Electronic Supplementary Information to paper

**Reaction-based Optical Fingerprinting Strategy for the Recognition of Fat-Soluble Samples: Discrimination of Motor Oils**

by Arseny A. Pypin, Anna V. Shik, Irina A. Stepanova, Irina A. Doroshenko, Tatyana A. Podrugina, Mikhail K. Beklemishev\*

**Contents of the ESI**

* **Synthesis of dye 2** (protocols and spectral data)
* **Table S1.** An example of a data table containing intensities of photograpihic images of samples.
* **Table S2**. An example of confusion matrix created by XLSTAT LDA software for evaluating the accuracy of discrimination by validation procedure.
* **Fig. S1.** Absorbance spectra at different reaction times for the indicator reactions named above the images.
* **Fig. S2.** Visible images of the reacting system *dye* ***2*** *+ HNO3* obtained by a smartphone camera.
* **Fig. S3.** Kinetic curves for the reaction between dye **1** and HNO3 plotted for three color channels (R, G, B).
* **Fig. S4.** Kinetic curves of indicator reactions (a) Dye **1** + t-BuOOH, (b) Dye **2** + t-BuOOH + Cu2+, in the presence of oil samples.
* **Accuracy of assignment of the sample as a whole (6 observations)**
* **References**

**Synthesis of dye 2** (2-((E)-2-((E)-2-(4-formylphenoxy)-3-((E)-2-(3-hexyl-1,1-dimethyl-1,3-dihydro-2H-benzo[e]indol-2-ylidene)ethylidene)cyclohex-1-en-1-yl)vinyl)-3-hexyl-1,1-dimethyl-1H-benzo[e]indol-3-ium iodide)



**2**

Compound **I** was obtained as described in [1]. Compounds **II** and **III** (dye **2**) were synthesized as shown in the Scheme and described below in more detail:





1H and 13C NMR spectra were recorded on Bruker Avance 400 (400 and 100 МHz, respectively) and Bruker Avance 600 (600 and 150 МHz). Residual signals of solvents were used as a reference (1Н: CDCl3, δ 7.26; CD2Cl2, δ 5.32. 13C: CDCl3, δ 77.1, CD2Cl2, δ 54.0).

IR spectra were recorded on a Nicolet IR 200 Fourier-transform spectrometer (Thermo Scientific) using an internal reflectance accessory with a ZnSe attenuated total reflection (ATR) element with an incidence angle of 45°. The resolution was 4 cm–1, the number of scans was 20.

High resolution mass spectra (HRMS) were recorded on an Agilent LC/MSD 1100 SL instrument with atmospheric pressure electrospray ionization (AP-ESI) in the positive ion detection mode (ion trap mass analyzer). Registration conditions: the nebulizer gas temperature (nitrogen) 300 °С at a rate of 12 L min–1, the source potential 5000 V, the capillary outlet potential 150 V, solvent acetonitrile.

Reaction progress and purity of chromatographically separated compounds were monitored by thin-layer chromatography on Silica gel 60 F254 plates (Merck).

Chromatographic separation was carried out on a flash chromatography systemBiotage Isolera Prime and on a column with MN Kieselgel 60 silica gel, 0.04—0.063 mm (230-400 mesh) ASTM and Interchim puriflash 60 si hp, 50 μm particle size.

**2-((E)-2-((E)-2-chloro-3-((E)-2-(3-hexyl-1,1-dimethyl-1,3-dihydro-2H-benzo[e]indol-2-ylidene)ethylidene)cyclohex-1-en-1-yl)vinyl)-3-hexyl-1,1-dimethyl-1H-benzo[e]indol-3-ium iodide (**compound **II)**

A mixture of 3-hexyl-1,1,2-trimethyl-1H-benzo[e]indol-3-ium iodide(0.50 g, 1.18 mmol), (E)-2-chloro-3-(hydroxymethylene)cyclohex-1-ene-1-carbaldehyde (0.10 g, 0.59 mmol) and (0.12 g, 1.42 mmol) of sodium acetate in 4 mL of ethanol was stirred at 50 °C for 5 h. After cooling to room temperature, an excess of diethyl ether was added, the precipitate was filtered off, washed on the filter with diethyl ether and dried. The dye was purified by column chromatography on silica gel in a mixture of eluents CH2Cl2: СH3OH (15:1, v/v) (Rf = 0.59).

The yield was 0.290 g (69 %). λabs = 824 nm (EtOH), λfl = 834 nm (EtOH), ɛ = 2.9·105 L mol–1 cm–1.

**1H NMR** (400 MHz, CDCl3, δ, ppm, J/Hz): 0.90 (t, *3JНН* = 7.09, 6H, 2CH3), 1.28 - 1.43 (m, 8H, 4CH2), 1.45 - 1.55 (m, 4H, 2CH2), 1.92 (quin, *3JНН* = 7.47, 4H, 2CH2), 2.04 (s, 14H, 2С(CH3)2, CH2), 2.79 (t, *3JНН* = 5.99, 4H, СН2SO3), 4.34 (t, *3JНН* = 7.40, 4H, 2NCH2), 6.28 (d, *3JНН* = 14.12, 2H, -CH=), 7.41 - 7.53 (m, 4H, Ar), 7.58 - 7.67 (m, 2H, Ar), 7.96 (d, *3JНН* = 8.99, 4H, Ar), 8.14 (d, *3JНН* = 8.56, 2H, -CH=), 8.45 (d, *3JНН* = 14.37, 2H, -CH=).

**13C NMR** (100 MHz, CDCl3 δ, ppm, J/Hz): 13.62 (CH3), 20.40, 20.73 (both CH2), 22.09 (CH3), 24.27, 26.27, 26.35, 27.30, 31.08 (all CH2), 44.84 (C(CH3)2), 50.75 (+NCH2), 100.57, 121.71, 124.78, 126.79, 127.42, 127.73, 130.42, 131.54, 139.29, 142.85 (all Ar), 110.49 (-CH=), 125.32 (-C(CH2)=), 129.79 (-C(CH2)=), 133.47 (C-Cl), 149.36 (-CH=), 173.22 (=C-N, C=N+).

**HRMS-ESI**: m/z [M+] calculated for С50H60N2Cl723.4440, found 723.4431.

**IR**, ν/cm–1: 1474.8 (N=C-CH=), 1544.7 (С=С-N).

**2-((E)-2-((E)-2-(4-formylphenoxy)-3-((E)-2-(3-hexyl-1,1-dimethyl-1,3-dihydro-2H-benzo[e]indol-2-ylidene)ethylidene)cyclohex-1-en-1-yl)vinyl)-3-hexyl-1,1-dimethyl-1H-benzo[e]indol-3-ium iodide (**compound **III =** dye **2)**

A mixture of freshly prepared CH3ONa 0.029 g (0.54 mmol, 3 eq.) and para-hydroxybenzaldehyde 0.020 g (0.18 mmol, 1 eq.) was dissolved in a minimum amount of methanol. The solution was stirred at room temperature for 15 min. Then the solvent was evaporated and a light yellow residue was formed. The resulting sodium 4-formylphenoate dissolved in 2 mL of DMFA was added to 0.150 g of **compound II** (0.18 mmol, 1 eq.) dissolved in 2 mL of DMFA in an argon atmosphere. The mixture was stirred for 4 h at room temperature, and DMFA was evaporated at reduced pressure. The obtained **compound III** was purified by column chromatography on silica gel in a mixture of eluents CH2Cl2:CH3OH (20:1, v/v), *R*f = 0.48. The isolated fraction was washed with diethyl ether.

The yield was 0.032 g (19 %). λabs = 815 nm (EtOH), λfl = 831 nm (EtOH), ε = 4.3·105 L mol–1 cm–1.

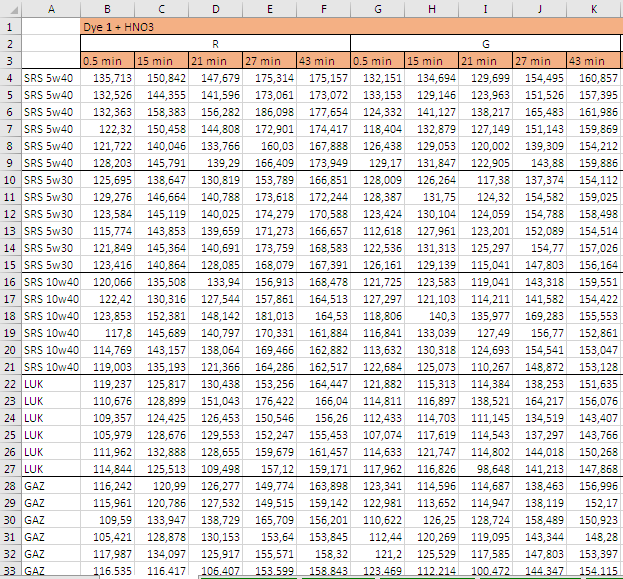
**1H NMR** (600 MHz, CD2Cl2, δ, ppm, J/Hz): 0.91 (t, *3JНН* = 7.06, 6H, 2CH3), 1.32 - 1.40 (m, 8H, 4CH2), 1.46 (quin, *3JНН* = 7.29, 4H, 2CH2), 1.64 (s, 12H, 4CH3), 1.85 (quin, *3JНН* = 7.50, 4H, 2CH2), 2.11 (quin, *3JНН* = 6.03, 2H, CH2), 2.78 (t, *3JНН* = 6.05, 4H, 2СН2SO3), 4.13 (t, *3JНН* = 7.57, 4H, 2NCH2), 6.09 (d, *3JНН* = 14.21, 2H, -CH=), 7.34 (d, *3JНН* = 8.53, 2H, Ar), 7.39 (d, *3JНН* = 8.71, 2H, Ar), 7.47 (t, *3JНН* = 7.34, 2H, Ar), 7.59 (t, *3JНН* = 7.57, 2H, Ar), 7.92 - 7.98 (m, 6H, Ar), 8.01 (t, *3JНН* = 8.48, 4H, Ar), 9.94 (s, 1H, CHO).

**13C NMR** (151 MHz, CD2Cl2 δ, ppm, J/Hz): 14.28 (2CH3), 21.69 (CH2), 23.02 (2CH2), 24.93 (CH2), 27.12 (2CH2), 27.81 (4CH3), 27.99 (2C(CH3)2), 31.95 (2CH2), 45.30 (2CH2), 51.42 (2+NCH2), 100.22 (2CH), 111.22 (Ar), 115.97, (С=С(O)-C), 121.83, 122.65, 125.68, 128.26, 128.55, 130.54, 131.16 (all Ar), 132.02 (C-CHO), 132.44, 133.12, 134.45, 140.07 (all Ar), 141.14 (2CH), 162.82, 164.53 (both С=С(O)-C), 174.13 (2С=N), 190.92 (CHO).

**HRMS-ESI**: m/z [M+] calculated for C57H65N2O2 809.5041, found 809.5034.

**IR,** ν/cm–1: 1468.05 (N=C-CH=), 1557.72 (С=С-N), 1693.68 (C=O).

**Table S1.** An example of a data table containing intensities of photograpihic images of samples listed in column A. Rows correspond to observations (wells of the plate) and columns correspond to different reaction times and different color channels (R – red, G – green) of the indicator reaction of Dye 1 with nitric acid. Only a part of the table is shown



**Table S2**. An example of confusion matrix created by XLSTAT LDA software for evaluating the accuracy of discrimination by validation procedure

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| True value | Predicted value | | | | | | Total num-ber | % accu­racy |
| EVE | GAZ | LUK | SRS 10w40 | SRS 5w30 | SRS 5w40 |
| EVE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | – |
| GAZ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | – |
| LUK | 0 | 0 | **2** | 0 | 0 | 0 | 2 | 100,0% |
| SRS 10w40 | 0 | 0 | 0 | **2** | **1** | 0 | 3 | 66,7% |
| SRS 5w30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | – |
| SRS 5w40 | 0 | 0 | 0 | 0 | 0 | **1** | 1 | 100,0% |
| **Total** | 0 | 0 | 2 | 2 | 1 | 1 | **6** | **83,3%** |

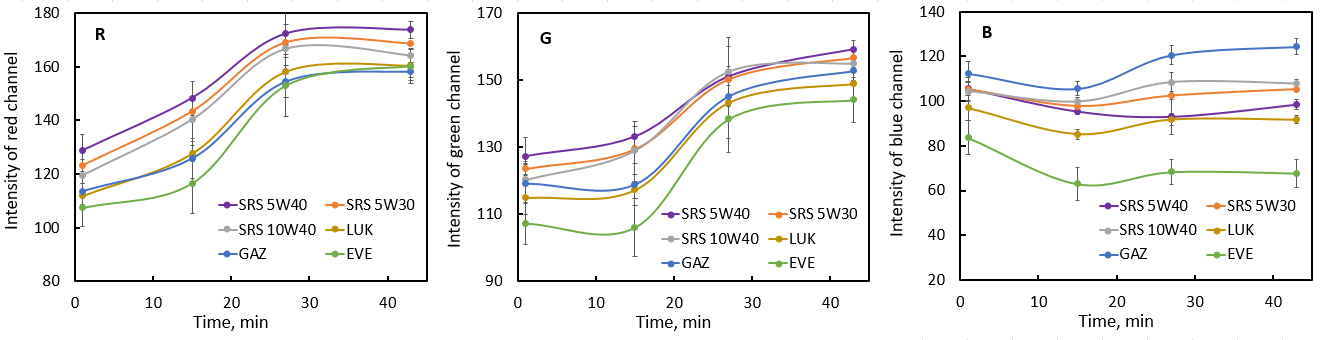
*Note*. Six randomly selected observations were used as validation set. Accuracy was calculated as the number of correctly assigned observations to the total number of observations in the validation set (in this example, it was 5 / 6 = 0.833).

|  |  |
| --- | --- |
| a) Dye **1** – aqua regia | b) Dye **1** – O2 |
|  |  |
| c) Dye **2** – t-BuOOH | d) Due **2** – t-BuOOH – Cu2+ |
|  |  |

**Fig. S1.** Absorbance spectra at different reaction times for the indicator reactions named above the images. The reactant amounts were as follows: a) 20 µL of 0.1 g/L dye **1**, 560 µL of EtOH, 4 µL of aqua regia; b) 20 µL of 0.1 g/L dye **1**, 440 µL of EtOH, 120 µL of concentrated HCl; c) 20 µL of 0.1 g/L dye **2**, 760 µL of EtOH, 40 µL of 1 M HCl, 80 µL of 0.5 M t-BuOOH; d) 20 µL of 0.1 g/L dye **2**, 760 µL of EtOH, 40 µL of 1 M HCl, 40 µL 10-4 М Cu2+, 60 µL of 0.5 M t-BuOOH.

|  |  |  |  |
| --- | --- | --- | --- |
| **5W40**  **5W30**  **10W40**  **LUK**  **GAZ**  **EVE** |  |  |  |
|  | **2 min** | **19 min** | **31 min** |

**Fig. S2.** Visible images of the reacting system *dye* ***2*** *+ HNO3* obtained by a smartphone camera. The replicate experiments for the same sample are in the rows. Time after the reaction start is shown under each image.

**Fig. S3.** Kinetic curves for the reaction between dye **1** and HNO3 plotted for three color channels (R, G, B). The results were averaged for the 6 parallel runs of each sample shown in the legend. The error bars represent a standard deviation.

|  |  |
| --- | --- |
| a) Dye **1** + t-BuOOH (NIR) | b) Dye **2** + t-BuOOH + Cu2+ (NIR) |
|  |  |

**Fig. S4.** Kinetic curves of indicator reactions: (a) Dye **1** + t-BuOOH, (b) Dye **2** + t-BuOOH + Cu2+, in the presence of oil samples shown in the legend. Each curve is a result of averaging of 6 replicate runs, and the error bars correspond to the standard deviations. Ordinate is the average intensity for the near-IR photographs.

**Accuracy of assignment of the sample as a whole** (6 parallel observations)based on the known accuracy of assignment of a single observation (as received from LDA or kNN techniques)

In this example, let the accuracy of a single observation be 93%, the general sample be *N =* 1000 observations, that contains *K* = 930 correct observations and *N – K* = 70 incorrectly assigned observations. The whole sample of *n =* 6 observations will be recognized incorrectly if  *k =* 3, 4, 5 or 6 out of 6 observations are incorrect. Let us calculate these probabilities using a known formula (<https://www.matburo.ru/tvart_sub.php?p=calc_gg_item>, accessed on July 16th, 2023):

Here *P* is the target probability of the incorrect assignment and is the number of combinations of *k* in *K* that equals (in different countries the notations may vary: <https://en.wikipedia.org/wiki/Combination>)

We have to sequentially calculate the probabilities of incorrect assignment for *k =* 3, 4, 5 or 6 out of 6 observations and sum them up. The calculation was performed using the online service provided by the website mentioned above. For *k* = 3 we obtained:

Similarly, for *k* = 4 we obtained *P*4 = 0.00028 and for *k* = 5, *P*5 = 0.00000. Therefore, the overall probability will be *P* = *P*3 + *P*4 = 0.0056, or 0.56%. So, the accuracy of discrimination of the whole sample of 6 observations will be **99.44%**. Using the provided formulas or the online service, the accuracies for other conditions can be easily obtained.

**Reference**

[1] E. Lima, A.G. Barroso, M.A. Sousa, O. Ferreira, R.E. Boto, J.R. Fernandes, P. Almeida, S.M. Silvestre, Ad.O. Santos, L.V. Reis, Picolylamine-functionalized benz[e]indole squaraine dyes: Synthetic approach, characterization and in vitro efficacy as potential anticancer phototherapeutic agents, *Eur. J. Med. Chem.* **229** (2022) 114071. https://doi.org/10.1016/j.ejmech.2021.114071.