Electronic Supplementary Information

Part 1 of 2

to the article

Ranking REACH registered neutral, ionizable and ionic organic chemicals based on their aquatic persistency and mobility

by

Hans Peter H. Arp,*[a] Trevor N. Brown,[b] Urs Berger[c]

and Sarah. E. Hale[a]

a. Norwegian Geotechnical Institute, Postboks 3930 Ullevål Stadion, NO-0806 Oslo, Norway.

b. ARC Arnot Research and Consulting Inc., 5536 Sackville St., Halifax, Nova Scotia, Canada

c. Department of Analytical Chemistry, Helmholtz Centre for Environmental Research – UFZ, Permoserstr. 15, DE-04318 Leipzig, Germany.

* Corresponding author: Email: hpa@ngi.no, Tel: + 47 950 20 667

Contents: 13 pages, 12 Tables, 2 Figures

Part 2 of the ESI is a spreadsheet containing substances considered in the PMOC ranking and identification approach, as well as relevant property information and literature data.

Section S1. Volatilization half-life calculations

1) **Volatilization** – Though volatilization half-lives are not commonly reported properties for compounds in the REACH dossier or elsewhere, volatilization half-lives can be estimated based on molecular properties as well as assumptions regarding the air and water flow conditions. The most conservative assumption in terms of persistency is to assume no turbulence in the water phase (including waves) and that atmosphere contains no wind. In this conservative case, volatilization can be estimated using the following model, based on pages 914-916 in Schwarzenbach et al (2013)[1]:

$$t_{1/2,volatilization} = 0.69/(v_{aw} * h) \qquad (V1)$$

Where $t_{1/2,volatilization}$, v_{aw} is the air-water exchange velocity and h the depth of the water. The v_{aw} term is determined by the following equation:

$$1/v_{aw} = 1/v_w + 1/(v_a * K_{aw}) \qquad (V2)$$

Where v_w is the mass transfer velocity of a substance in water, the v_a is the mass transfer velocity of a substance in air and K_{aw} is the dimensionless Henry's Law constant (adjusted for 12 °C, if possible based on data availability). The term v_a at 0 m/s windspeed is calculated as

$$v_a = (D_a/D_{water\ a})^{0.67} + v_{water\ a} \qquad (V3)$$

Where D_a is the diffusion coefficient of the compound in air ($D_a = 0.26 * (MW/18)^{-0.5}$, where MW is the molecular weight), $D_{water\ a}$ is the diffusion coefficient of water vapours in air (0.26 cm^2/s), and $v_{water\ a}$ is the velocity of water vapors in air at 0 m/s wind speed (0.3 cm/s).

The term v_w at 0 m/s windspeed is calculated as

$$v_w = (Sc_w/600)^{0.67} + v_{CO2\ w} \qquad (V4)$$

Where Sc_w is the Schmidt number of the compound in water ($Sc_w = 0.00893/(0.0000192*(MW/18)^{-0.5}$ at 0 m/s wind speed) and $v_{CO2\ w}$ is the mass transfer velocity of CO_2 in water (0.00065 cm/s).

Section S2. IFS QSARs for the P and M scores

The following text is a basic description of the multiple-linear regressions of individual molecular fragments and experimentally determined P and M scores, as used to calibrate the final IFS QSAR model to estimate P and M categories. Substances that were categorized as P and M were given a PMOC score of 4/5 (due to model uncertainty), substances with P and intM or intP and M were given a score of 2/4 (due to model uncertainty), and all other substances were given a score of 1.

Interpretation of the intercepts: For the M Score, the intercept is close to the maximum M Score of 5 (4.41). This means that for very small molecules, with no fragments present in the QSAR, a prediction of "mobile" will be given by default. As more atoms are added the possibility of becoming immobile increases depending on the functional groups added. For both of the P Scores the intercept is close to the minimum score of 1 (1.20 and 1.16). This means that very small molecules with no fragments in the QSAR would be non-persistent by default, and as functional groups are added the possibility of becoming persistent increases. These results are intuitive and are consistent with other QSARs previously developed.[2,3]

Interpretation of the fragments: As discussed in previous papers, interpretation of the fragments is not always straight forward.[2,4] This is because fragments are often overlapping and the contributions from each fragment need to be properly summed to compare between different QSARs. For some complex fragments comparison between different QSARs may not be possible. As an example, the effect of an aliphatic substituted chlorine and an aromatic substituted chlorine are compared for the M and P(G) scores.:

<u>M score aliphatic Cl:</u>

fragment #16: -0.14 (any chlorine atom)

total effect from each aliphatic Cl: -0.14

<u>M score aromatic Cl:</u>

fragment #16: -0.14 (any chlorine atom)

fragment #15: +0.11 (aromatic carbon with any functional group attached)

total effect from each aromatic Cl: -0.03

In general chlorine atoms slightly decrease the M score, however, specific substitution patterns on aromatic rings also play a role. Fragment #21 adds an additional -0.24 for a specific chlorine substitution pattern. Other more general fragments (9,10,25,34,36,39) have positive or negative regression coefficients for general aromatic substitution patterns, which could include chlorine atoms. It is also debatable if the effect of fragment #18 should be included in this comparison (any carbon atom, -0.15). The overall effect of a C-Cl group will include this additional -0.15 contribution, but comparing vs. an unsubstituted carbon or comparing aromatic vs. aliphatic substitution it does not make sense to include this contribution as the effect will cancel out.

<u>P(G) score aliphatic Cl:</u>

fragment #3: +2.49 (any aliphatic atom)

fragment #43: -1.79 (any chlorine atom)

total effect from each aliphatic Cl: +0.7

P(G) score aromatic Cl:

fragment #3: +2.49 (any aliphatic atom)

fragment #5: +1.45 (aromatic carbon - chlorine bond)

fragment #43: -1.79 (any chlorine atom)

total effect from each aromatic Cl: +2.15

Accounting for the total effect of a C-Cl group is more complicated than the M score. To include the carbon atom these additional factors need to be included:

P(G) score aliphatic C of C-Cl group:

fragment #3: +2.49 (any aliphatic atom)

fragment #10: +0.65 (any carbon atom)

fragment #38: -1.19 (any bond between atoms)

total effect from each aliphatic Cl: +0.7 +1.95 = 2.65

P(G) score aromatic C of c-Cl group:

fragment #2: +2.51 (any aromatic atom)

fragment #10: +0.65 (any carbon atom)

fragment #38: -1.19 (any bond between atoms)

total effect from each aromatic Cl: +2.15 +1.97 = 4.12

A more relevant comparison may be the effect of replacing one aliphatic or aromatic hydrogen with one chlorine:

P(G) score remove one aliphatic hydrogen:

fragment #29: -(-0.57) (bond between hydrogen and an aliphatic carbon)

fragment #40: -(-1.45) (any hydrogen atom)

total effect from each aliphatic Cl: +0.7 +2.02 = 2.72

P(G) score remove one aromatic hydrogen:

fragment #8: -(+0.68) (bond between hydrogen and an aromatic carbon)

fragment #40: -(-1.45) (any hydrogen atom)

total effect from each aromatic Cl: +2.15 +0.77 = 2.92

This calculation may be even more complex if the type of aliphatic carbon is changed by removing a hydrogen, for example if the carbon becomes a quaternary carbon (fragment #16) an additional +0.27 contribution is gained, or if the carbon was a methyl group (fragment #22) then the contribution of -0.15 is lost.

Interpretation of regression coefficients: In general the regression coefficients of M score seem reasonable. However, the P(G) and P(S) scores have both large positive and large negative regression coefficients that tend to balance out to give a score in the range 1-4. This is a common sign of over-fitting, and the prediction results may be especially unstable as the regression models are extrapolated outside of their training domain. The predictions from these QSARs should be treated with careful skepticism.

S2.1 PMOC QSAR Validation Statistics

Table S1: M Score Training Dataset Results Summary

	Predicted Mobile	Predicted Intermediate	Predicted Not Mobile
Expected Mobile	148	22	0
Expected Intermediate	67	300	33
Expected Not Mobile	0	20	73

Table S2: M Score Validation Dataset Results Summary

	Predicted Mobile	Predicted Intermediate	Predicted Not Mobile
Expected Mobile	109	26	3
Expected Intermediate	90	302	43
Expected Not Mobile	5	19	60

Table S3: M Score Training and Validation Summary Statistics

	Training Dataset	Validation Dataset
% Mobile Predictions Correct	68.8	53.4
% False Positives[a]	31.2	46.6
% Intermediate Predictions Correct	87.7	87
% Intermediate False Negatives[b]	6.4	7.5
% Not Mobile Predictions Correct	68.9	56.6
% Not Mobile False Negatives[c]	0	2.8
% Total Correct	78.6	71.7

[a] Mobile predictions for chemicals expected to be intermediate or not mobile.
[b] Intermediate predictions for chemicals expected to be mobile. Remainder are intermediate predictions for chemicals expected to be not mobile.
[c] Not mobile predictions for chemicals expected to be mobile. Remainder are not mobile predictions for chemicals expected to be intermediate.

Table S4: P(G) Score Training Dataset Results Summary

	Predicted Persistent	Predicted Labile
Expected Persistent	79	34
Expected Labile	52	231

Table S5: P(G) Score Validation Dataset Results Summary

	Predicted Persistent	Predicted Labile
Expected Persistent	67	66
Expected Labile	69	236

Table S6: P(G) Score Training and Validation Summary Statistics

	Training Dataset	Validation Dataset
% Persistent Predictions Correct	60.3	49.3
% False Positives[a]	39.7	50.7
% Labile Predictions Correct	87.2	78.1
% False Negatives[b]	12.8	21.9
% Total Correct	78.3	69.2

[a] Persistent predictions for chemicals expected to be labile.
[b] Labile predictions for chemicals expected to be persistent.

Table S7: P(S) Score Training Dataset Results Summary

	Predicted Persistent	Predicted Labile
Expected Persistent	67	36
Expected Labile	49	238

Table S8: P(S) Score Validation Dataset Results Summary

	Predicted Persistent	Predicted Labile
Expected Persistent	63	63
Expected Labile	70	238

Table S9: P(S) Score Training and Validation Summary Statistics

	Training Dataset	Validation Dataset
% Persistent Predictions Correct	57.8	47.4
% False Positives[a]	42.2	52.6
% Labile Predictions Correct	86.9	79.1
% False Negatives[b]	13.1	20.9
% Total Correct	78.2	69.4

[a] Persistent predictions for chemicals expected to be labile.
[b] Labile predictions for chemicals expected to be persistent.

S2.2 IFS QSAR Regression coefficients for the selected fragments

Table S10: M Score QSAR fragments and coefficients

#	Description	SMARTS code	Regression Coefficient	Std. Err.
1	aromatic nitrogen with hydrogen	[nX3H1+0]	1.04	0.25
2	chloro-aldehyde	[ClX1H0]-[CX3H0]=[OX1H0+0]	0.74	0.25
3	carbon-nitrogen double bond attached to an alkyl chain	[CX4H2]-[CX4H2]-[CX3H0]=[NX2H0+0]	0.63	0.24
4	ethene group, one substituent on either side	[CX3H1]=[CX3H1]	0.60	0.12
5	ethyne group	[CX2H0]#[CX2H0]	0.57	0.29
6	aromatic nitro group, two unsubstituted neighbouring carbons	[cX3H1]:[cX3H1]:[cX3H0;$(*-A)]-[NX3H0+0](=[OX1H0+0])=[OX1H0+0]	0.52	0.20
7	any oxygen atom	[5]	0.51	0.04
8	aromatic methoxy group	[CX4H3]-[OX2H0+0]-[cX3H0;$(*-A)]	0.48	0.12
9	aromatic methyl group beside another aliphatic substituent	[CX4H3]-[cX3H0;$(*-A)]:[cX3H0;$(*-A)]:[cX3H1]	0.45	0.17
10	aromatic secondary amine, para to another aliphatic substituent	[NX3H1+0]-[cX3H0;$(*-A)]:[cX3H1]:[cX3H1]:[cX3H0;$(*-A)]:[cX3H1]	0.36	0.08
11	aliphatic ether with a neighbouring substitution	[CX4H2]-[CX4H1]-[OX2H0+0]	0.26	0.05
12	alcohol group attached to quaternary carbon	[CX4H3]-[CX4H0]-[OX2H1+0]	0.23	0.11
13	aliphatic ether	[CX4H2]-[OX2H0+0]	0.21	0.03
14	any nitrogen atom	[6]	0.11	0.04
15	aromatic carbon with any aliphatic substituent	[cX3H0;$(*-A)]	0.11	0.03
16	any chlorine atom	[ClX1H0]	-0.14	0.06
17	two fused aromatic aromatic carbons with three neighboring unsubsituted positions	[cX3H0;!$(*-a);!$(*~A)]:[cX3H0;!$(*-a);!$(*~A)]:[cX3H1]:[cX3H1]:[cX3H1]	-0.15	0.07
18	any carbon atom	[7]	-0.15	0.01
19	ester group	[OX2H0+0]-[CX3H0]=[OX1H0+0]	-0.22	0.06
20	aromatic tertiary carbon with no ortho, meta or para substituents	[CX4H0]-[cX3H0;$(*-A)]:[cX3H1]:[cX3H1]:[cX3H1]	-0.23	0.13
21	aromatic chlorine with no ortho or meta substituents on one side	[ClX1H0]-[cX3H0;$(*-A)]:[cX3H1]:[cX3H1]	-0.24	0.11
22	isopropyl group	[CX4H3]-[CX4H1]-[CX4H3]	-0.26	0.08
23	any bromine atom	[BrX1H0]	-0.27	0.06
24	aliphatic alcohol group with three neighbouring substituents	[CX4H1]-[CX4H1]-[CX4H1]-[OX2H1+0]	-0.27	0.07
25	aromatic alkyl chain with aliphatic para substituent	[CX4H2]-[cX3H0;$(*-A)]:[cX3H1]:[cX3H1]:[cX3H0;$(*-A)]	-0.27	0.10
26	quaternary cabon with a methyl and an isopropyl attached	[CX4H3]-[CX4H0]-[CX4H1](-[CX4H2])-[CX4H2]	-0.31	0.09
27	n-butyl group (also counted for any longer chains)	[CX4H2]-[CX4H2]-[CX4H2]-[CX4H3]	-0.32	0.07
28	silicon with no hydrogens attached	[SiX4H0]	-0.36	0.10
29	linear alkyl chain with some substitutions	[CX4H1]-[CX4H2]-[CX4H1]-[CX4H1]-[CX4H2]	-0.38	0.17
30	terminal ethene group, aliphatic attachment	[CX4H2]-[CX3H1]=[CX3H2]	-0.38	0.14

31	any double bonded pair of aliphatic atoms in a ring	[A!#1x2+0]=[A!#1x2+0]	-0.42	0.15
32	aromatic carbonyl group with no ortho or para substituents	[OX1H0+0]=[CX3H0]-[cX3H0;$(*-A)](:[cX3H1]:[cX3H1]):[cX3H1]:[cX3H1]	-0.42	0.12
33	any ether	[OX2H0+0]	-0.48	0.06
34	aromatic ether with one ortho substituent and no meta substiuents	[cX3H1]:[cX3H0;$(*-A)]:[cX3H0;$(*-A)](-[OX2H0+0]):[cX3H1]:[cX3H1]	-0.48	0.16
35	ethene group, alkyl chain on one side and two substituents on the other	[CX4H2]-[CX4H2]-[CX3H1]=[CX3H0]	-0.53	0.22
36	aromatic primary amine with a meta substituent	[NX3H2+0]-[cX3H0;$(*-A)]:[cX3H1]:[cX3H0;$(*-A)]	-0.58	0.23
37	trifluoro methyl group	[FX1H0]-[CX4H0](-[FX1H0])-[FX1H0]	-0.66	0.20
38	sulfate group	[OX1H0+0]=[SX4H0]=[OX1H0+0]	-0.80	0.22
39	aromatic nitro group with one ortho substituent	[cX3H1]:[cX3H1]:[cX3H0;$(*-A)]:[cX3H0;$(*-A)](:[cX3H1])-[NX3H0+0](=[OX1H0+0])=[OX1H0+0]	-0.93	0.32
40	any nitro group	[OX1H0+0]=[NX3H0+0]=[OX1H0+0]	-1.51	0.16
	intercept		4.41	0.07

Table S11: P(G) Score QSAR fragments and coefficients

#	Description	SMARTS code	Regression Coefficient	Std. Err.
1	any boron	[8]	3.02	1.63
2	any aromatic atom	[a]	2.51	1.42
3	any aliphatic atom	[A]	2.49	1.30
4	any fused aromatic carbon	[cX3H0;!$(*-a);!$(*~A)]	2.18	0.74
5	aromatic chlorine	c-Cl	1.45	0.47
6	siloxane (Si-O-Si)	[SiX4]-[OX2H0]-[SiX4]	1.34	0.52
7	sulfur with aromatic attachment	S-c	0.75	0.86
8	aromatic carbon - hydrogen bond	c-[9]	0.68	0.73
9	three neighbouring substituted aromatic carbons	[cX3H1]:[cX3H1]:[cX3H1]	0.65	0.41
10	any carbon	[7]	0.65	1.08
11	any aliphatic nitrogen attached to an aromatic carbon	c-N	0.54	0.40
12	tertiary amine with any three carbon attachments	[7]-[NX3](-[7])-[7]	0.50	0.48
13	aromatic ether	c-O	0.39	0.33
14	phosphorus-oxygen single bond	P-O	0.37	0.47
15	aromatic-aliphatic carbon carbon bond	c-C	0.35	0.41
16	quaternary carbon	[CX4H0]	0.27	0.15
17	aliphatic ketone	CC(=O)C	0.22	0.26
18	aliphatic ester	CC(=O)OC	0.15	0.17
19	any aromatic attached group also double bonded to oxygen	[cX3H0;$(*-A)]-*=[OX1H0+0]	0.12	0.24
20	non-terminal propyl chain	[CX4H2]-[CX4H2]-[CX4H2]	0.10	0.08
21	number of rings		0.07	0.06
22	methyl	[CX4H3]	-0.15	0.11
23	non-terminal ethyl chain	[CX4H2]-[CX4H2]	-0.15	0.10
24	anhydrous phthalate group	[OX1H0+0]=[CX3H0]-[OX2H0+0]-[CX3H0]=[OX1H0+0]	-0.32	0.58
25	aliphatic secondary amine	C[NH]C	-0.36	0.41
26	aliphatic carbon-carbon bond	C-C	-0.45	0.25
27	carbon-nitrogen double bond	C=N	-0.48	0.76

28	nitrogen-oxygen double bond	O=N	-0.52	0.50
29	aliphatic carbon - hydrogen bond	C-[9]	-0.57	0.36
30	aliphatic ether	COC	-0.68	0.26
31	aliphatic primary amine	C[NH2]	-0.79	0.59
32	nitrogen-nitrogen single bond	N-N	-0.82	0.67
33	any oxygen	[5]	-0.97	0.91
34	any nitrogen	[6]	-1.02	0.88
35	ortho unsubstituted aromatic carbons	[cX3H1]:[cX3H1]	-1.04	0.77
36	alcohol group	OH	-1.04	0.37
37	peroxy group	O-O	-1.06	0.39
38	any bond	*~*	-1.19	0.57
39	cabon-nitrogen aromatic bond	c:n	-1.24	0.77
40	any hydrogen	[9]	-1.45	1.19
41	carbon-oxygen aromatic bond	c:o	-1.65	0.81
42	any bromine	[10]	-1.66	1.14
43	any chlorine	[11]	-1.79	1.07
44	any fluorine	[12]	-1.94	1.07
45	carbon-carbon double bond	C=C	-1.95	0.79
46	any sulfur	[13]	-1.99	1.02
47	any silicon	[14]	-1.99	0.83
48	carbon-carbon aromatic bond	c:c	-1.99	0.93
49	carbonyl group	C=O	-2.10	0.66
50	cyano group	C#N	-2.43	1.25
51	phosphorus-oxygen double bond	P=O	-2.48	1.18
52	any iodine	[15]	-2.52	2.07
53	carbon-carbon triple bond	C#C	-3.23	1.61
	intercept		1.20	0.19

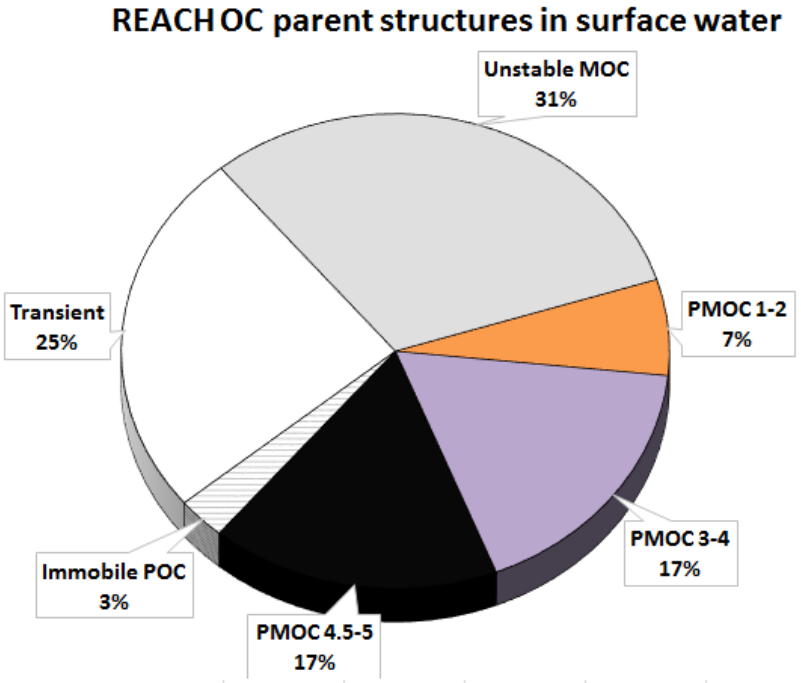
Table S12: P(S) Score QSAR fragments and coefficients

#	Description	SMARTS code	Regression Coefficient	Std. Err.
1	aromatic chlorine	c-Cl	3.93	1.23
2	any fused aromatic carbon	[cX3H0;!$(*-a);!$(*~A)]	3.78	1.22
3	single bond between two aromatic carbons (eg. biphenyl bridge)	c-c	3.40	1.97
4	sulfur with aromatic attachement	S-c	2.70	1.24
5	aromatic carbon - hydrogen bond	c-[9]	2.30	1.02
6	aromatic-aliphatic carbon carbon bond	c-C	2.29	0.97
7	aromatic ether	c-O	1.88	0.93
8	any aliphatic nitrogen attached to an aromatic carbon	c-N	1.71	0.97
9	any aliphatic atom	[A]	1.49	1.02
10	sulfur double bonded to carbon	S=C	1.47	1.05
11	any boron	[8]	1.25	1.40
12	any aromatic atom	[a]	1.09	1.59
13	tertiary amine with any three carbon attachments	[7]-[NX3](-[7])-[7]	0.93	0.56
14	aliphatic chlorine	C-Cl	0.79	0.71
15	aromatic primary or secondary amine	c[NX3;H1,H2]	0.65	0.56
16	three neighbouring unsubstituted aromatic carbons	[cX3H1]:[cX3H1]:[cX3H1]	0.60	0.74
17	any phosphorus	[16]	0.52	1.40

18	four neighbouring unsubstituted aromatic carbons	[cX3H1]:[cX3H1]:[cX3H1]:[cX3H1]	0.45	0.92
19	quaternary carbon	[CX4H0]	0.37	0.28
20	carbon-nirtogen double bond	C=N	0.29	0.39
21	sulfur single bonded to an aliphatic carbon	S-C	0.22	0.44
22	aliphatic ketone	CC(=O)C	0.20	0.23
23	aliphatic ester	CC(=O)OC	0.16	0.14
24	aliphatic alcohol	C[OH]	0.16	0.38
25	silicon with single bond to aliphatic carbon	[Si]-C	0.16	0.37
26	CH2 group	[CX4H2]	0.06	0.09
27	CH1 group	[CX4H1]	0.05	0.18
28	number of rings	rings	0.02	0.06
29	aliphatic tertiary amine	CN(C)C	-0.20	0.56
30	any bond	*~*	-0.20	0.34
31	cyano group	C#N	-0.21	0.42
32	aliphic primary amine	C[NH2]	-0.22	0.45
33	anhydrous phthalate group	[OX1H0+0]=[CX3H0]-[OX2H0+0]-[CX3H0]=[OX1H0+0]	-0.29	0.54
34	nitrogen-oxygen double bond	O=N	-0.34	0.39
35	aliphic ether	COC	-0.37	0.21
36	nitrogen-nitrogen single bond	N-N	-0.68	0.60
37	alcohol	O-[9]	-0.69	0.36
38	carbonyl group	C=O	-0.77	0.27
39	any oxygen	[5]	-0.79	0.87
40	carbon-oxygen aromatic bond	c:o	-0.87	1.23
41	cabon-nitrogen aromatic bond	c:n	-0.94	0.92
42	peroxy group	O-O	-0.99	0.45
43	any bromine	[10]	-1.00	0.99
44	any nitrogen	[6]	-1.04	0.85
45	ortho unsubstituted aromatic carbons	[cX3H1]:[cX3H1]	-1.14	0.83
46	any carbon	[7]	-1.22	0.83
47	any fluorine	[12]	-1.26	0.93
48	any hydrogem	[9]	-1.33	0.89
49	carbon-carbon aromatic bond	c:c	-1.52	1.04
50	any iodine	[15]	-1.70	1.89
51	any silicon	[14]	-1.82	1.17
52	and sulfur	[13]	-1.90	1.02
53	any chlorine	[11]	-2.16	1.15
54	phosphorus-oxygen double bond	P=O	-2.47	1.25
	intercept		1.16	0.19

Section S3. Distribution of PS, M and PMOC scores for surface water

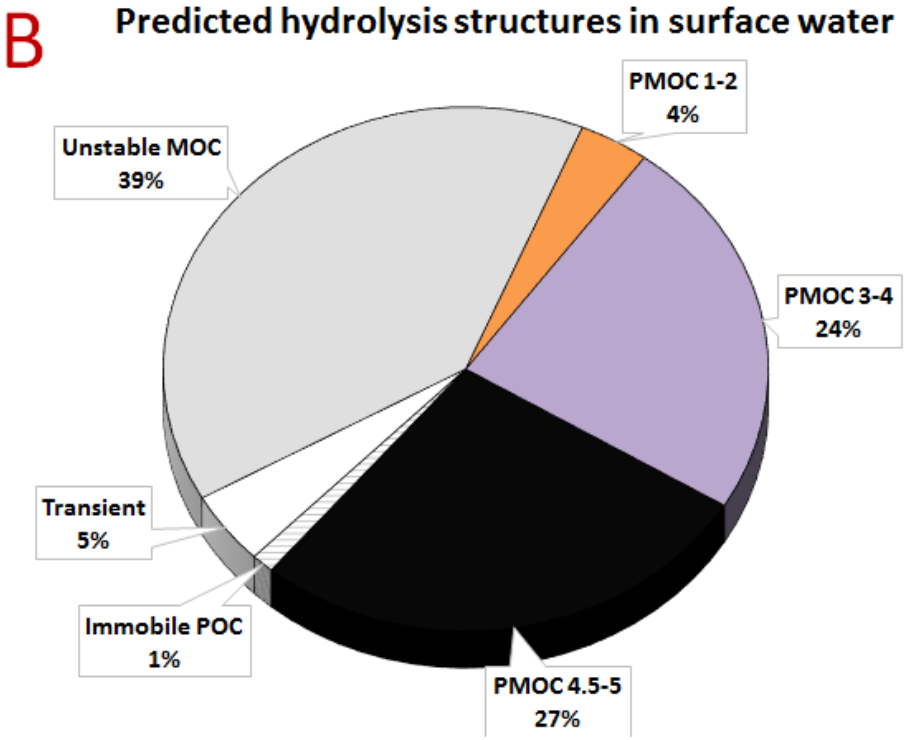
A



Sum		443	823	812	767	2298	*n*
P4	1	138	218	177	**151**	**379**	1064
P3	0	10	38	48	**36**	**126**	258
P2	1	24	60	87	106	315	593
P1	1	259	475	472	440	1164	2811
no P data	9	12	32	28	34	314	429
P vs M	**no M data**	**M1**	**M2**	**M3**	**M4**	**M5**	**Sum**
Sum	0%	9%	16%	16%	15%	45%	%
4	0%	3%	4%	3%	**3%**	**7%**	21%
3	0%	0%	1%	1%	**1%**	**2%**	5%
2	0%	0%	1%	2%	2%	6%	12%
1	0%	5%	9%	9%	9%	23%	55%
no P data	0%	0%	1%	1%	1%	6%	8%
P vs M	**no M data**	**1**	**2**	**3**	**4**	**5**	**Sum**

Figure S3A. Distribution of PMOC and non-PMOC categories in surface water for all structures considered in this study as pie charts, as well as the distribution of P vs M-scores following the PMOC scoring chart as presented in Figure 1 for the 5515 unique REACH OC structures considered.

B



Sum		75	276	325	452	3914	*n*
P4	0	55	143	115	**153**	**921**	1387
P3	0	6	14	38	**50**	**251**	359
P2	0	4	38	73	98	882	1095
P1	0	10	80	95	144	1847	2176
no P data	1	0	1	4	7	13	26
P vs M	**no M data**	**M1**	**M2**	**M3**	**M4**	**M5**	**Sum**
Sum	0%	1%	5%	6%	9%	76%	*%*
4	0%	1%	3%	2%	**3%**	**18%**	27%
3	0%	0%	0%	1%	**1%**	**5%**	7%
2	0%	0%	1%	1%	2%	17%	21%
1	0%	0%	2%	2%	3%	36%	42%
no P data	0%	0%	0%	0%	0%	0%	1%
P vs M	**no M data**	**1**	**2**	**3**	**4**	**5**	**Sum**

Figure S3B. Distribution of PMOC and non-PMOC categories in surface water for all structures considered in this study as pie charts, as well as the distribution of P vs M-scores following the PMOC scoring chart as presented in Figure 1 for the 5043 unique hydrolysis structures.

References

1. Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M., *Environmental Organic Chemistry*. 2. ed.; John Wiley & Sons: Hoboken, 2003.
2. Brown, T. N.; Arnot, J. A.; Wania, F., Iterative fragment selection: A group contribution approach to predicting fish biotransformation half-lives. *Environ. Sci. Technol.* **2012,** *46*, (15), 8253-8260.
3. Arnot, J. A.; Brown, T. N.; Wania, F., Estimating screening-level organic chemical half-lives in humans. *Environ. Sci. Technol.* **2013,** *48*, (1), 723-730.
4. Brown, T. N., Predicting hexadecane–air equilibrium partition coefficients (L) using a group contribution approach constructed from high quality data. *SAR and QSAR in Environmental Research* **2013,** *25*, (1), 51-71.
5. Breedveld, G. D.; Pelletier, E.; St Louis, R.; Cornelissen, G., Sorption characteristics of polycyclic aromatic hydrocarbons in aluminum smelter residues. *Environ. Sci. Technol.* **2007,** *41*, (7), 2542-2547.
6. Cornelissen, G.; Pettersen, A.; Broman, D.; Mayer, P.; Breedveld, G. D., Field testing of equilibrium passive samplers to determine freely dissolved native polycyclic aromatic hydrocarbon concentrations. *Environ. Toxicol. Chem.* **2008,** *27*, (3), 499-508.
7. Cornelissen, G.; Pettersen, A.; Nesse, E.; Eek, E.; Helland, A.; Breedveld, G. D., The contribution of urban runoff to organic contaminant levels in harbour sediments near two Norwegian cities. *Mar. Pollut. Bull.* **2008,** *56*, (3), 565-573.
8. Barthe, M.; Pelletier, E.; Breedveld, G. D.; Cornelissen, G., Passive samplers versus surfactant extraction for the evaluation of PAH availability in sediments with variable levels of contamination. *Chemosphere* **2008,** *71*, (8), 1486-1493.
9. Cornelissen, G.; Wiberg, K.; Broman, D.; Arp, H. P. H.; Persson, Y.; Sundqvist, K.; Jonsson, P., Freely Dissolved Concentrations and Sediment-Water Activity Ratios of PCDD/Fs and PCBs in the Open Baltic Sea. *Environ. Sci. Technol.* **2008,** *42*, (23), 8733-8739.
10. Yeo, H. G.; Choi, M.; Chun, M. Y.; Sunwoo, Y., Gas/particle concentrations and partitioning of PCBs in the atmosphere of Korea. *Atmos. Environ.* **2003,** *37*, (25), 3561-3570.
11. van Noort, P. C. M.; Cornelissen, G.; ten Hulscher, T. E. M.; Vrind, B. A.; Rigterink, H.; Belfroid, A., Slow and very slow desorption of organic compounds from sediment: influence of sorbate planarity. *Water Res.* **2003,** *37*, (10), 2317-2322.
12. Oen, A. M. R.; Breedveld, G. D.; Kalaitzidis, S.; Christanis, K.; Cornelissen, G., How quality and quantity of organic matter affect polycyclic aromatic hydrocarbon desorption from Norwegian harbor sediments. *Environ. Toxicol. Chem.* **2006,** *25*, (5), 1258-1267.
13. Cornelissen, G.; Elmquist, M.; Groth, I.; Gustafsson, O., Effect of sorbate planarity on environmental black carbon sorption. *Environ. Sci. Technol.* **2004,** *38*, (13), 3574-3580.
14. Cornelissen, G.; Breedveld, G. D.; Kalaitzidis, S.; Christanis, K.; Kibsgaard, A.; Oen, A. M. P., Strong sorption of native PAHs to pyrogenic and unburned carbonaceous geosorbents in sediments. *Environ. Sci. Technol.* **2006,** *40*, (4), 1197-1203.
15. Arp, H. P. H.; Schwarzenbach, R. P.; Goss, K. U., Equilibrium sorption of gaseous organic chemicals to fiber filters used for aerosol studies. *Atmos. Environ.* **2007,** *41*, (37), 8241-8252.
16. Schenker, U.; MacLeod, M.; Scheringer, M.; Hungerbuhler, K., Improving data quality for environmental fate models: A least-squares adjustment procedure for harmonizing physicochemical properties of organic compounds. *Environ. Sci. Technol.* **2005,** *39*, (21), 8434-8441.