the pure DAD and MS spectra ($\mathbf{S}^{\mathbf{T}}_{\mathbf{aug}}$) (see **Figure 2d**), which are common to all chromatographic runs.

This approach is very powerful and describes the entire system with even higher reliability. **Table 2** shows the MCR-ALS results of the analysis of the $\mathbf{D}_{supaug}$ data matrix using the lower irradiation power condition experiment ($\mathbf{D}_{400DAD,400MS,supaug}$). Constraints used in this MCR-ALS analysis are given in the table. As in the LC-DAD analysis for the 400 W/m^2 experiment, only five among the seven MCR-ALS resolved components were finally assigned to the investigated photodegradation pathway. The two extra MCR-ALS components were needed to model the presence of the baseline and the background contributions. The results obtained for the 765 W/m^2 experiment matrix ($\mathbf{D}_{765DAD,765MS,supaug}$) were very close to those obtained from the 400 W/m^2 experiment. Although, in this case, a sixth extra species was recovered by MCR, but it was not further characterized because no information about a pure MS spectrum containing m/z 367 as molecular ion was found in the literature.

Figure 6 near here

Figure 6 shows the pure mass spectra ($\mathbf{S}^T_{MS}$ submatrix) resolved by MCR-ALS for the lower irradiation power condition LC-DAD-MS monitoring experiment. In ESI+ mode, the protonated molecular ion $[M+H]^+$ was formed for both parental compounds and photoproducts. MS spectra of the first and second MCR-ALS components (blue and green) confirmed that TP1 should be the (*E*)-isomer of TAM, with the same base peak at m/z 372 $[M+H]^+$ and with identical molecular formula ($C_{26}H_{29}NO$), but different m/z fragmentation ratio of the confirmation ion (m/z 72), assigned to dimethylaminopropylene ion produced by cleavage of the side-chain [21, 23, 24]. This different m/z fragmentation ratio allowed distinguishing between these two compounds. The UV-VIS spectra obtained in this analysis ($\mathbf{S}^T_{DAD}$ submatrix) were practically identical to those obtained previously in the preceding LC-DAD data analysis (see **Figure 4**).

Despite of having the same molecular ion $[M+H]^+$ at m/z 370 and very similar UV-VIS spectra [21, 23, 24], photoproducts TP2 and TP3 were identified as Phenanthrene II and I, respectively. They gave a different fragmentation pattern, due to a loss of a methyl group in the case of Phenanthrene I (-CH_3, 15 Da). Thus, the cleavage of the common molecular ion of the photoproducts produced, for Phenanthrene II, a fragment ion at m/z 72. In contrast, m/z 58 fragmentation seemed to be more favorable for Phenanthrene I. Both photoproducts have, however, the same molecular formula $C_{26}H_{27}NO$.

According to the literature [45], *cis–trans* photoisomerization is a typical alkene photochemical fast reaction, and the formation of products such as Phenanthrene I or II due

to a photocyclization reaction can be observed in the photochemistry of stilbene-type compounds. In fact, the photocyclization reaction proceed only by absorption of a photon of UV-light by the *cis*-isomer, but not by the *trans*-isomer of the stilbene derivative [45]. This implies that the synthetically more accessible and stable (*Z*)-TAM (compound used as cytostatic drug), which is a *trans*-isomer stilbene derivative, suffers initially a reversible *cis-trans* photoisomerization, to give the mechanistically required *cis*-isomer ((*E*)-TAM) for starting the photocyclization [45]. Because of this regioselectivity preference, it is observed that (see **Figure 5**), in the photodegradation of TAM drug, Phenanthrene II (TP2, yellow), which comes from the *cis*-isomer ((*E*)-TAM), appeared before and reached higher concentration than Phenanthrene I (TP3, magenta), which is formed from the other isomer. Photocyclization reaction, in fact, occurs via a dihydrophenanthrene intermediate which may be undetected especially under oxidative conditions [21, 45]. Therefore, when TAM is UV-light irradiated, due to the presence of geminal phenyl rings, both isomers are capable of dehydrogenation leading to two possible phenanthrene products. This kind of reactions (pericyclic reactions) are an important part of the stilbene chemistry and, specifically, photocyclization is very useful in synthetic routes as final step to generate a fused aromatic ring at a benzylic position.

Photodegradation of TAM in presence of oxygen, according to the literature, also produces a photooxygenation reaction with the formation of a benzophenone derivative with a molecular ion $[M+H]^+$ at m/z 270 and molecular formula $C_{17}H_{19}NO_2$ [21, 24]. This benzophenone derivative can be assigned to the TP4 (red component). Fragment ions at m/z 72, 105 and 211 from this photoproduct were also obtained and confirm this identification. Its pure UV-VIS spectrum was already given (see **Figure 4**). Most probably, photooxygenation was, in this case, a secondary reaction occurring during the sample aliquots removing process, due to the presence of atmospheric oxygen in the solution. This photoproduct did not appear, indeed, during the irradiation experiments continuously monitored by UV spectroscopic because the UV measurement cuvette was always stopped/closed with little air/oxygen interaction. Photooxygenation leading to the ketone derivative is a minor reaction which appears not to involve singlet oxygen reactive species ($^1O_2$). More probably, the excited molecules of TAM were trapped by diradical ground-state molecular oxygen ($^3O_2$) [21]. Extraction of β-hydrogen and shift of the double bond would lead to a hydroperoxide intermediate which, probably via a Hock-cleavage, would give this benzophenone derivative [21]. The formation of this by-product can reduce the useful UV-light flux for the light-dependent reactions occurring in the photodegradation

process, since benzophenone and, most likely, its derivative as well, acts as a filter for UV radiation and it is able to absorb this radiation and dissipate it as heat.

6. Conclusions

The photodegradation of the antiestrogen drug TAM in aqueous solution was investigated in detail by a combination of spectrometric and chromatographic techniques and chemometric data analysis. Five different photoproducts with different kinetic profiles were resolved and identified.

For all the components, their pure UV-VIS and MS spectra and kinetic profile were estimated. Partial MS fragmentation, using high cone voltage, of the obtained products confirmed the proposed structures. The photodegradation pathway of TAM showed a first isomerization of the drug followed by a cyclization reaction of both isomers. An additional photoproduct can be formed when the isomers are excited by UV-light in presence of oxygen through a photooxygenation reaction. No evidence of a different photodegradation pathway was observed when a higher irradiation power was selected, only an increase in the degradation rate was detected.

MCR-ALS method was able to resolve the mixture of products formed during the photodegradation reaction despite their incomplete chromatographic separation at isocratic conditions. The simultaneous analysis of fused DAD-MS data from these two different instrumental techniques provided a better resolution of the species formed during the photodegradation process. Due to slightly shifted chromatographic elution, minor differences in UV-VIS spectra, and different m/z fragmentation ratio, isomeric compounds ((*Z*)-TAM and (*E*)-TAM, and Phenanthrene I and II, respectively) were finally correctly identified.

Acknowledgments

A research grant from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) for project CTQ2015-66254-C2-1-P is acknowledged by the authors as financial support. Marc Marín-García also acknowledges the FPI grant BES-2016-076678 from MINECO.

References

718 [1] J.P. Besse, J.F. Latour, J. Garric, Anticancer drugs in surface waters. What can we say
719 about the occurrence and environmental significance of cytotoxic, cytostatic and endocrine
720 therapy drugs?, Environment International, 39 (2012) 73-86.
721 [2] H. Franquet-Griell, C. Gómez-Canela, F. Ventura, S. Lacorte, Predicting
722 concentrations of cytostatic drugs in sewage effluents and surface waters of Catalonia (NE
723 Spain), Environmental Research, 138 (2015) 161-172.
724 [3] C. Gómez-Canela, F. Ventura, J. Caixach, S. Lacorte, Occurrence of cytostatic
725 compounds in hospital effluents and wastewaters, determined by liquid chromatography
726 coupled to high-resolution mass spectrometry, Analytical and Bioanalytical Chemistry,
727 406 (2014) 3801-3814.
728 [4] A. Nikolaou, S. Meric, D. Fatta, Occurrence patterns of pharmaceuticals in water and
729 wastewater environments, Analytical and Bioanalytical Chemistry, 387 (2007) 1225-1234.
730 [5] S.D. Richardson, T.A. Ternes, Water analysis: emerging contaminants and current
731 issues, Anal Chem, 86 (2014) 2813-2848.
732 [6] J. Zhang, V.W. Chang, A. Giannis, J.Y. Wang, Removal of cytostatic drugs from
733 aquatic environment: a review, Sci Total Environ, 445-446 (2013) 281-298.
734 [7] P.H. Roberts, K.V. Thomas, The occurrence of selected pharmaceuticals in wastewater
735 effluent and surface waters of the lower Tyne catchment, Science of The Total
736 Environment, 356 (2006) 143-153.
737 [8] T. Kosjek, E. Heath, B. Kompare, Removal of pharmaceutical residues in a pilot
738 wastewater treatment plant, Analytical and Bioanalytical Chemistry, 387 (2007) 1379-
739 1387.
740 [9] C.J. Halsall, Environmental Organic Chemistry, in: R. Harrison (Ed.) Principles of
741 Environmental Chemistry, Royal Society of Chemistry, Cambridge, 2007, pp. 279-313.
742 [10] M. Clemons, S. Danson, A. Howell, Tamoxifen ('Nolvadex'): a review: Antitumour
743 treatment, Cancer Treatment Reviews, 28 (2002) 165-180.
744 [11] V. Craig Jordan, The role of tamoxifen in the treatment and prevention of breast
745 cancer, Current Problems in Cancer, 16 (1992) 134-176.
746 [12] V.C. Jordan, Tamoxifen (ICI46,474) as a targeted therapy to treat and prevent breast
747 cancer, Br J Pharmacol, 147 Suppl 1 (2006) S269-276.
748 [13] R.K. Pandey, R.D. Wakharkar, P. Kumar, Wittig–Horner Approach for the Synthesis
749 of Tamoxifen, Synthetic Communications, 35 (2005) 2795-2800.
750 [14] D.W. Robertson, J.A. Katzenellenbogen, Synthesis of the (E) and (Z) isomers of the
751 antiestrogen tamoxifen and its metabolite, hydroxytamoxifen, in tritium-labeled form, The
752 Journal of Organic Chemistry, 47 (1982) 2387-2393.
753 [15] L. Ferrando-Climent, S. Rodriguez-Mozaz, D. Barceló, Incidence of anticancer drugs
754 in an aquatic urban system: From hospital effluents through urban wastewater to natural
755 environment, Environmental Pollution, 193 (2014) 216-223.
756 [16] N. Negreira, M.L. de Alda, D. Barceló, Cytostatic drugs and metabolites in municipal
757 and hospital wastewaters in Spain: Filtration, occurrence, and environmental risk, Science
758 of The Total Environment, 497–498 (2014) 68-77.
759 [17] C. Gómez-Canela, N. Cortés-Francisco, F. Ventura, J. Caixach, S. Lacorte, Liquid
760 chromatography coupled to tandem mass spectrometry and high resolution mass
761 spectrometry as analytical tools to characterize multi-class cytostatic compounds, Journal
762 of Chromatography A, 1276 (2013) 78-94.
763 [18] T. Kosjek, E. Heath, Occurrence, fate and determination of cytostatic pharmaceuticals
764 in the environment, TrAC Trends in Analytical Chemistry, 30 (2011) 1065-1087.
765 [19] Z. Chen, G. Park, P. Herckes, P. Westerhoff, Physicochemical treatment of three
766 chemotherapy drugs: Irinotecan, tamoxifen, and cyclophosphamide, J. Adv. Oxid.
767 Technol., 11 (2008) 254-260.

[20] T. Kojima, S. Onoue, F. Katoh, R. Teraoka, Y. Matsuda, S. Kitagawa, M. Tsuhako, Effect of spectroscopic properties on photostability of tamoxifen citrate polymorphs, International Journal of Pharmaceutics, 336 (2007) 346-351.
[21] M. DellaGreca, M.R. Iesce, M. Isidori, A. Nardelli, L. Previtera, M. Rubino, Phototransformation products of tamoxifen by sunlight in water. Toxicity of the drug and its derivatives on aquatic organisms, Chemosphere, 67 (2007) 1933-1939.
[22] N. Negreira, J. Regueiro, M. López de Alda, D. Barceló, Transformation of tamoxifen and its major metabolites during water chlorination: Identification and in silico toxicity assessment of their disinfection byproducts, Water Research, 85 (2015) 199-207.
[23] J. Šalamoun, M. Macka, M. Nechvátal, M. Matoušek, L. Knesel, Identification of products formed during UV irradiation of tamoxifen and their use for fluorescence detection in high-performance liquid chromatography, Journal of Chromatography A, 514 (1990) 179-187.
[24] L. Ferrando-Climent, R. Gonzalez-Olmos, A. Anfruns, I. Aymerich, L. Corominas, D. Barceló, S. Rodriguez-Mozaz, Elimination study of the chemotherapy drug tamoxifen by different advanced oxidation processes: Transformation products and toxicity assessment, Chemosphere, 168 (2017) 284-292.
[25] M. De Luca, G. Ioele, S. Mas, R. Tauler, G. Ragno, A study of pH-dependent photodegradation of amiloride by a multivariate curve resolution approach to combined kinetic and acid-base titration UV data, Analyst, 137 (2012) 5428-5435.
[26] M. De Luca, S. Mas, G. Ioele, F. Oliverio, G. Ragno, R. Tauler, Kinetic studies of nitrofurazone photodegradation by multivariate curve resolution applied to UV-spectral data, International Journal of Pharmaceutics, 386 (2010) 99-107.
[27] M. De Luca, G. Ragno, G. Ioele, R. Tauler, Multivariate curve resolution of incomplete fused multiset data from chromatographic and spectrophotometric analyses for drug photostability studies, Anal Chim Acta, 837 (2014) 31-37.
[28] M. De Luca, R. Tauler, G. Ioele, G. Ragno, Study of photodegradation kinetics of melatonin by multivariate curve resolution (MCR) with estimation of feasible band boundaries, Drug Testing and Analysis, 5 (2013) 96-102.
[29] C. Gómez-Canela, G. Bolivar-Subirats, R. Tauler, S. Lacorte, Powerful combination of analytical and chemometric methods for the photodegradation of 5-Fluorouracil, Journal of Pharmaceutical and Biomedical Analysis, 137 (2017) 33-41.
[30] A. Jayaraman, S. Mas, R. Tauler, A. de Juan, Study of the photodegradation of 2-bromophenol under UV and sunlight by spectroscopic, chromatographic and chemometric techniques, Journal of Chromatography B, 910 (2012) 138-148.
[31] S. Mas, A. de Juan, S. Lacorte, R. Tauler, Photodegradation study of decabromodiphenyl ether by UV spectrophotometry and a hybrid hard- and soft-modelling approach, Analytica Chimica Acta, 618 (2008) 18-28.
[32] S. Mas, R. Tauler, A. de Juan, Chromatographic and spectroscopic data fusion analysis for interpretation of photodegradation processes, Journal of Chromatography A, 1218 (2011) 9260-9268.
[33] Q1B Stability Testing: Photostability Testing of New Drug Substances and Products, in: I.C.o.H. (ICH) (Ed.), 1996.
[34] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemometrics and Intelligent Laboratory Systems, 140 (2015) 1-12.
[35] R. Tauler, Multivariate curve resolution applied to second order data, Chemometrics and Intelligent Laboratory Systems, 30 (1995) 133-146.
[36] R. Tauler, M. Maeder, A. de Juan, Multiset Data Analysis: Extended Multivariate Curve Resolution, in: T.R. Brown S., Walczak R. (Ed.) Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 473-505.
[37] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry, 36 (1964) 1627-1639.

[38] A. de Juan, S.C. Rutan, R. Tauler, Two-Way Data Analysis: Multivariate Curve Resolution – Iterative Resolution Methods, in: T.R. Brown S., Walczak R. (Ed.) Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 325-344.
[39] A. de Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures, Analytica Chimica Acta, 500 (2003) 195-210.
[40] G.H. Golub, Van Loan, C.F., Matrix Computations, 2nd ed., The John Hopkins University Press, London, 1989.
[41] W. Windig, J. Guilment, Interactive Self-Modeling Mixture Analysis, Analytical Chemistry, 63 (1991) 1425-1432.
[42] W. Windig, S. Markel, Simple-to-use interactive self-modeling mixture analysis of FTIR microscopy data, Journal of Molecular Structure, 292 (1993) 161-170.
[43] R. Tauler, M. Maeder, Two-Way Data Analysis: Multivariate Curve Resolution – Error in Curve Resolution, in: T.R. Brown S., Walczak R. (Ed.) Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 345-363.
[44] A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou, R. Tauler, Second-order multivariate curve resolution applied to rank-deficient data obtained from acid-base spectrophotometric titrations of mixtures of nucleic bases, Chemometrics and Intelligent Laboratory Systems, 38 (1997) 183-196.
[45] F.B. Mallory, C.W. Mallory, Photocyclization of Stilbenes and Related Molecules, in: W.G.e.a. Dauben (Ed.) Organic Reactions, John Wiley & Sons, Inc., Bryn Mawr College, Bryn Mawr, Pennsylvania, 1984, pp. 1-456.

843 **Table 1.** Names and dimensions of the different data matrices obtained in the TAM
844 photodegradation experiments.

W/m^2	Data matrix	Nr. Rows (Reaction or elution time values)	Nr. Columns (Wavelengths or m/z)
400	$\mathbf{D}_{d400sp}$	50	131
	$\mathbf{D}_{400DADn}$ [a]	985[b]	131
	$\mathbf{D}_{400DAD,aug}$	10244	131
	$\mathbf{D}_{d400sp,400DAD,aug}$	10294	131
	$\mathbf{D}_{400MSn}$ [a]	985[b]	351
	$\mathbf{D}_{400MS,aug}$	10244	351
	$\mathbf{D}_{400DAD,400MS,supaug}$	10244	482
765	$\mathbf{D}_{d765sp}$	61	131
	$\mathbf{D}_{765DADn}$ [a]	1042	131
	$\mathbf{D}_{765DAD,aug}$	11462	131
	$\mathbf{D}_{d765sp,765DAD,aug}$	11523	131
	$\mathbf{D}_{765MSn}$ [a]	1042	351
	$\mathbf{D}_{765MS,aug}$	11462	351
	$\mathbf{D}_{765DAD,765MS,supaug}$	11462	482
400 and 765	$\mathbf{D}_{d400sp,d765sp}$	111	131

[a] n=1,...,11 sample aliquots (chromatographic runs)

[b] 394 time values, between 15.7-22.5 min of elution time, for 0 min of UV-light irradiation reaction time (n=1), and 985 time values, between 13-30 min of elution time, for the remainder aliquots collected (n=2,…,11)

845
846

847 **Table 2.** Summary of the MCR-ALS results obtained in the analysis of the data matrices
848 from the different TAM photodegradation experiments.

Matrix	Reaction time (min)	Number of components (NC)	Constraints	Lack of fit (%)	Explained variance (%)
$\mathbf{D}_{d400sp}$	**0-160**	4	NN, C, Sel	0.53	99.99
$\mathbf{D}_{d765sp}$	**0-120**	4	NN, C, Sel	0.28	99.99
$\mathbf{D}_{400DADn}$ n=1	**0**	2	NN, U, Sel	0.71	99.91
	2	3		0.99	99.99
	6	5		0.69	99.99
	8	5		0.82	99.99
	15	6		0.92	99.99
	20	6		1.17	99.98
	40	5		0.85	99.99
	60	6		1.14	99.98
	80	7		0.73	99.99
	140	6		0.65	99.99
n=11	**190**	5		0.73	99.99
$\mathbf{D}_{400DAD,aug}$	**0-190**	7	NN, U, Corr	6.05	99.63
$\mathbf{D}_{d400sp,400DAD,aug}$	**0-160, 0-190**	7	NN, U, Corr	9.94	99.01
$\mathbf{D}_{400MSn}$ n=1	**0**	2	NN, U, Sel	2.02	99.96
	2	3		1.85	99.97
	6	4		1.80	99.96
	8	5		1.88	99.96
	15	4		2.27	99.95
	20	5		1.84	99.96
	40	6		1.68	99.97
	60	6		2.36	99.94
	80	6		3.13	99.90
	140	6		2.88	99.91
n=11	**190**	6		3.16	99.89
$\mathbf{D}_{400MS,aug}$	**0-190**	8	NN, U, Corr	17.71	96.86
$\mathbf{D}_{400DAD,400MS,supaug}$	**0-190**	7	NN, U, Corr	5.40	99.71

849 **NN**: Non-negativity, **U**: Unimodality, **C**: Closure, **Sel**: Selectivity, **Corr**: Correspondence of
850 species

851

Figure captions

Fig. 1. Matrix arrangements of the different data sets obtained in the TAM photodegradation experiment at 400 W/m^2 and analyzed by MCR-ALS.

Fig. 2. MCR-ALS bilinear models for the analysis of the different data matrices generated from the experimental data collected during TAM photodegradation study: **a)** for the UV spectrophotometric monitoring experiment or for a single chromatographic run, **b)** for the simultaneous analysis of the set of data matrices at the different chromatographic runs (n=1,...,11) using the UV-DAD detector, **c)** for the simultaneous analysis of the UV spectroscopic monitoring experiment and of the whole set of the LC-DAD chromatographic runs, and **d)** for the simultaneous chemometric analysis of the different sample aliquots using both chromatographic detection systems (UV-DAD and MS).

Fig. 3. Spectral evolution along time for the UV spectrophotometric monitoring of TAM photodegradation experiments at 400 and 850 W/m^2 ($\mathbf{D_d}$ matrices) and their MCR-ALS resolved kinetic profiles ($\mathbf{C}$) and pure UV-VIS spectra ($\mathbf{S^T}$).

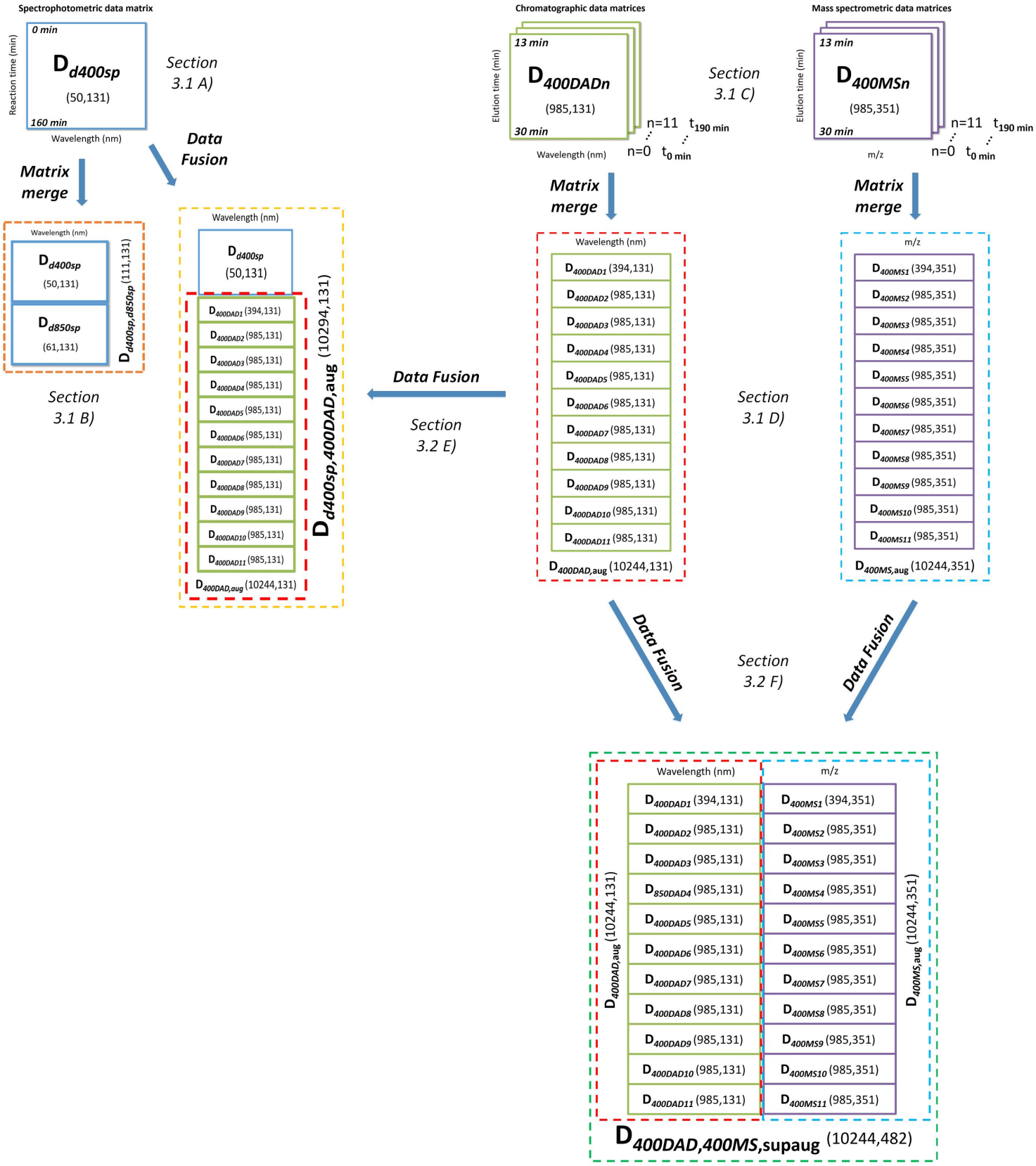
Fig. 4. Elution profiles ($\mathbf{C_{aug}}$) and pure UV-VIS spectra ($\mathbf{S^T}$) resolved by simultaneous MCR-ALS analysis of the LC-DAD data of the 11 removed sample aliquots from 400 W/m^2 TAM photodegradation experiment ($\mathbf{D_{\mathit{400DAD}, aug}}$).

Fig. 5. Evolution of the LC-DAD chromatographic peaks area as a function of the reaction time (kinetic profiles) from $\mathbf{D_{\mathit{400DAD}, aug}}$. Data points are the markers and lines are the kinetic evolution estimations.

Fig. 6. Identification of photoproducts from their pure MS spectra ($\mathbf{S^T_{\mathit{MS}}}$ submatrix) resolved by simultaneous MCR-ALS analysis of the fused HPLC-DAD-MS data at 400 W/m^2 TAM photodegradation experiment ($\mathbf{D_{\mathit{400DAD, 400MS}, supaug}}$).

881 **Figure 1**

882

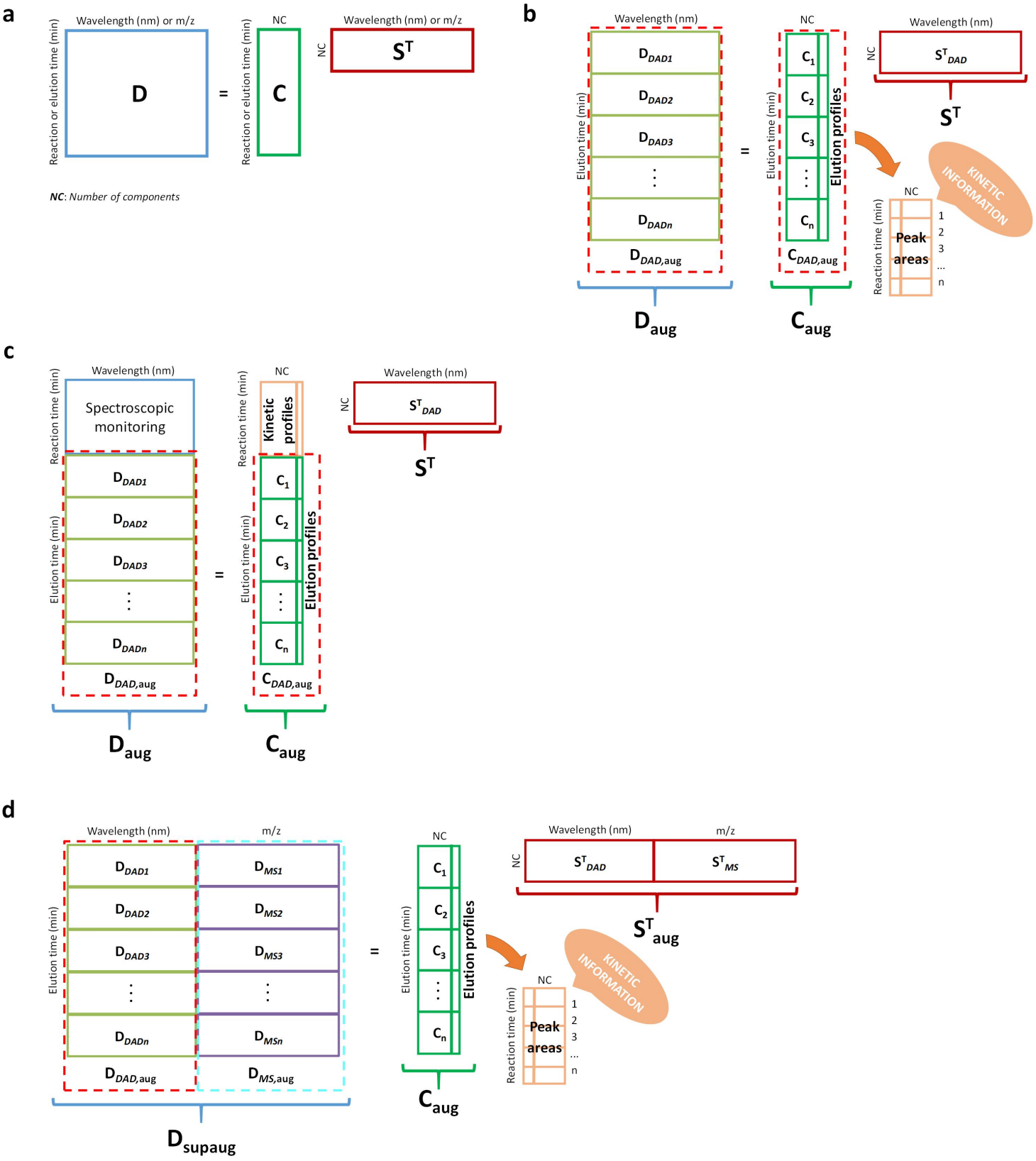


883

884

885 **Figure 2**

886

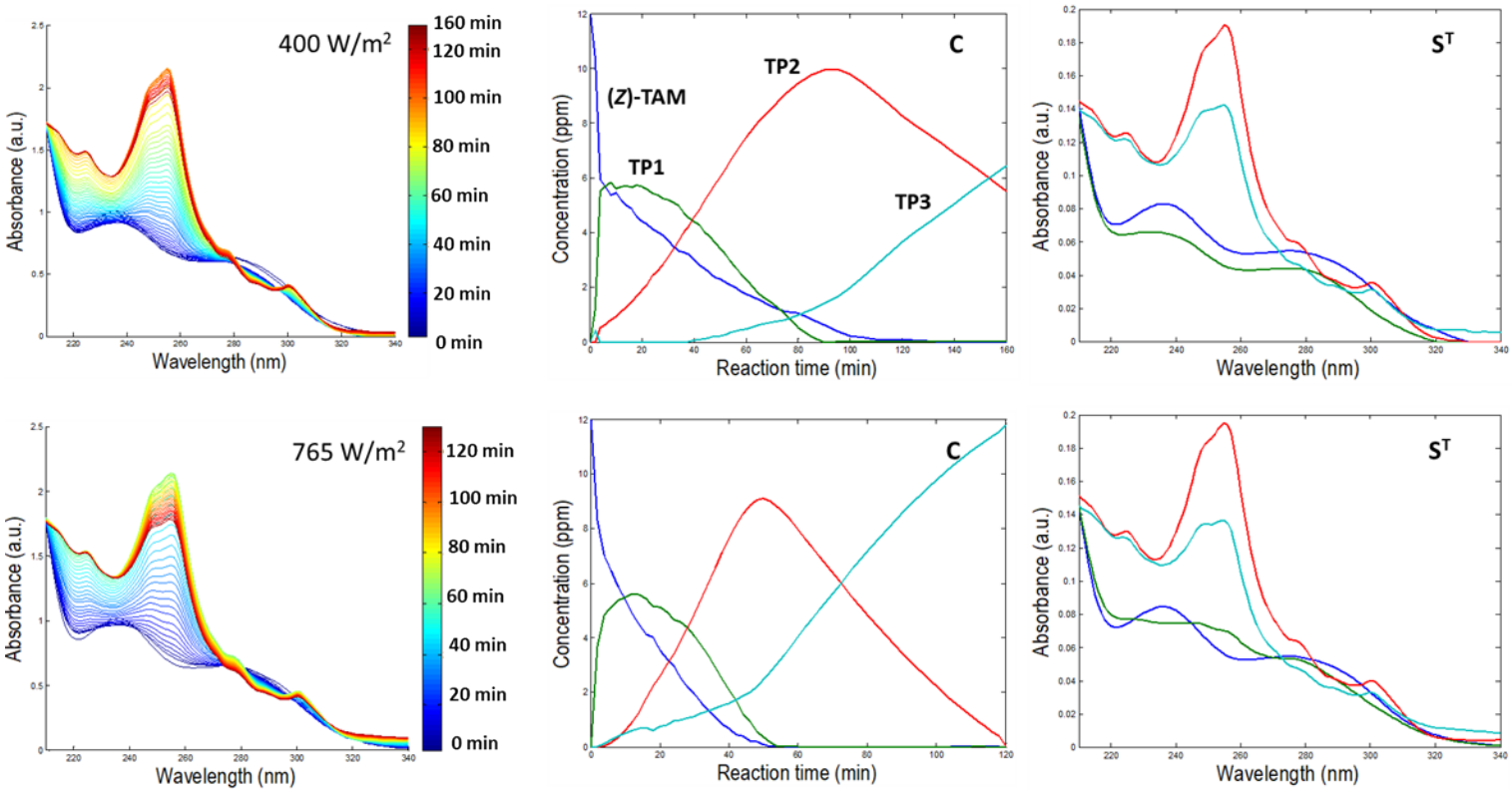


887

888

889 **Figure 3**

890



891

892

Figure 4

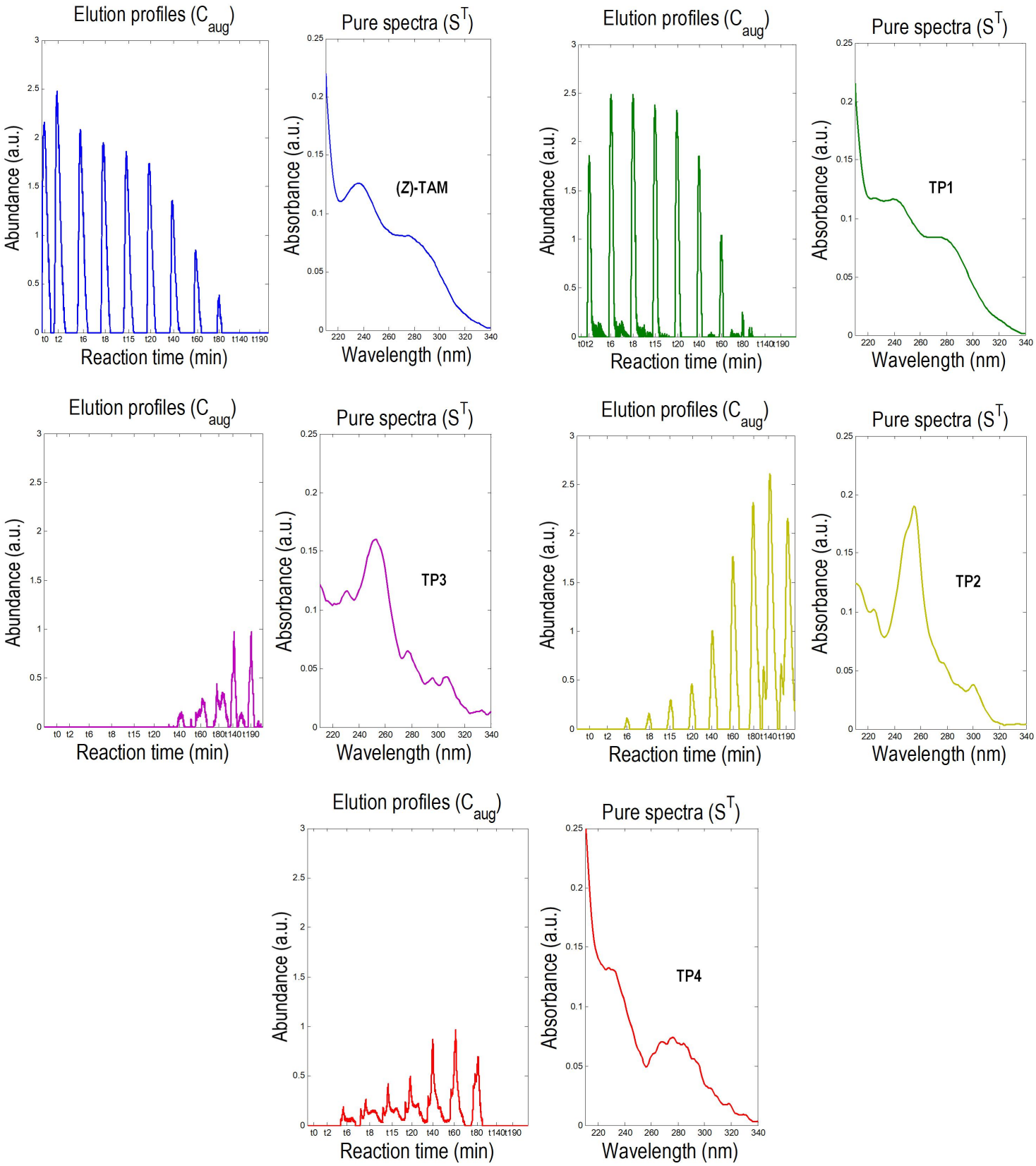
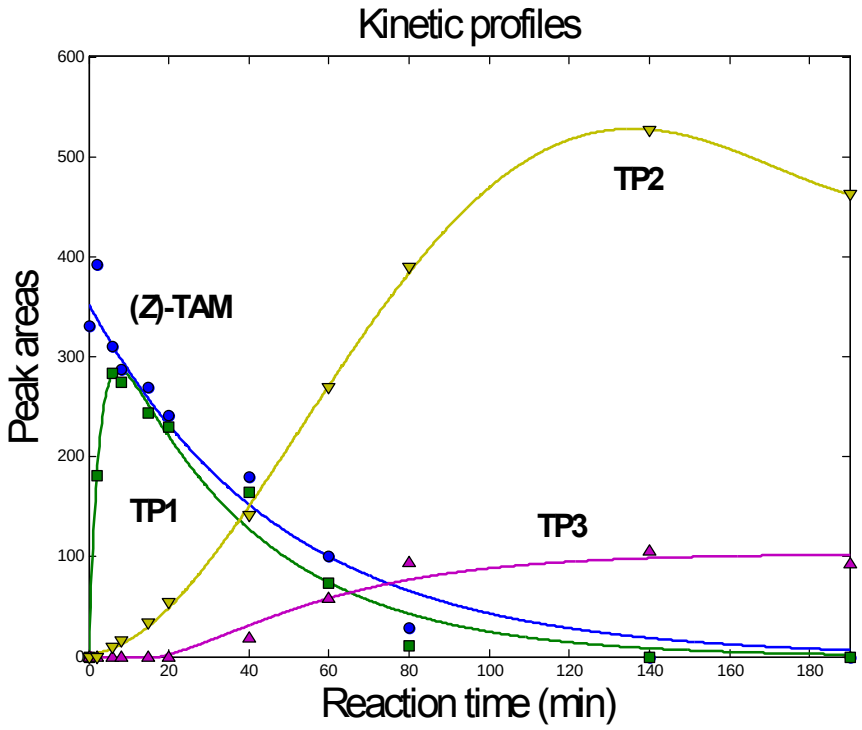
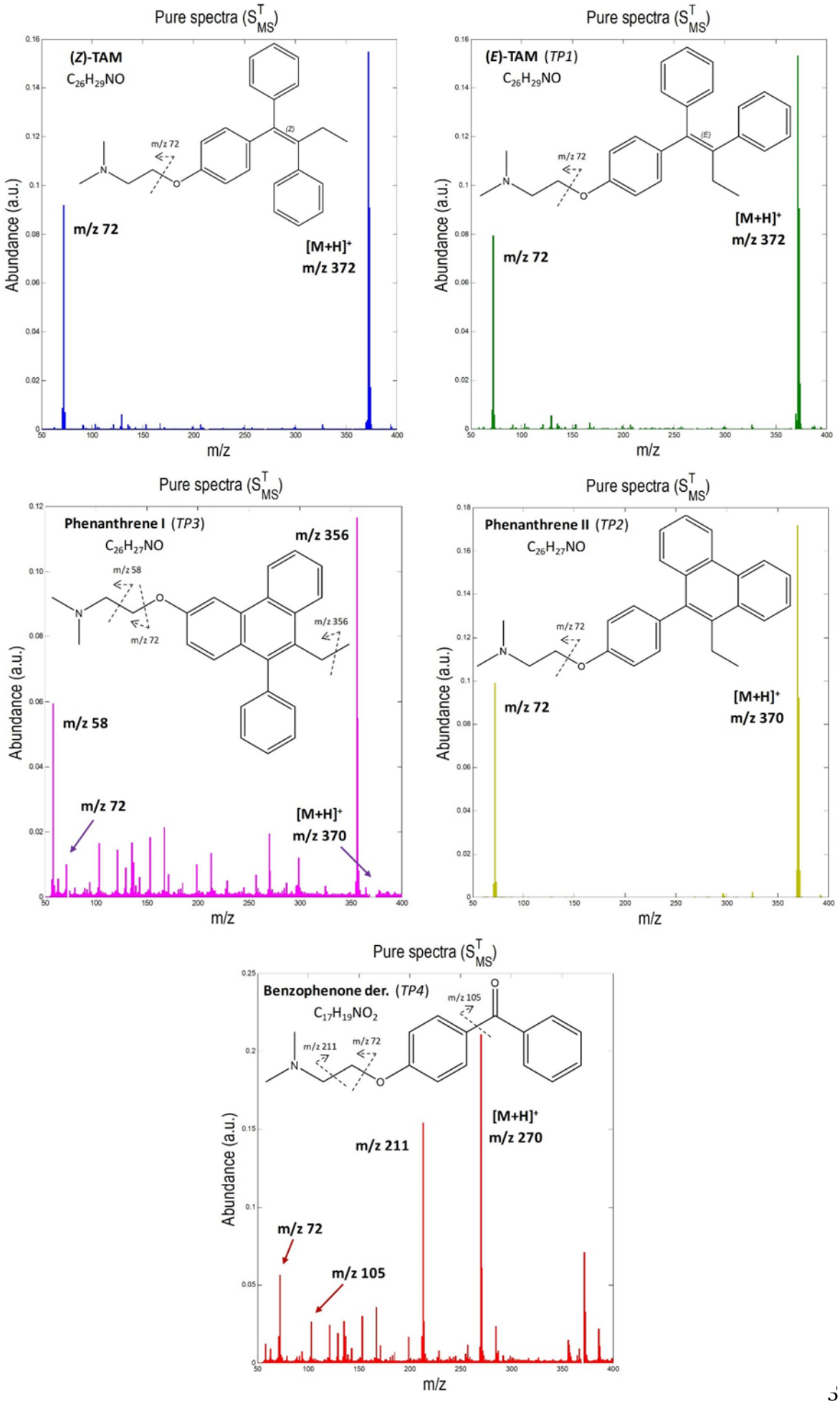

Elution profiles (C aug)
Abundance (a.u.)
Reaction time (min)
t0 t2 t6 t8 t15 t20 t40 t60 t80 t140 t190
Pure spectra (S T)
Absorbance (a.u.)
Wavelength (nm)
(Z)-TAM
TP1
TP3
TP2
TP4

Figure 5


Kinetic profiles
Peak areas
Reaction time (min)
(Z)-TAM
TP1
TP2
TP3
0
100
200
300
400
500
600
0
20
40
60
80
100
120
140
160
180

901 **Figure 6**


Pure spectra (S^T_{MS})
(Z)-TAM
$C_{26}H_{29}NO$
m/z 72
m/z 72
[M+H]+
m/z 372
Abundance (a.u.)
m/z
Pure spectra (S^T_{MS})
(E)-TAM (TP1)
$C_{26}H_{29}NO$
m/z 72
[M+H]+
m/z 372
m/z 72
Abundance (a.u.)
m/z
Pure spectra (S^T_{MS})
Phenanthrene I (TP3)
$C_{26}H_{27}NO$
m/z 356
m/z 58
m/z 72
m/z 356
m/z 58
m/z 72
[M+H]+
m/z 370
Abundance (a.u.)
m/z
Pure spectra (S^T_{MS})
Phenanthrene II (TP2)
$C_{26}H_{27}NO$
m/z 72
m/z 72
[M+H]+
m/z 370
Abundance (a.u.)
m/z
Pure spectra (S^T_{MS})
Benzophenone der. (TP4)
$C_{17}H_{19}NO_2$
m/z 105
m/z 211
m/z 72
[M+H]+
m/z 270
m/z 211
m/z 72
m/z 105
Abundance (a.u.)
m/z