Supplementary Information

Table S1. Full component SMILES library for hydrogen bond acceptors and donors

Molecule Name	SMILES	Notes
Choline Chloride	C[N+](C)(C)CCO.[Cl-]	Ionic
Betaine Chloride	C[N+](C)(C)CC(=O)O.[Cl-]	Ionic
Sodium Propionate	CCC(=O)[O-].[Na+]	Ionic
L-Proline	OC(=O)[C@@H]1CCCN1	Amino acid
Betaine	C[N+](C)(C)CC(=O)[O-]	Zwitterion
Urea	C(=O)(N)N	Amide
Glycerine	C(C(CO)O)O	Polyol
Ethylene Glycol	C(CO)O	Diol
1,2-Butanediol	CCC(CO)O	Diol
1,3-Butanediol	CC(CCO)O	Diol
1,4-Butanediol	C(CCO)CO	Diol
2,3-Butanediol	CC(C(C)O)O	Diol
1,6-Hexanediol	C(CCCO)CCO	Diol
Malic Acid	C(C(C(=O)O)O)C(=O)O	Carboxylic acid
Malonic Acid	C(C(=O)O)C(=O)O	Carboxylic acid
Citric Acid	C(C(=O)O)C(CC(=O)O)(C(=O)O)O	Carboxylic acid
Levulinic Acid	CC(=O)CCC(=O)O	Keto acid
Lactic Acid	CC(C(=O)O)O	Hydroxy acid
Oxalic Acid	C(=O)(C(=O)O)O	Carboxylic acid
Fructose	C1[C@H]([C@H]([C@@H](C(O1)(CO)O)O)O)O	Sugar (Ketose)
Glucose	C([C@@H]1[C@H]([C@@H]([C@H](C(O1)O)O)O)O)O	Sugar (Aldose)
Sucrose	C([C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O[C@]2([C@H]([C@@H]([C@H](O2)CO)O)O)CO)O)O)O)O	Disaccharide
Sorbitol	C([C@H]([C@H]([C@@H]([C@H](CO)O)O)O)O)O	Sugar alcohol

Menthol	CC1CCC(C(C1)C(C)C)O	Terpenoid
Thymol	CC1=CC(=C(C=C1)C(C)C)O	Phenolic
N4444Br	CCCC[N+](CCCC)(CCCC)CCCC.[Br-]	Tetrabutylammonium bromide
Lauric Acid	CCCCCCCCCCCC(=O)O	Fatty acid (C_{12})
Dodecanol	CCCCCCCCCCCCO	Fatty alcohol (C_{12})
Octanoic Acid	CCCCCCCC(=O)O	Fatty acid (C_8)

Table S2. SMARTS patterns utilized for functional group recognition

Group Name	SMARTS Pattern	F_{di}	F_{pi}^2	E_{hi}	V_{mi}
Methyl (-CH_3)	[CX4H3]	420	0	0	33.5
Methylene (-CH_2-)	[CX4H2]	270	0	0	16.1
Methine (>CH-)	[CX4H1]	80	0	0	-1
Quaternary Carbon (>C<)	[CX4H0]	70	0	0	-19.2
Alkenyl (=CH-)	[CX3H1]	200	0	0	13.5
Alkenyl (>C=)	[CX3H0]	70	0	0	-5.5
Carboxyl (-COOH)	[CX3](=O)[OX2H1]	530	420	10000	28.5
Ester (-COO-)	[CX3](=O)[OX2H0]	390	490	7000	18
Alcohol (-OH)	[OX2H1;!$(OC=O)]	210	500	20000	10
Ether/Sugar (-O-)	[OX2H0;!$(OC=O)]	100	400	3000	3.8
Ketone (>C=O)	[CX3H0](=O)	290	770	2000	10.8
Aromatic Carbon	[c]	211.6	18.3	0	8.7
Chloride (-Cl)	[Cl]	450	550	400	24
Amine (-NH_2)	[NX3H2]	280	0	8400	19.2
Amide (-NH-)	[NX3H1]	160	210	3100	4.5
Tertiary Amine (>N-)	[NX3H0]	20	800	5000	-9

S3. Complete High-Throughput Screening (HTS) output dataset

The complete High-Throughput Screening output dataset is open-source and freely accessible via GitHub at https://github.com/iehoshva/early-screening-des.