Electronic Supplementary Information

Computer-aided solvent screening for the fractionation of wet microalgae biomass

Laura König-Mattern, Steffen Linke, Liisa Rihko-Struckmann, Kai Sundmacher

1 Materials and experimental methods

1.1 Cultivation and harvest

Phaeodactylum tricornutum was purchased as 1090-1b from the Culture Collection of Algae at Göttingen University (SAG). The cells were maintained in liquid culture in Mann and Myers Medium as used by Derwenskus et al. in shaker flasks.[1] Exponentially growing cultures served as inoculum for a 28L flat panel airlift photobioreactor (Subitec) equipped with warm-white LEDs (Sanlight e.U., P4W). Light intensity was measured with a PAR-sensor (LI-COR, LI19OR-DWC) and adjusted with increasing culture density. The pH was adjusted to 7.4 by automatic addition of NaOH and CO_2. The temperature was maintained at 20°C. Dissolved oxygen was measured using an optical pO_2 electrode (Hamilton, Visiferm DD ARC 12 H0). Nitrate and phosphate contents in the medium were daily measured using ion exchange chromatography (Methrohm, Compact IC Flex oven/SeS/PP/Deg) equipped with an anion column (Metrosepp A Supp 5/150/4.0). In case of low nitrogen and phosphate concentrations, the medium was supplemented by the respective anions. At the end of the exponential phase, cells were harvested by centrifugation and washed with MilliQ water. The resulting algal paste was aliquoted to 50 ml Falcon tubes and stored at 4°C in order to prevent cell disruption by freezing, and were rapidly consumed for experiments.

1.2 Determination of water content

1 g of wet biomass was subjected to dried and pre-weighed porcellain crucibles. The biomass weight was recorded before and after drying. The amount of water in the biomass was determined by subtracting the dry weight of the biomass from the initial weight. The moisture content is then calculated as the amount of water divided by the dry weight.

1.3 Solvents

For extraction of the algae biomass, the following solvents were used: hexane (98.0 %, Merck), ethanol (99.9 %, Merck), ethyl acetate (99.8 %, Merck), 1-butanol (99.8 %, Merck), 2-butanol (99.5 %, Merck), methyl acetate (99.5 %, Sigma-Aldrich), methyl propionate (98.0 %, Sigma-Aldrich), chloroform (99.8 %, Merck), methanol (99.8 %, Merck), and ethyl formate (97.0 %, Sigma-Aldrich).

1.4 FTIR spectroscopy of lipid extracts

Dried lipid extracts were redissolved in chloroform:methanol (1:1). The solution was pipetted onto a silicon microtiter plate with 96 spots (PIKE Technologies) until a smooth film of lipid extract was visible. The absorbance of the lipid extracts was recorded on a Nicolet 6700 (Thermo) equipped with XY-stage (PIKE Technologies) with DTGS detector. Spectra were recorded within a range of 3000-750 cm^{-1} wavelength. The device was flushed with N_2. The blank was read against air. 32 scans per sample were taken. The spectra were baseline-corrected and smoothed using the OMNIC software (Thermo). CO_2 peaks were excluded. Spectra were normalised to the absorbance of the peak at a wavelength of 2925 cm^{-1}

2 FTIR spectra of the lipid extracts

Fig. 1 shows the FTIR spectra of the extracts obtained by different solvents. In each spectrum, a lipid peak at 1745 cm^{-1} is clearly visible.

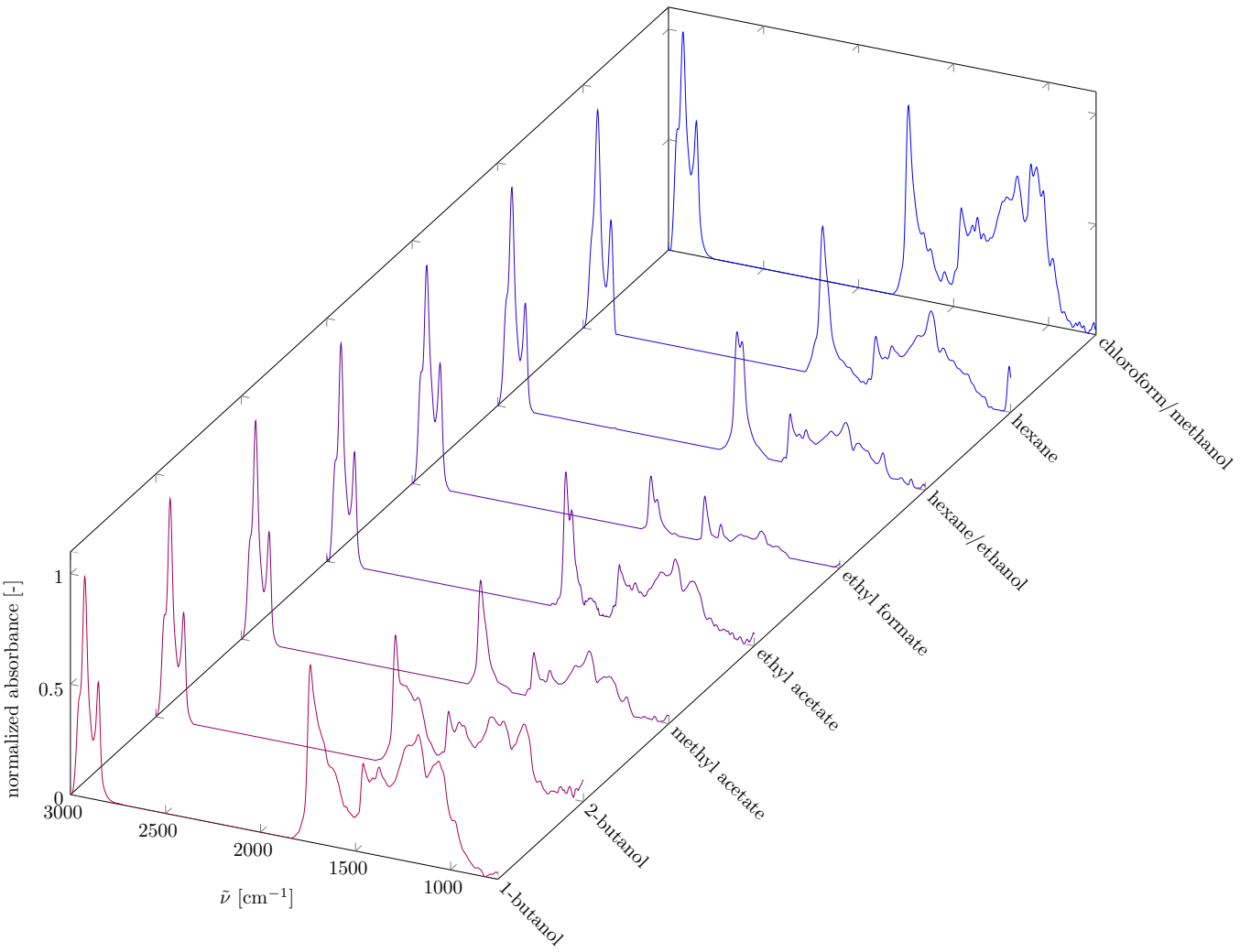


Figure 1: FTIR spectra of the extracts obtained by different solvents.

3 Computational solvent screening

3.1 Soft- and hardware

For molecules not present in the COSMObase13-01 and COSMObaseIL-19-01, quantum chemical density functional theory calculations were performed. In a first step, RDKit was used to generate conformers according to the approach proposed by Eberjer[2] using a force field. The resulting geometries were further optimised at quantum mechanical (QM) level using TURBOMOLE 3.7 and its *calculate* interface (version 2.1, 2009). In the calculations, the BP-86 functional with the def-TZVP bases set was used with the COSMO boundary condition and the standard COSMO cavity construction. Afterwards, a single point calculation with the optimised geometry was performed using the more accurate def2-TZVPD bases set for cavity construction at the FINE level. The automated screening was implemented in a python script (Python 3.7). The database described in section 3.2.1 of the main manuscript was loaded as a pandas dataframe (pandas 0.25.1) from which unsuitable solvent candidates were subsequently deleted. MP and BP were read from PubChem using PubChemPy 1.0.4. Missing BP and FP data as well as solubilities were calulated using CosmoPy 19.10. EHS properties were predicted using VEGA QSAR 1.1.5 whereby Tab. 3 of the main manuscript shows the models used. For LLE and

partition coefficient calculations, COSMOtherm was called from the python script via command line using automatically generated input files to start calculations. All results were stored in SQLite3 databases using sqlite 3.30.0. All COSMO-RS predictions were performed using COSMOtherm $v19$[3] and the BP_TZVPD_FINE_19.ctd parameter set. Calculations including ILs and DESs were implemented according to the so-called electroneutral approach, such that each IL and DES was treated as two separate compounds in a stoichiometric mixture.[4–6] Note that this approach is commonly chosen for predictions with ILs and DESs but cannot capture their stability in aqueous media. All calculations were performed on a Linux Ubuntu 16.04 computer (Intel i5-8500 processer at 3.00 GHz and 16 GB RAM).

3.2 List of all molecules contained in the database

A list of all molecules in the database can be found in the supplementary file `database.xlsx`. The solvent name, CAS-number (if applicable) and SMILES is given. For DESs, the molar fractions in the mixture of the two components are additionally stated.

3.3 List of manually added candidates after screening for green EHS properties

For some solvents, VEGA predicted unsuitable EHS properties, although, according to their safety data sheets, no health hazard, accute toxicity or environmental hazard was given. These solvents were manually re-added to the screening procedure. The list of manually added solvents can be found in the file `green_screening_manual_add.xlsx`.

3.4 Lists of solvent candidates after prescreening

After the prescreening, for solvents can be assigned to each biomass fraction. Lists of the solvents can be found in the following files:

- proteins: `proteins_prescreening.xlsx`
- carbohydrates: `carbohydrates_prescreening.xlsx`
- neutral lipids: `neutral_lipids_prescreening.xlsx`
- polar lipids: `polar_lipids_prescreening.xlsx`
- pigments: `pigments_prescreening.xlsx`

3.5 Lists of solvent candidates after screening for partial miscibility

After the prescreening step, the solvents were screened for partial miscibility with water, and suitable partition coefficients. The resulting solvents are given in the following files:

- neutral lipids: `neutral_lipids_partial_miscibility.xlsx`
- polar lipids: `polar_lipids_partial_miscibility.xlsx`
- pigments: `pigments_partial_miscibility.xlsx`

3.6 Solvent screening for hydrophilic fractions

For the screening of the hydrophilic fractions, two scenarios for the separation of carbohydrates and proteins were evaluated. In the first scenario, a suitable solvent has a high solubility for one of the hydrophilic fractions but not for all other fractions. In this way, the solube hydrophilic fraction would be transferred to the solvent whereas the insoluble fraction remains in the biomass. Hence, a solvent with a high selectivity for either carbohydrates or proteins is desired. After solubility calculations, as can be seen in Fig. 2, 16 solvents remain for the carbohydrates fraction and 18 for the protein fraction. Lists of solvents are given in the supplementary files `proteins_prescreening.xlsx` and `carbohydrates_prescreening.xlsx`. Most solvents do not have a high selectivity to one of the hydrophilic fractions or must be excluded due to toxicity. For some solvents, the MP was not given in the database and had to be excluded after manual checking. One IL was found (1-butyl-3-methylimidazolium dibutylphosphonium), that has no LLE with water but high carbohydrate solubility. Nevertheless, this IL also solubilises polar lipids very well and is therefore not selective enough. In the second scenario, carbohydrates and proteins are both extracted from the algae and then separated between two liquid phases. In this case, one fraction is transferred to the solvent while the other fraction remains in the aqueous phase. For the second option, two ILs were identified which have an LLE with water. The identified ILs are Cyphos IL 103 and 104. Nevertheless, after phase separation, carbohydrates and proteins remain in the aqueous phase together as predicted by their partition coefficients. In addition, both ILs also had a high solubility for pigments and were hence, not selective. As a consequence, no solvent could be identified for

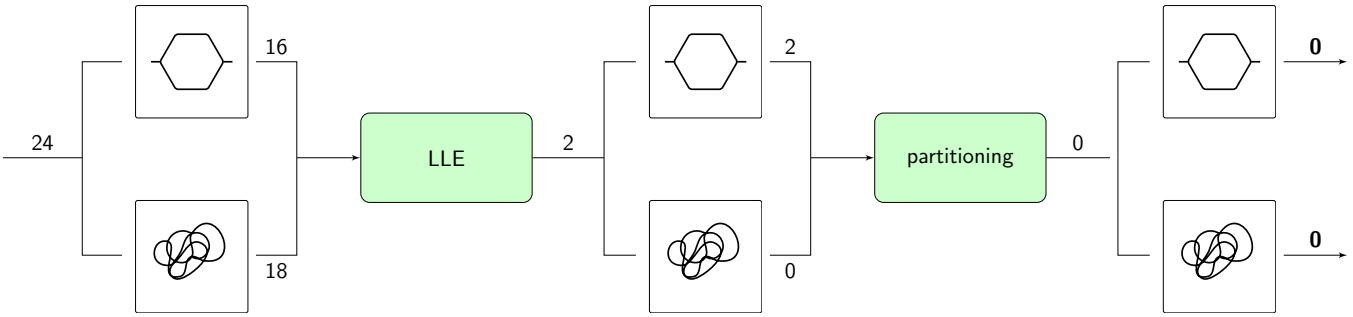


Figure 2: Solvent screening for the hydrophilic fractions: 24 solvents passed the prescreening step and were subsequently screened for the existence of an LLE, as well as the possibility to separate carbohydrates from proteins as predicted by partition coefficients.

the separation of carbohydrates and proteins from algae biomass, nor from aqueous solution. Hence for both fractions, water remains as the chosen solvent. Water is green, cheap, and readily available such that there is no advantage of using another solvent unless separation of carbohydrates and proteins is possible. Recently, aqueous two phase systems (ATPSs) were studied for the separation of carbohydrates and proteins from aqueous solutions and were already applied to microalgal biomass.[7,8] In this approach, an aqueous solvent solution forms a biphasic system after salt addition. Under certain conditions, this can lead to partitioning of the proteins to the organic phase while the carbohydrates remain in the aqueous phase.

4 Partition coefficients

The partition coefficients of proteins, carbohydrates, pigments and polar lipids in all solvents identified in the screening are shown in Fig. 3. Neutral lipids are omitted, since their solubility in solvents being partially miscible with water is predicted to be very low. For all fractions, partition coefficients approach zero with increasing water content.

5 Influence of water content in organic solvents on solubility and yield

The water content in the organic phase $x^{\text{org}}_{\text{H}_2\text{O}}$ correlates with lower solubility of lipophilic target molecules, see Fig 4a in which this was exemplified by the polar lipid fraction for the solvents which were evaluated in the lab (2-butanol, 1-butanol, methyl acetate, ethyl acetate, ethyl formate). Hence, one would expect that solvents predited to have a low solubility due to their high $x^{\text{org}}_{\text{H}_2\text{O}}$ result in lower yields. However, we observed the opposite effect, see Fig. 4b. We propose that a high $x^{\text{org}}_{\text{H}_2\text{O}}$ leads to increased accessibility to the lipophilic target molecules, and thus increased yield, see Fig. 4c. To obtain high yields in algae wet extraction, accessibility of the solvent to the cells is a key factor, which we achieved by using partially miscible solvents. As a consequence for the computational screening, the COSMO-RS predictions for LLEs must be accurate, which is given for the solvent-water systems we evaluated in the lab, see Fig. 4d.

6 Accuracy of solubility predictions

Fig. 5 summarises the solubilities predicted by COSMO-RS at 298.15 K and experimental data as taken from literature in the range of 293.15 and 298.15 K for glucose, l-alanine, palmitic acid and β-carotene in different solvents.[9–11] β-carotene was used instead of fucoxanthin since there is more experimental data available in literature. Additional QM-calculations for β-carotene were performed as described in section 3.1 of the ESI. Although COSMO-RS does not deliver quantitatively correct results, the majority of the predictions are qualitatively correct. This is the reason why we parameterized the solubility thresholds in the compuational screening relative to the reference solvents hexane, ethanol and water.

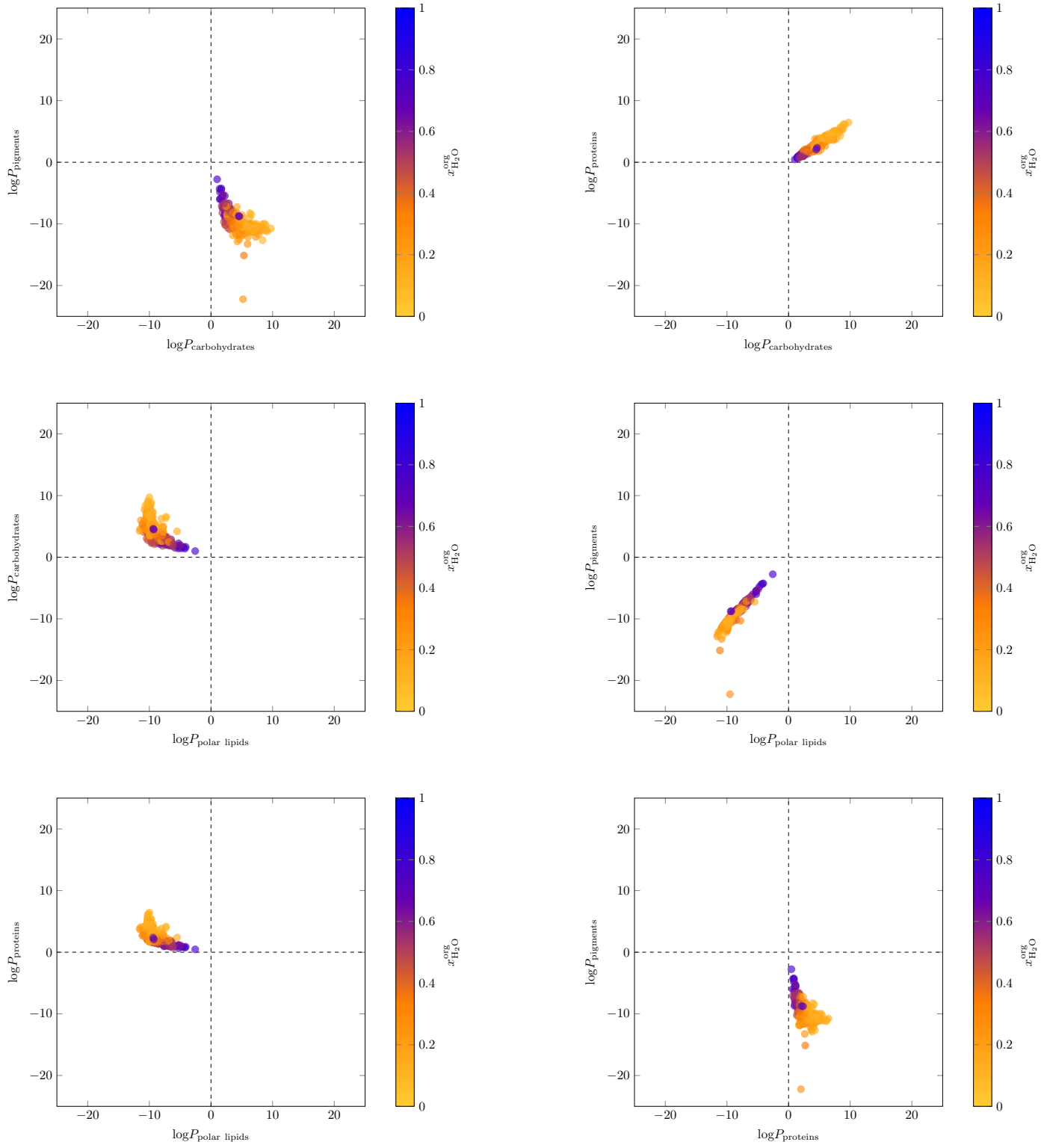
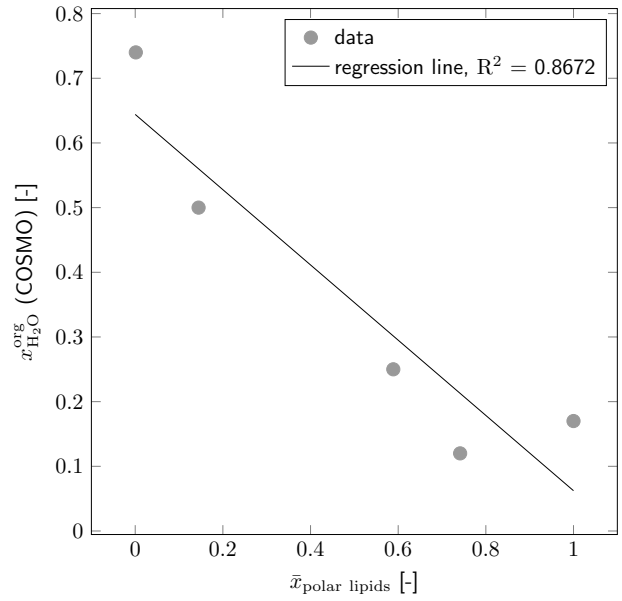
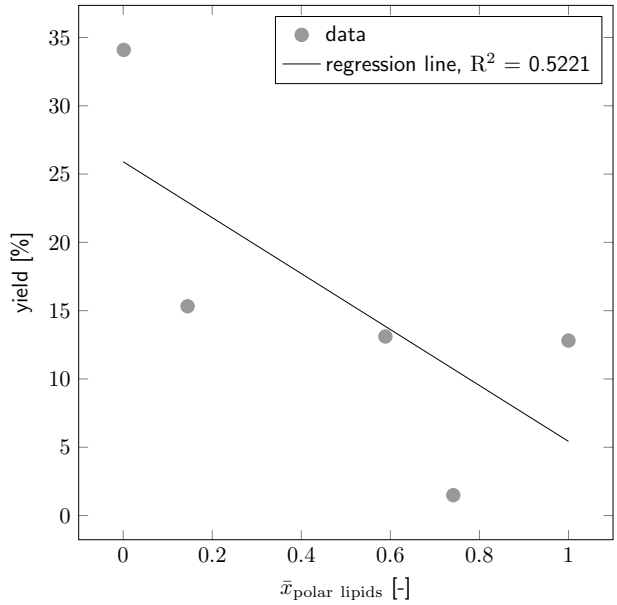


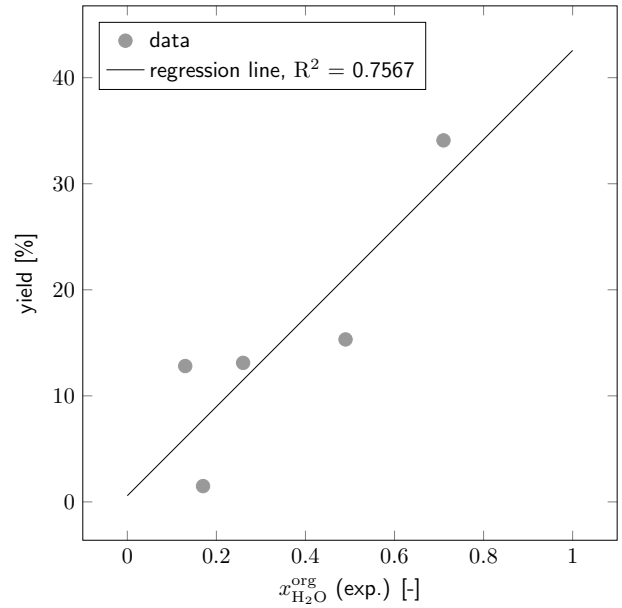
Figure 3: Partition coefficients of different biomass fractions extracted from the biomass by solvents being partially miscible with water. The colorbar represents the water content in the organic phase $x^{\text{org}}_{\text{H}_2\text{O}}$.



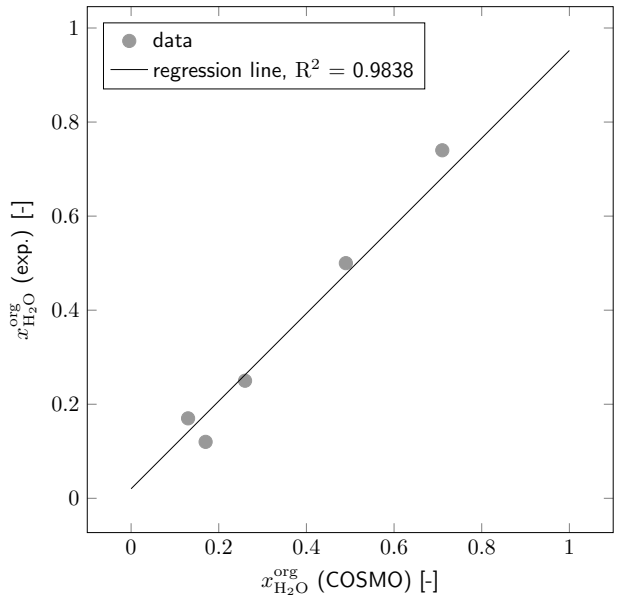
(a) Linear regression of the solubility of the polar lipid fraction $\bar{x}_{\text{polar lipids}}$ and the water content of the solvent $x^{\text{org}}_{\text{H}_2\text{O}}$, both predicted by COSMO-RS.



(b) Linear regression of the solubility of the polar lipid fraction $\bar{x}_{\text{polar lipids}}$ as predicted by COSMO-RS and the yield of all lipophilic compounds as determined in experiments.

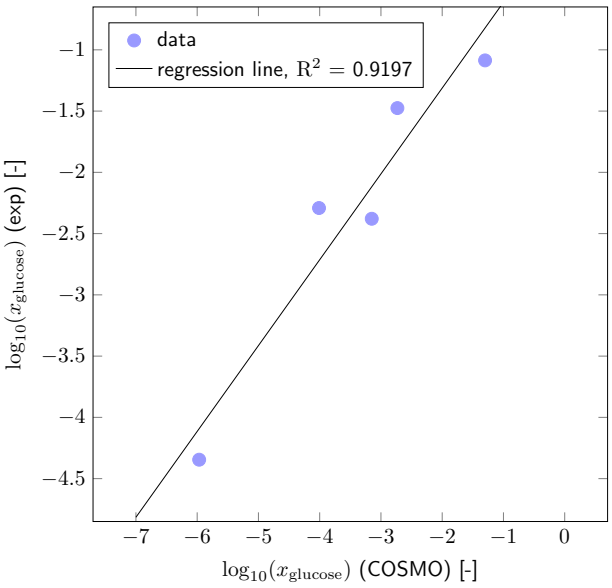


(c) Linear regression of the experimental values of $x^{\text{org}}_{\text{H}_2\text{O}}$ taken from literature (see manuscript) and the yield of all lipophilic compounds as determined in experiments.

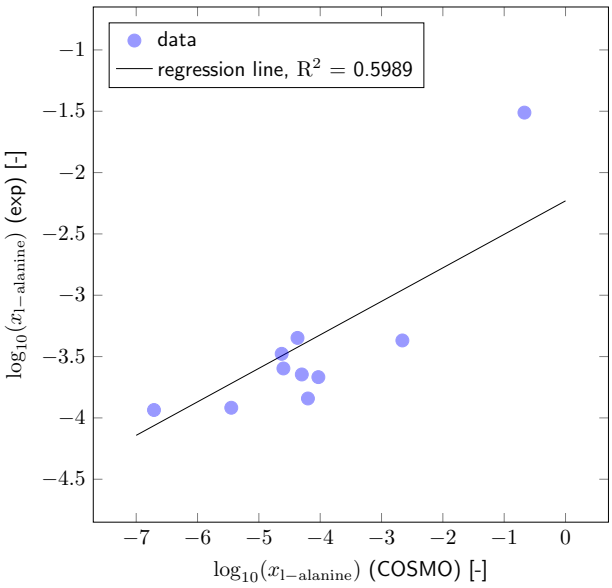


(d) Linear regression of the experimental values of $x^{\text{org}}_{\text{H}_2\text{O}}$ as predicted by COSMO-RS and experimental values taken from literature (see manuscript).

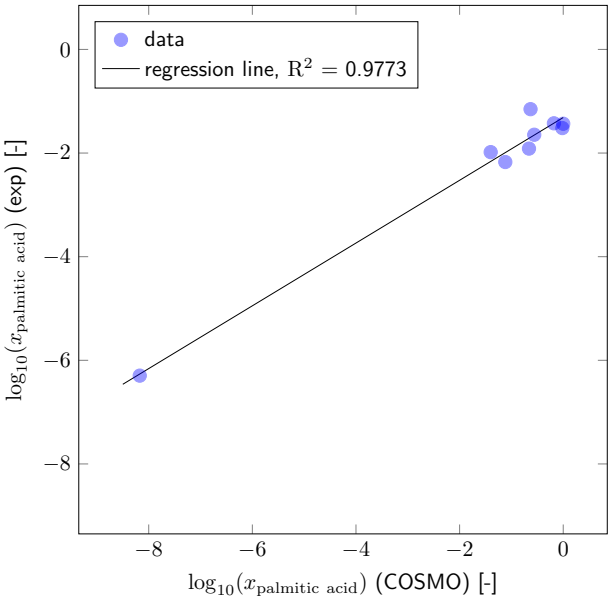
Figure 4: Regression analysis of values predicted by COSMO-RS and experimental data.



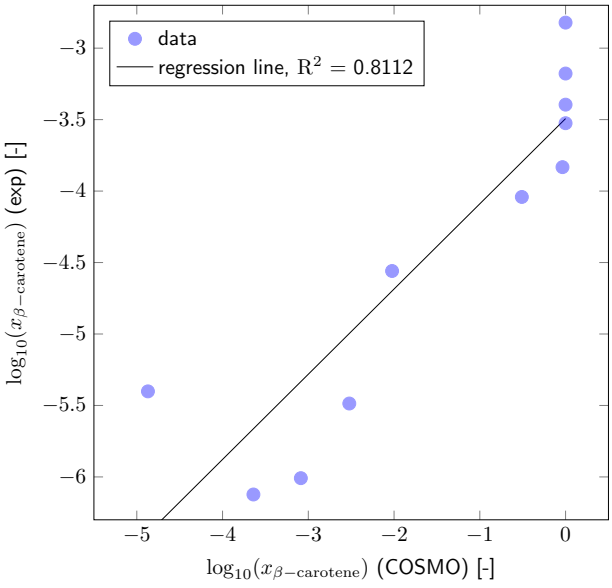
(a) Regression analysis of COSMO-RS solubility predictions for glucose at 298.15 K and experimental values at temperatures between 293.15 and 298.15 K.[11]



(b) Regression analysis of COSMO-RS solubility predictions for l-alanine at 298.15 K and experimental values at a temperature of 293.15 K.[9]



(c) Regression analysis of COSMO-RS solubility predictions for palmitic acid at 298.15 K and experimental values at temperatures between 293.15 and 298.15 K.[11]



(d) Regression analysis of COSMO-RS solubility predictions for β-carotene at 298.15 K and experimental values at ambient temperature.[10]

Figure 5: COSMO-RS solubility predictions and experimental results for different representative molecules from each biomass fraction. Note that for the pigment fraction β-carotene was used, since there was only limited experimental data for fucoxanthin available.

References

[1] F. Derwenskus, B. Schäfer, J. Müller, K. Frick, A. Gille, K. Briviba, U. Schmid-Staiger and T. Hirth, *Chem. Ing. Tech.*, 2020, **92**, 1780–1789.

[2] J.-P. Ebejer, G. M. Morris and C. M. Deane, *J. Chem. Inf. Model.*, 2012, **52**, 1146–1158.

[3] *COSMOtherm, Release 19, © 2019 COSMOlogic GmbH & Co. KG, a Dassault Systèmes Company*.

[4] F. Bezold, M. E. Weinberger and M. Minceva, *J. Chromatogr. A*, 2017, **1491**, 153–158.

[5] H. F. Hizaddin, A. Ramalingam, M. A. Hashim and M. K. Hadj-Kali, *J. Chem. Eng. Data*, 2014, **59**, 3470–3487.

[6] M. Diedenhofen and A. Klamt, *Fluid Phase Equilib.*, 2010, **294**, 31–38.

[7] C. A. Suarez Ruiz, J. Kwaijtaal, O. C. Peinado, C. van den Berg, R. H. Wijffels and M. H. Eppink, *ACS Sustainable Chem. Eng.*, 2020, **8**, 2441–2452.

[8] E. Suarez Garcia, C. A. Suarez Ruiz, T. Tilaye, M. H. Eppink, R. H. Wijffels and C. van den Berg, *Sep. Purif. Technol.*, 2018, **204**, 56–65.

[9] M. An, J. Qiu, D. Yi, H. Liu, S. Hu, J. Han, H. Huang, H. He, C. Liu, Z. Zhao, Y. Shi and P. Wang, *J. Chem. Eng. Data*, 2020, **65**, 549–560.

[10] N. E. Craft and J. H. Soares, *J. Agric. Food Chem.*, 1992, **40**, 431–434.

[11] *Chemister Database*, `http://chemister.ru/Database/search-en.php`, (accessed October 27, 2021).