# PARAFAC Modeling of Fluorescence Excitation−Emission Spectra of Fish Bile for Rapid En Route Screening of PAC Exposure

JAN H. CHRISTENSEN,*,† GIORGIO TOMASI,† JAKOB STRAND,‡ AND OLE ANDERSEN§

*Department of Basic Sciences and Environment, the Faculty of Life Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark, Department of Marine Ecology, National Environmental Research Institute, University of Aarhus, Frederiksborgvej 399, 4000 Roskilde, Denmark, and Department of Science, Systems and Models, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark*

*Received January 19, 2009. Revised manuscript received April 9, 2009. Accepted April 28, 2009.*

Polycyclic aromatic compound (PAC) metabolites in fish bile can be used as biomarkers for recent environmental exposure to PACs. Here, a novel method for rapid screening of nonhydrolyzed fish bile is presented. The method is based on excitation−emission fluorescence spectroscopy combined with parallel factor analysis (PARAFAC) and may constitute an alternative to fixed wavelength fluorescence and synchronous fluorescence spectroscopy (SFS). PARAFAC was applied to excitation−emission matrices (EEMs) of bile samples of shorthorn sculpins and European eels collected in Greenland and Denmark. The EEMs were decomposed into a four-factor PARAFAC model. The comparison of the PARAFAC factors with the EEMs of PAC metabolites and amino acids suggests that two factors are related to PAC metabolites and two correspond to fluorescent residues of tryptophan and tyrosine in bile proteins. A new standardization procedure based on the mean of the scores for the biological factors was used to correct for feeding status and sample dilution and, upon such normalization, the score plots of PARAFAC factors showed a clear distinction between exposed and nonexposed fish. A good correlation was found between the factor scores and 1-hydroxypyrene equivalents determined by SFS for high contamination levels, whereas the sensitivity was better for the EEM method.

## 1. Introduction

Polycyclic aromatic compounds (PACs) are natural constituents of crude oil and petroleum products, and they are also formed by incomplete combustion of organic materials during human activities and natural processes (*1*). They can enter the aquatic environment via several routes, where atmospheric wet and dry deposition, surface runoffs and petroleum spills are the most prominent and ubiquitous sources (*2*). The measurement of tissue concentrations in invertebrates and vertebrates is a direct measure of exposure of marine organisms to these organic contaminants, although the relationship between exposure and tissue concentrations in aquatic organisms is highly dependent on the rate of uptake, biotransformation, and excretion for the individual species (*3*). Since invertebrates like bivalves have low biotransformation rates, they are the preferred pollution indicator organisms in environmental surveys (*3*). In contrast, fish and other marine vertebrates have a more efficient metabolic system for xenobiotics, which enables them to efficiently metabolize and excrete PACs. Specifically, PACs are rapidly metabolized to epoxy or hydroxy derivatives (phase-I metabolism), which are further transformed into highly water-soluble conjugates (phase-II metabolism) that are stored in the gallbladder before excretion from the organism (*4*). Therefore, while tissue levels of parent compounds are unlikely to reflect exposure of fish (*3, 5*) concentrations of PAC metabolites in their bile fluid are better and more direct indicators of recent PAC uptake, having been used in several studies employing various analytical approaches (*6, 7*). Modern biomonitoring of PACs in marine vertebrates often consists of two-tier analytical approaches. Namely, an initial screening of a large number of samples to obtain a semiquantitative measure of the level of PAC exposure followed by a more in-depth study of selected samples where individual metabolites are separated, identified, and quantified to gain more detailed information on the level and type of exposure (*8*). The detailed analysis is normally based on high-performance liquid chromatography with fluorescence detection (HPLC-F) or gas chromatography with mass spectrometric detection (GC-MS) of hydrolyzed bile (the latter requires also derivatization) followed by quantification of individual hydroxylated PACs (*9*−*11*). Furthermore, high resolution liquid chromatography with tandem mass spectrometric detection (LC-MS/MS) has been proposed for separation and identification of phase-II metabolites in nonhydrolyzed bile (*8*).

Fixed wavelength fluorescence spectroscopy (FF), synchronous fluorescence spectroscopy (SFS) and HPLC-F, are the preferred methods for initial screening and semiquantitative analysis of 2−5 ring PAC metabolites because the methods are rapid, easy-to-use and can be applied directly to nonhydrolyzed bile (*8, 12, 13*). The FF and SFS approaches are less time-consuming and can easily be implemented en-route. Hence, since they have been found to correlate well with the HPLC-F results, they are often preferred for biomonitoring purposes (*13, 14*). However, bile fluids contain compounds of biological origin and other aromatic contaminants that fluoresce at the selected wavelength pairs and in the synchronous scan regions used for PAC metabolite quantification in the FF and SFS methods, respectively (*8, 14*). These interferences make the FF and SFS approaches prone to systematic errors as it is often not possible to separate contributions from PAC metabolites from those of the interfering compounds.

Full-scan excitation−emission spectra (also known as EEMs, excitation−emission matrices, or fluorescence landscapes) contain more chemical information than FF and SFS (see Figure 1), but so far, they have not been used for PAC metabolite quantification. Multiway methods such as parallel factor (PARAFAC) analysis of EEMs can be used to resolve mixtures of fluorescent compounds into their individual constituents, thereby reducing the interference problem and allowing a more accurate quantification (*15*−*17*).

The objective of this study was therefore to develop a novel method for en-route rapid screening of nonhydrolyzed

* Corresponding author phone: +45-35332456; fax: +45-35332398; e-mail: jch@life.ku.dk.
† University of Copenhagen.
‡ University of Aarhus.
§ Roskilde University.

10.1021/es900187x CCC: $40.75 
Published on Web 05/13/2009

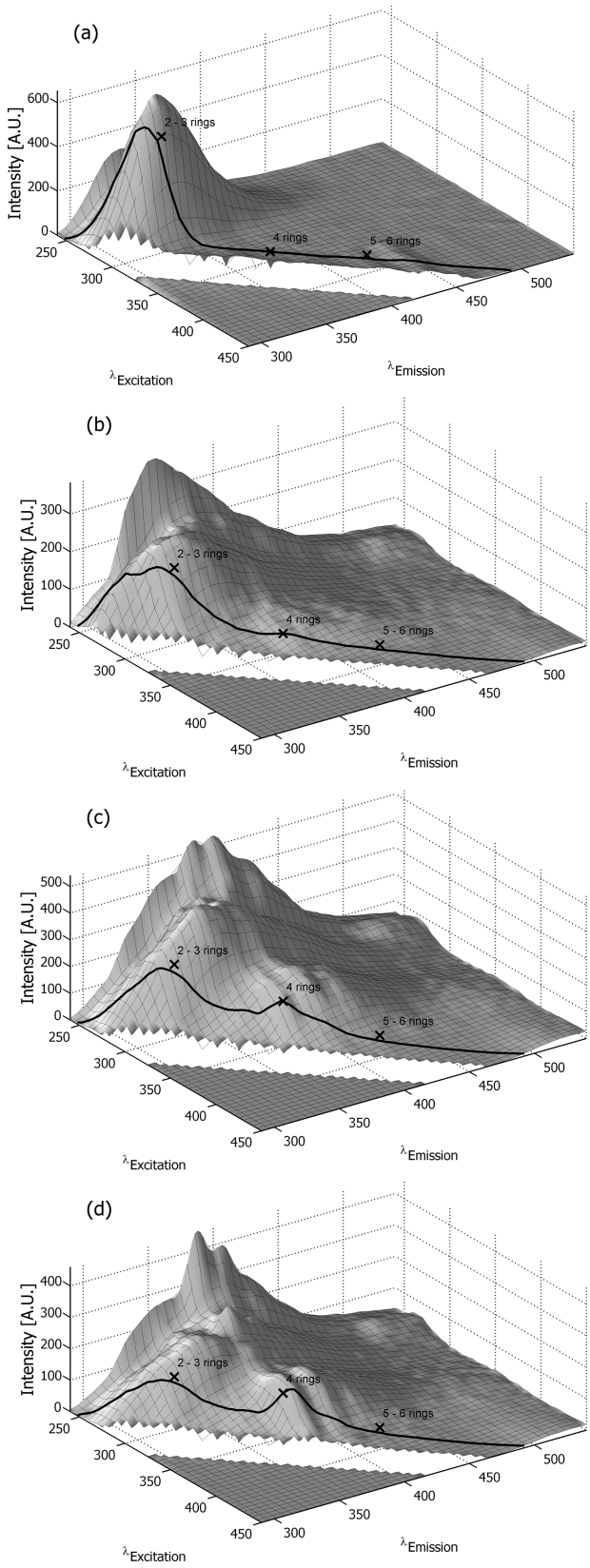


**FIGURE 1. Fluorescence excitation−emission spectra of four fish bile samples: (a) Narsasuaq, (b) Nanortalik, (c) Copenhagen harbor, and (d) Copenhagen harbor spiked with 1-hydroxypyrene. The wavelength pairs 290/335 nm (2- or 3-ring PACs), 341/383 nm (4-ring PACs), 380/430 nm (5-ring PACs); and the SFS with $\Delta\lambda = 42$ nm are marked in the plots.**

bile samples based on the hypothesis that, through the use of the entire fluorescence landscapes and of the PARAFAC model, one can obtain a more comprehensive characterization of the PAC exposure less prone to interference from biological bile constituents.

## 2. Materials and Methods

**2.1. Sample Collection.** Fish samples were collected in August 2006 during the Danish research cruise Galathea3 (www.Galathea3.dk/UK). European eels *(Anguilla anguilla)* were caught from inside Copenhagen harbor, Denmark, whereas shorthorn sculpins (*Myoxocephalus scorpius*) were caught from 15 coastal sampling sites in Greenland (see Supporting Information (SI) Figure S1 and Table S1). The sampling sites in Greenland were located inside local harbors close to pollution hot spots (boating activities with minor gasoline and light fuel oil spills), outside harbors close to local boating activities and at reference sites far away from any human activities (e.g., Narsasuaq-3, cf. SI Table S1). The bile fluid was collected by piercing the exposed gall bladder wall with a needle fitted to a 1 mL disposable syringe. The samples were stored in 1 mL eppendorf tubes at −20 °C in freezers on-board the research vessel.

**2.2. Experimental Section.** The bile samples were vial mixed for 1 min and centrifuged for 1 min at 10 000 rpm. Then standard dilutions between 400 and 1600 times were prepared in 1:1 mixture of methanol (Merck, Darmstadt, Germany) and deionized water (v/v) to avoid saturation of the detector and to reduce the effects of quenching, inner filter and energy transfer processes on the measured fluorescence landscapes. The detector voltage was set to 950 V to obtain a sufficiently high fluorescence signal at these dilutions.

The EEMs were measured on a Varian Eclipse fluorescence spectrophotometer in scan mode. A collection of emission scans from 290 to 560 nm with 2 nm increments was obtained at varying excitation wavelengths ranging from 240 to 450 nm with 5 nm increment. The bandwidths were 5 nm for excitation and 2.5 nm for emission, and the scan rate 1200 nm $min^{-1}$, the latter leading to a scan time of less than 10 min per sample. Each scan was comprised of 136 emission and 43 excitation wavelengths.

The SFS at constant wavelength difference ($\Delta\lambda = 42$) for the excitation range 248−450 nm were extracted from the EEMs using linear interpolation (each spectrum comprised of 102 points, with 2 nm increments). A 6-point external calibration curve of 1-hydroxypyrene in the range of 0.02 to 1.9 ng $mL^{-1}$ was used for the semiquantitative analysis of 1-hydroxypyrene equivalents (1OHPy).

**2.3. Data.** A total of 42 fish bile samples and two blanks were analyzed by fluorescence spectroscopy. EEM regions with low fluorescence at high emission wavelength ($\lambda > 540$ nm) were removed, resulting in a reduced three-way array of the size 42 fish bile samples × 126 emission wavelengths × 43 excitation wavelengths, with emission ranging from 290 to 540 nm and excitation from 240 to 450 nm.

In addition, EEMs of three solutions of aromatic amino acids (phenylalanine, tryptophan and tyrosine) in 0.1 M HCl (aq), and of four hydroxy-PACs (2,3-dihydroxyphenanthrene, 9-hydroxyphenanthrene, 1-hydroxypyrene, and 3-hydroxybenzo(a)pyrene) in methanol/water (1:1, v/v) were acquired at appropriate dilutions (2−110 ng $mL^{-1}$ ng/mL for amino acids and 30−340 ng $mL^{-1}$ for hydroxy-PACs).

**2.4. Data Preprocessing.** The Rayleigh and Raman scattering signals increase the number of PARAFAC components required to model EEM data and are therefore dealt with prior to the actual data analysis (*18*). The following procedure was employed here to reduce the influence of scattering: the EEM of the solvent blank measured closest in time was subtracted from each fluorescence landscape, the values in a 38 nm band (28 nm below and 10 nm above the excitation wavelength) in each emission scan were set to missing and

the other values below the excitation wavelength in each scan were set to zero (*16*). The EEMs were then corrected for instrument biases using an excitation correction spectrum derived from a concentrated solution of rhodamine B and an emission correction spectrum derived from a ground quartz diffuser (*19*).

**2.5. Parallel Factor Analysis.** If fluorescence EEMs are stacked to form a three-way array **X** of dimensions $I \times J \times K$, where $I$ is number of samples, $J$ number of emission wavelengths, and $K$ number of excitation wavelengths, they can be modeled using PARAFAC. In this model, each element $x_{ijk}$ of **X** is approximated by the sum of $F$ products of three parameters $a_{if}$, $b_{jf}$, and $c_{kf}$ (eq 1), which are grouped in three matrices **A** (the score matrix), **B**, and **C** (loading matrices, cf. eq 1).

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \text{ for } i = 1...I,\ j = 1...J,\ k = 1...K \quad (1)$$

where $x_{ijk}$ is the intensity of fluorescence for the $i$th sample, at the $j$th emission wavelength and at the $k$th excitation wavelength. The number of columns in the loading matrices ($F$) is the number of PARAFAC factors and $e_{ijk}$ the residuals, i.e., the unaccounted variability.

Under certain conditions, the PARAFAC model corresponds to the underlying physical model of fluorescence data (*20*); when that is the case, the $f$th column (score vector) of **A** is directly proportional to the concentration of the $f$th fluorescent species in the measured samples and $\mathbf{b}_f$ and $\mathbf{c}_f$ (i.e., the $f$th columns of **B** and **C**, respectively) are estimates of its excitation and emission spectra. However, fish bile contains numerous fluorescent species and PARAFAC is more likely to yield factors that refer to homogeneous classes of compounds (e.g., the homologues of 2 or 3 ring PACs).

The PARAFAC model was fitted in MATLAB version 7.4.0 using the N-way toolbox and an alternating least-squares algorithm that uses expectation maximization to deal with the missing values due to both scattering and cross validation (*18*).

## 3. Results and Discussion

**3.1. Fluorescence Data.** The preprocessed EEMs of three diluted bile samples from stations with different expected contamination levels (viz. Narsasuaq, reference station; Nanortalik, small harbor in Southern Greenland, and Copenhagen harbor) are shown in Figure 1a−c together with a fish bile sample from Copenhagen harbor spiked with additional 50 ng $mL^{-1}$ of 1-hydroxypyrene (Figure 1d). Note that contamination levels in fish bile samples cannot be directly related to contamination levels at the sites when different species are compared (viz. European eels in Copenhagen harbor vs shorthorn sculpins in Greenland), since these can have different feeding habits and the relationship between exposure and tissue concentrations depend on their rate of uptake, biotransformation, and excretion.

The wavelength pairs recommended by Ariese et al. (2005) for semiquantitative assessment of PACs in fish by FF and SFS at $\Delta\lambda = 42$ nm (*8*) are also shown in the figure.

Fluorescence signals above the baseline are observed at the excitation and emission pairs used in FF and in the SFS for all the samples. The emission intensities in the spectral region of 4-, 5-, and 6-ring PACs are more pronounced for the samples with the highest expected contamination level and close to 0 for the reference site. However, it is also apparent that the fluorescence landscapes contain spectral information that is not immediately encompassed by either the selected FF wavelengths or the SFS. In particular, a large peak is also present in the sample from the reference site (Figure 1a), which coincides with the selected wavelengths for quantification of 2 and 3 ring PAC metabolites. Since univariate methods cannot account for the presence of any interfering compound, the estimates for the concentrations of these PACs would be overestimated. Employing multivariate techniques on synchronous spectra could still yield accurate estimates as long as the interfering compounds are present in the calibration set, although no information on their nature could be gained. Another interesting feature is that the fluorescence peak maxima of conjugated 4-ring PAC metabolites in the SFS of the eel bile sample from Copenhagen harbor (341/383 nm) is blue-shifted compared to the fluorescence peak maxima obtained after standard addition of 1-hydroxypyrene (Figure 1c and d). The conjugation of a phenolic OH group (e.g., by coupling to glucuronic acid) leads to only a minor perturbation of the electronic system of the chromophore, but while the excitation and emission spectra of free and conjugated hydroxy-PACs are similar in shape, the conjugation has been found to lead to a slight spectral blue-shift (i.e., change to lower wavelengths) (*12, 21*).

The presence of conjugated PAC metabolites in fish bile samples from the Greenland environment was confirmed by an UPLC-qTOF (ultra performance liquid chromatograph interfaced to a quadrupole time-of-flight tandem mass spectrometer) screening of selected bile samples from this study. In particular, two groups of conjugated PAC metabolites: (glucuronide conjugated 3-ring PAC metabolites, $m/z$ 369.097; and sulfate conjugated 4-ring hydroxy-PAC, $m/z$ 313.017) were tentatively identified in the highly PAC exposed sample from Nuuk (see SI Figure S2), whereas no such compound could be detected in the reference sample from Narsasuaq (see the Supporting Information for experimental details).

Concentration equivalents of 1-hydroxypyrene were estimated by SFS for the 42 fish bile samples included in this study. The concentrations ranged between <1 and <30 ng 1OHPy $mL^{-1}$ depending on dilution factors and $2 \times 10^2$ ng 1OHPy $mL^{-1}$ bile for shorthorn sculpins from the Greenland environment, and between $9 \times 10^1$ and $6 \times 10^2$ ng 1OHPy $mL^{-1}$ for European eels from the Copenhagen harbor. Note that these values are very approximate estimates of the pollution levels because of the low specificity of the SFS method. Even though a correction factor is often used to account for the higher fluorescence intensities of PAC conjugates compared to 1-hydroxypyrene, its value is highly uncertain and depends on the nature of the PAC conjugates in the bile. For example, whereas Ariese et al. (1993) using SFS found a $2.2 \pm 0.1$ correction factor ($\Delta\lambda = 37$) based on pyrene-1-glucuronide (19), Stroomberg et al. (1999) found that the ratios between the emission maxima of four conjugates isolated from the gut of the isopod *Porcellio Scaber* (two were identified as pyrene-1-sulfate and pyrene-1-glucoside) and of 1-hydroxypyrene were between 1.6 and 1.8 (*12, 21*). Therefore, considering also that the aim of this study is not to compare pollution levels with other works, no correction was employed here.

**3.2. PARAFAC Modeling.** PARAFAC models with 1−6 factors were calculated using the complete preprocessed data set (42 fish bile samples × 126 emission wavelengths × 43 excitation wavelengths). A convergence criterion of $10^{-8}$ relative decrease in the least-squares loss function was employed and a maximum of 10 000 iterations was allowed. The models were estimated first without constraints and then enforcing nonnegativity on all three modes; these constraints are in accordance with chemical a priori knowledge (i.e., negative concentrations and fluorescence intensities are physically meaningless) and turned out to be necessary to yield stable and meaningful solutions (*18*). The nonnegativity constraints were initially validated by observing that the explained variance, the core consistency and the other diagnostics (*22*) were largely unaffected when nonnegativity

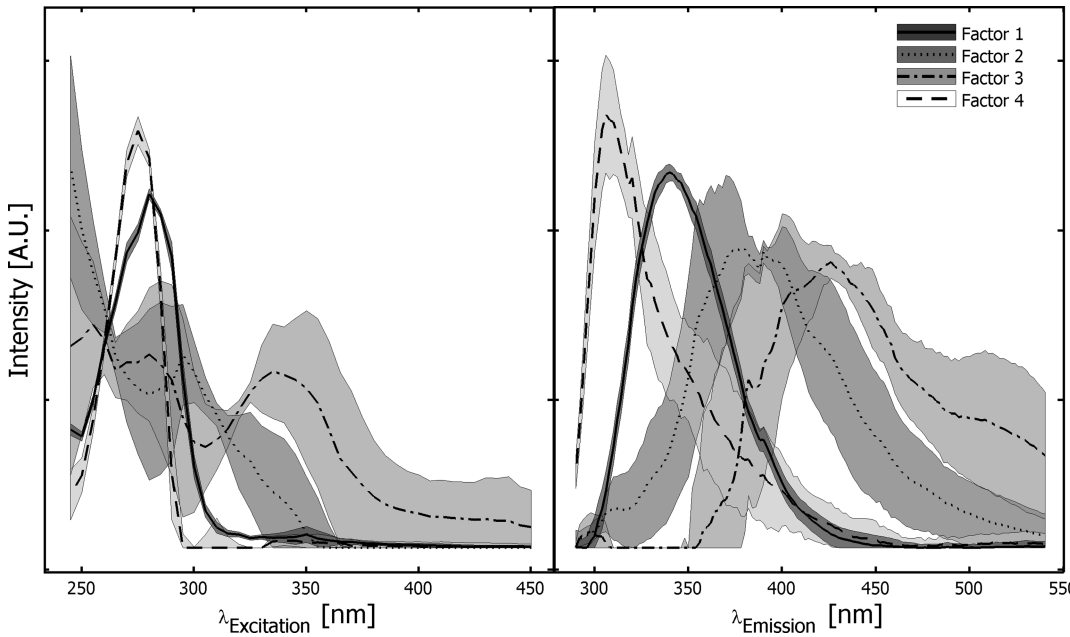


**FIGURE 2. Excitation and emission loadings for the four-factor model with confidence intervals based on jack-knifing ($t_{0.975,42}$ × standard error).**

constraints were imposed (not shown). The fluorescence spectra measured at $\lambda_{ex} = 240$ nm were particularly noisy and distorted and the model residuals were generally high and systematic for this wavelength (not shown). This could be ascribed to the conditions of the xenon lamp (*23*) and, to yield more stable results, the spectra measured at this excitation wavelength were removed.

The core consistency diagnostics decreases significantly from a four-factor model (>80%) to a five-factor one (<35%), but the presence of a fifth component cannot be excluded (see SI Figure S3a). The explained variance increases abruptly when two factors are used instead of one, but the improvements are minor when additional factors are used (see SI Figure S3b). Cross validation yields similar conclusions and did not lead to further insight into the correct number of factors. More information in this sense was obtained from split-half analysis and jack-knifing (leaving out one sample at a time): both show that there are two very stable factors (1 and 4, cf. Figure 2 and SI Figure S3c) with simple excitation and emission loadings and 2, whose profiles change considerably depending on the composition of the data set. However, the values for both the core consistency and the explained variation do not significantly differ in either case, which strongly supports the conclusion that at least four factors can be reliably extracted from the data set.

Residual and leverage analyses were used for outlier detection in PARAFAC modeling together with jack-knifing (*22*), but no clear outliers could be identified. Moreover, albeit some systematic behavior could be observed in the residuals of the four-factor model (but also for the five-factor one), all the EEMs were well described by the model and were thus maintained. Nonetheless, the confidence intervals obtained from the jack-knifing (calculated as jack-knifing standard error × 2.045, i.e., the quantile of a Student distribution with 42 degrees of freedom, and $\alpha = 0.95$) were particularly wide for factors 2 and 3, which appear to be related to the level and type of contamination. The relative scarcity of heavily contaminated samples and the small size of the data set are the most likely reason for this instability. In particular, the dominating PAC metabolites might be quite different in contaminated samples, leading to appreciable shifts in spectral features. This is also confirmed by the great stability of the other two components, which can be associated to biological factors (see below) and thus are expected to be less affected by the environment.

**3.3. Chemical Interpretation of PARAFAC Factors.** The chemical interpretation of the four factors in the PARAFAC model was based on comparisons of the factor loadings with the fluorescence excitation−emission landscapes of aromatic amino acids and of PAC metabolites (Figure 3), the spectral shape and maxima of 2−6 ring PACs (*16, 24*), and the scores of bile samples with different expected contamination levels (Figure 4).

The amino acid residues of tryptophan, tyrosine, and phenylalanine from bile proteins are the most likely "biological" fluorophores in fish bile (*25*). As can be seen from Figure 3a, the excitation and emission loadings for the two biological factors are compatible with the excitation and emission spectra of tryptophan (factor 1) and tyrosine (factor 4), whereas no factors can be associated to phenylalanine. This is particularly evident for tyrosine, as its maximum absorption and emission correspond to the maxima of the corresponding factor loadings and the correlation coefficient ($r$) between the excitation and emission spectra of tyrosine and the factor loadings are 0.995 and 0.945, respectively. The small red shift for the excitation profile and the visible tailing of the emission loading of factor 4 with respect to pure tyrosine can be explained by matrix effects favoring the formation of tyrosinate from the excited state, which has maximum emission around 345 nm (*25*). The differences between factor 1 loadings and the tryptophan spectra are larger: 5 nm red-shift in the near UV absorption and 12 nm blue-shift in emission and the correlation between the excitation and emission spectra of tryptophan and the factor loadings are 0.985 for excitation and 0.930 for emission. However, the tryptophan spectra are affected by the surrounding environment and shifts can originate from a combination of the different conditions in which the indole residue is found within the solution (*25*). In general, the shifts appear to be compatible with those observed for tryptophan in protein studies (*25*), although additional analyses are required to further validate the identification.

The concentration of bile components steadily increases between meals, which implies that expressing the concentration of PAC metabolites per volume of bile can

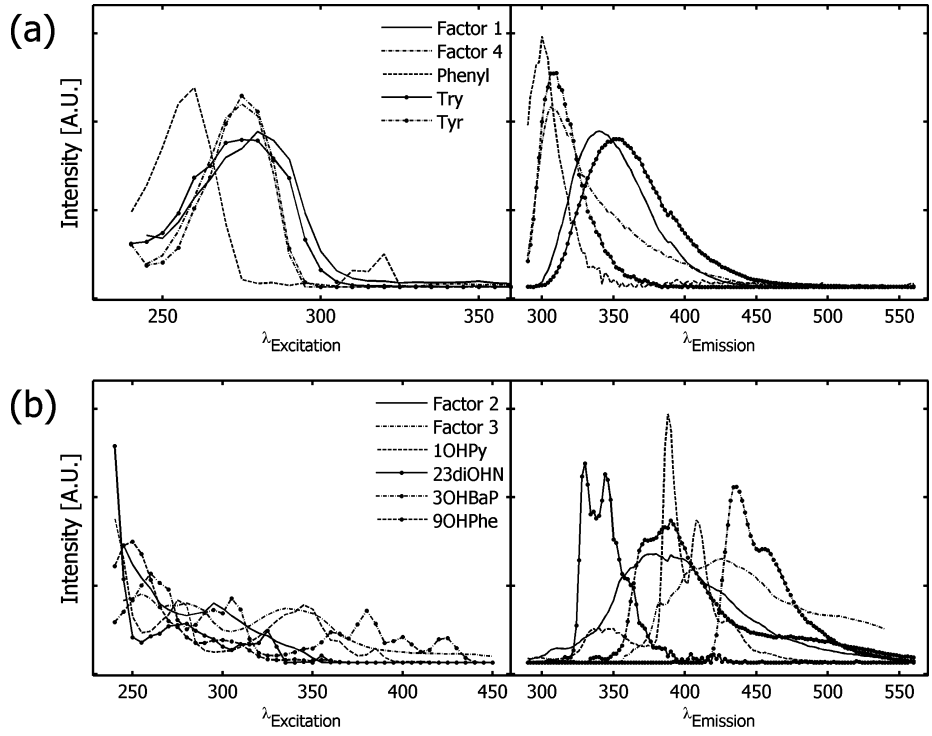


**FIGURE 3. Comparison of excitation and emission loadings for (a) Factor 1 and 4 with fluorescence spectra of phenylalanine (Phenyl), tryptophan (Try), and tyrosine (Tyr); and (b) Factor 2 and 3 with fluorescence spectra of 2,3-dihydroxynaphthalene (23diOHN), 9-hydroxyphenanthrene (9OHPhe), 1-hydroxypyrene (1OHPy), and 3-hydroxybenzo(a)pyrene (3OHBaP).**

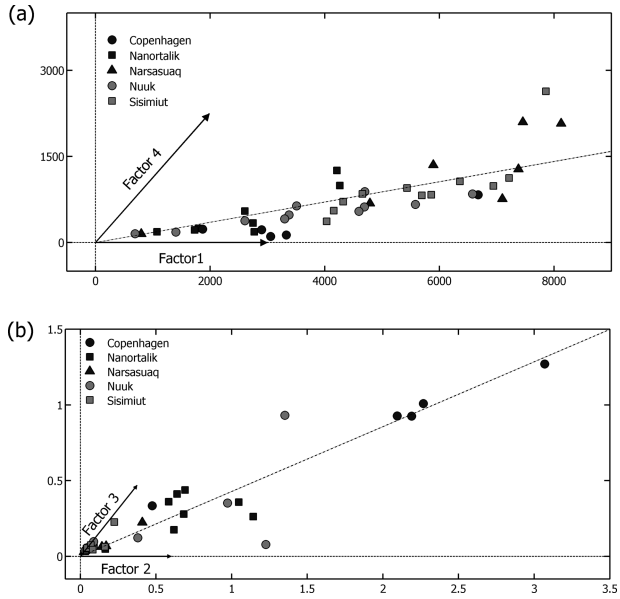


**FIGURE 4. PARAFAC score plots. (a) Factor 1 vs Factor 4. Symbols for bile samples are explained in the legend. The plot axes are an orthogonal basis for the space spanned by the two factors (*29*). The dash-dot line indicates the ODR line. (b) Factor 2 vs Factor 3 after normalization to the sum of the biological factor scores.**

change over time, or between individuals, depending on their feeding status and irrespective of their actual exposure to PACs (*26*). Internal standardization to the bile pigment biliverdin or to the protein content are the most frequent choices as correction factors for FF and SFS methods (*8, 26*). However, it has recently been shown that normalization to biliverdin is not suitable to correct for feeding status since the metabolic pathways leading to biliverdin are affected by the exposure to other contaminants and due to the limited knowledge on the effects of starvation on the biliverdin excretion (*26*).

Proteins contain fluorescent residues (mostly tryptophan and tyrosine) (*25*) and could provide a better alternative for internal standardization (*26, 27*). Assuming that the chemical interpretation of the biological factors 1 and 4 is correct, their scores are expected to correlate to the protein content; this is supported by the positive correlation ($r = 0.71$, Figure 4a) between the two score vectors. The orthogonal projection of samples on the orthogonal distance regression (ODR) (*28*) line between the two score vectors (dash-dot line in the figure) would then be proportional to the protein content and may provide a good normalization factor for the contamination related scores to account for the feeding status. The mean of the two sample scores for each fish bile sample was used for normalization in this study, as it was found to be proportional to the values of such projections and was a more operational and straightforward approach.

Factors 2 and 3 are clearly associated to PAC exposure, as can be seen for example from the high scores for the samples from contaminated sites (Figure 4b), but their chemical interpretation is not as straightforward as for the biological factors. The spectra of four hydroxy-PAC metabolites with 2−5 aromatic rings are displayed in Figure 3b together with the excitation and emission loadings of the two "chemical" factors. Although they are of minor importance in fish bile, phenolic PACs were used in the comparison since only few conjugated metabolites are commercially available.

Even though some overlap exists between the loadings and the spectra of phenolic PACs, no clear correspondence with a specific class of PAC compounds can be observed. In particular, neither chemical factor has extensive overlap with the fluorescence spectra of 2,3-dihydroxynaphthalene (Figure 3b), whereas in emission, the overlaps between the loading for factor 2 and the spectra of 9-hydroxyphenanthrene and 1-hydroxypyrene and between the

factor 3 loading and the spectrum of 3-hydroxybenzo-(a)pyrene suggest that the PARAFAC model can separate the contributions of 3 and 4 ring PAC metabolites from PAC metabolites with 5- and more aromatic rings. In fact, the correlation coefficient ($r$) between the excitation and emission spectra of 9-hydroxyphenanthrene and the factor 2 loadings are 0.908 and 0.881, respectively. The broader peaks without vibrational fine-structure could be explained by the large number of similar metabolites described by each component, whereas conjugation could explain the spectral blue shifts (*16, 21*). However, the relatively poor overlap between the excitation loadings and the spectra of hydroxy-PACs and the relative instability of these factors suggests that their interpretation is not so clear-cut and that especially the 4-ring PAC metabolites are somewhat split between the two factors. Furthermore, the residuals for the most polluted samples (e.g., Nuuk-3 and Copenhagen) contain some systematic variation in the spectral regions where the 2-ring and 6-ring PAC metabolites are expected to emit (i.e., $\lambda_{ex} \leq 300$ and $\lambda_{em} \leq 350$ for the former, and $\lambda_{ex} \geq 300$ and $\lambda_{em} \geq 350$ for the latter, not shown), which indicates that these classes of compounds are not well modeled. This supports the existence of at least a fifth component, which cannot be separated due to the relatively small size of the data set.

**3.3. Interpretation of the Contamination Levels and Sources.** Notwithstanding the lack of a direct interpretation of the chemical factors, the PARAFAC model explains 99% of the variation in the data set and thus can be seen as a data compression tool. The position of the samples in the score plot of factor 2 vs factor 3 (Figure 4b) therefore reflects the PAC exposure and possibly give some indication on the type of exposure. Compared to the biological factors, the chemical ones show a higher correlation ($r = 0.84$) as can also be seen from the orthogonalized score plot (Figure 4b). The ODR line between the two score vectors seems associated to the level of PAC metabolites in the bile, which is particularly high for European eel samples from Copenhagen harbor and for some samples of shorthorn sculpin from Nuuk and Nanortalik. A few samples, however, deviate from this simple behavior: e.g., the Nuuk-4 sample is particularly high for factor 2 and low for factor 3 (this is further enhanced without the orthogonalization of the coordinates, SI Figure S4). This second direction of variance may be related to the type of exposure, since petrogenic sources are rich in parent and alkylated PACs with fewer aromatic rings (two or three), whereas pyrogenic ones (incomplete combustion of organic matter and fossil fuels) are mostly comprised of parent PACs with four to six aromatic rings. Although, based on few samples, these conclusions are supported by visual observations at the sampling sites (e.g., Nuuk-4 is a small heavily oil polluted harbor) and from the relative concentrations of parent and alkylated two to six ring PACs in blue mussels collected from the same sites (unpublished data).

**3.4. Comparison of the EEM and the SFS Approach.** A further validation of the EEM approach for screening of PAC exposure in fish was made by predicting the 1OHPy equivalents determined by SFS from the chemical scores. The ODR model is displayed in Figure 5 and shows that for relatively high contamination levels, the two methods are in good accordance. However, several low exposure samples from the Greenland environment (see insert) deviate between the two methods. While the SFS measurements for these samples mainly from Nanortalik are not distinguishable from the baseline noise, the chemical score values determined by the EEM approach are clearly above those observed for samples from reference sites. These observations support the previous conclusions that the PARAFAC modeling of fluorescence excitation−emission spectra capture more information and could be more sensitive than the standard FF and SFS approaches.

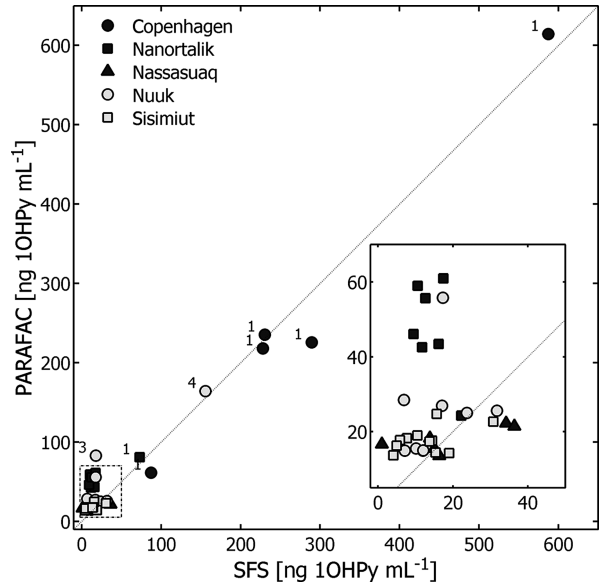


**FIGURE 5. Regression model of 1OHPy equivalents and normalized scores for Factors 2 and 3.**

## Acknowledgments

The present work was carried out as part of the Galathea3 expedition under the auspices of the Danish Expedition Foundation. The study was financially supported by the Danish Expedition foundation (providing accommodation and laboratory facilities on-board the research vessel), the Villum Kann Rasmussen foundation, the Torkil Holms Foundation and the Lundbeck foundation. G.T. gratefully acknowledges the financial support of the RECETO consortium. We thank the naval personal onboard the research vessel and Richard Sylvestersen, Troels Møller, Dorthe Petersen (Asiaq), and Jan Rasmussen (City of Copenhagen) for valuable help with sampling and sample preparation, and Martin Larsen for preparation of Figure S1 in the Supporting Information. The Galathea3 contribution no. is P32.

## Supporting Information Available

Additional information on the sampling sites, the diagnostic for determining the number of factors and the stability of the four-factor PARAFAC model, the nonorthogonalized score plot for factors 2 and 3, method details for the UPLC-qTOF analysis of selected bile samples, and SICs from a fish bile sample from Nuuk. This material is available free of charge via the Internet at http://pubs.acs.org.

## Literature Cited

(1) Lima, A. L. C.; Farrington, J. W.; Reddy, C. M. Combustion-Derived Polycyclic Aromatic Hydrocabons in the Environmnet-a Review. *Environ. Forensics.* **2005**, *6*, 109–131.

(2) Boehm, P. D. In *Environmental Forensics: Contaminant Specific Guide*, 1st ed.; Morrison, R. D., Murphy, B. L. , Eds.; Academic Press: New York 2008; Chapter 15.

(3) Stegeman, J. J.; Lech, J. J. Cytochrome-P-450 monooxygenase systems in aquatic species - carcinogen metabolism and biomarkers for carcinogen and pollutant exposure. *Environ. Health Perspect.* **1991**, *90*, 101–109.

(4) Foureman, G. L. In *Metabolism of Polycyclic Aromatic Hydrocarbons in the Aquatic Environment*, Chapter 5; Varanasi, U. Ed.; CRC Press: Boca Raton, FL, 1989.

(5) Baumard, P.; Budzinski, H.; Garrigues, P.; Sorbe, J. C.; Burgeot, T.; Bellocq, J. Concentrations of PAHs (polycyclic aromatic hydrocarbons) in various marine organisms in relation to those

in sediments and to trophic level. *Mar. Pollut. Bull.* **1998**, *36* (12), 951–960.
(6) Ariese, F.; Burgers, I.; Oudhoff, K.; Rutten, T.; Stroomberg, G. J.; Vethaak, D. *Comparison of Analytical Approaches for PAH Metabolites in Fish Bile Samples for Marine and Estuarine Monitoring*;, Universiteit, Institute of Environmental Studies: Amsterdam, 1997.
(7) Krahn, M. M.; Ylitalo, G. M.; Buzitis, J.; Chan, S. L.; Varanasi, U. Rapid high-performance liquid-chromatographic methods that screen for aromatic-compounds in environmental-samples. *J. Chromatogr., A* **1993**, *642* (1−2), 15–32.
(8) Ariese, F.; Beyer, J.; Jonsson, G.; Visa, C. P.; Krahn, M. M. Review of analytical methods for determining metabolites of polycyclic aromatic comounds (PACs) in fish bile. *ICES Tech. Mar. Environ. Sci.* **2005**, *39*, 1–49.
(9) Johnson-Restrepo, B.; Olivero-Verbel, J.; Lu, S. J.; Guette-Fernandez, J.; Baldiris-Avila, R.; O'Byrne-Hoyos, I.; Aldous, K. M.; Addink, R.; Kannan, K. Polycyclic aromatic hydrocarbons and their hydroxylated metabolites in fish bile and sediments from coastal waters of Colombia. *Environ. Pollut.* **2008**, *151* (3), 452–459.
(10) Krahn, M. M.; Burrows, D. G.; Ylitalo, G. M.; Brown, D. W.; Wigren, C. A.; Collier, T. K.; Chan, S. L.; Varanasi, U. Mass-spectrometric analysis for aromatic-compounds in bile of fish sampled after the Exxon-Valdez oil-spill. *Environ. Sci. Technol.* **1992**, *26* (1), 116–126.
(11) Ruddock, P. J.; Bird, D. J.; McEvoy, J.; Peters, L. D. Bile metabolites of polycyclic aromatic hydrocarbons (PAHs) in European eels Anguilla anguilla from United Kingdom estuaries. *Sci. Total Environ.* **2003**, *301* (1−3), 105–117.
(12) Ariese, F.; Kok, S. J.; Verkaik, M.; Gooijer, C.; Velthorst, N. H.; Hofstraat, J. W. Synchronous fluorescence spectrometry of fish bile—A rapid screening method for the biomonitoring of pah exposure. *Aquat. Toxicol.* **1993**, *26* (3−4), 273–286.
(13) Lin, E. L. C.; Cormier, S. M.; Torsella, J. A. Fish biliary polycyclic aromatic hydrocarbon metabolites estimated by fixed-wavelength fluorescence: Comparison with HPLC-fluorescent detection. *Ecotox. Environ. Safe.* **1996**, *35* (1), 16–23.
(14) Vuorinen, P. J.; Keinanen, M.; Vuontisjarvi, H.; Barsiene, J.; Broeg, K.; Forlin, L.; Gercken, J.; Kopecka, J.; Kohler, A.; Parkkonen, J.; Pempkowiak, J.; Schiedek, D. Use of biliary PAH metabolites as a biomarker of pollution in fish from the Baltic Sea. *Mar. Pollut. Bull.* **2006**, *53* (8−9), 479–487.
(15) Beltran, J. L.; Guiteras, J.; Ferrer, R. Parallel factor analysis of partially resolved chromatographic data—Determination of polycyclic aromatic hydrocarbons in water samples. *J. Chromatogr., A* **1998**, *802* (2), 263–275.
(16) Christensen, J. H.; Hansen, A. B.; Mortensen, J.; Andersen, O. Characterization and matching of oil samples using fluorescence spectroscopy and parallel factor analysis. *Anal. Chem.* **2005**, *77* (7), 2210–2217.
(17) Patra, D. Applications and new developments in fluorescence spectroscopic techniques for the analysis of polycyclic aromatic hydrocarbons. *Appl. Spectrosc. Rev.* **2003**, *38* (2), 155–185.
(18) Andersen, C. M.; Bro, R. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *J. Chemom.* **2003**, *17* (4), 200–215.
(19) Stedmon, C. A.; Markager, S.; Bro, R. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* **2003**, *82* (3−4), 239–254.
(20) Smilde, A. K.; Bro, R.; Geladi, P. *Multi-Way Analysis Applications in the Chemical Sciences*, 1st ed.; John Wiley & Sons Ltd: Chichester, England, 2004.
(21) Stroomberg, G. J.; de Knecht, J. A.; Ariese, F.; Van Gestel, C. A. M.; Velthorst, N. H. Pyrene metabolites in the hepatopancreas and gut of the isopod Porcellio scaber, a new biomarker for polycyclic aromatic hydrocarbon exposure in terrestrial ecosystems. *Environ. Toxicol. Chem.* **1999**, *18* (10), 2217–2224.
(22) Bro, R. Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications, Disputats, Universiteit van Amsterdam, 1998.
(23) Riu, J.; Bro, R. Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. *Chemom. Intell. Lab.Syst.* **2003**, *65* (1), 35–49.
(24) Beltran, J. L.; Ferrer, R.; Guiteras, J. Multivariate calibration of polycyclic aromatic hydrocarbon mixtures from excitation-emission fluorescence spectra. *Anal. Chim. Acta* **1998**, *373* (2−3), 311–319.
(25) Lakovicz, J. R. *Principles of Fluorescence Spectroscopy*, 2nd ed.; Kluwer Academic/Plenum Publishers: New York, 1999.
(26) van den Hurk, P. Bile fluorescence, heme oxygenase induction, and increased biliverdin excretion by mixtures of environmental toxicants. *Aquat. Toxicol.* **2006**, *77* (2), 202–209.
(27) Aitken, A. and Learmonth, M. P. In *The Protein Protocols Handbook*, 2nd ed.; Humana Press: New York, 2002.
(28) Massart, D. L.; Vandeginste, B. G.; Buydens, L. M. C.; DeJong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics, 2375*, Elsevier: Amsterdam1997.
(29) Kiers, H. A. L. Some procedures for displaying results from three-way methods. *J. Chemom.* **2000**, *14* (3), 151–170.

ES900187X