

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule ID	CAS	SMILES	Median Percent Removal [%]	Number of Measurements	Data Source	Included in Training
Molecule 1	7664-41-7	N	30.43	3665	Cali Set	FALSE
Molecule 2	1746-01-6	<chem>ClC=1C=C2OC3=CC(CI)=C(CI)C=C3OC2=CC1Cl</chem>	0	17	Cali Set	TRUE
Molecule 3	72-20-8	<chem>Cl[C@]12[C@@]3([C@]([C@](Cl)(C1(CI)Cl)C(CI)=C2Cl)([C@]4(C[C@]3([C@H]5[C@@H]4O5)[H])([H])([H])[H])</chem>	90	24	Cali Set	TRUE
Molecule 4	76-44-8	<chem>ClC1=C(CI)C2(CI)C3(CI)C=CC3C1(CI)C2(CI)Cl</chem>	90	24	Cali Set	TRUE
Molecule 5	58-89-9	<chem>Cl[C@H]1[C@@H](Cl)[C@H](Cl)[C@@H](Cl)[C@@H](Cl)[C@@H]1Cl</chem>	0	1	Cali Set	FALSE
Molecule 6	309-00-2	<chem>Cl[C@]12[C@@]3([C@]([C@](Cl)(C1(CI)Cl)C(CI)=C2Cl)([C@]4(C[C@]3(C=C4)[H])([H])([H])[H])</chem>	90	24	Cali Set	TRUE
Molecule 7	1024-57-3	<chem>Cl[C@]12[C@@]3([C@]([C@](Cl)(C1(CI)Cl)C(CI)=C2Cl)([C@@H](Cl)[C@@H]4[C@H]3O4)[H])([H])</chem>	90	24	Cali Set	TRUE
Molecule 8	1031-07-8	<chem>O=S1(=O)OCC2C(CO1)C3(CI)C(CI)=C(CI)C2(CI)C3(CI)Cl</chem>	90	24	Cali Set	TRUE
Molecule 9	319-85-7	<chem>Cl[C@H]1[C@H](Cl)[C@@H](Cl)[C@H](Cl)[C@@H](Cl)[C@@H]1Cl</chem>	90	24	Cali Set	FALSE
Molecule 10	72-43-5	<chem>ClC(Cl)(Cl)C(C1=CC=C(OC)C=C1)C2=CC=C(OC)C=C2</chem>	87.64	22	Cali Set	TRUE
Molecule 11	319-84-6	<chem>Cl[C@H]1[C@@H](Cl)[C@H](Cl)[C@@H](Cl)[C@H](Cl)[C@@H]1Cl</chem>	90	24	Cali Set	FALSE
Molecule 12	75-27-4	<chem>ClC(Cl)Br</chem>	0	151	Cali Set	TRUE
Molecule 13	319-86-8	<chem>Cl[C@H]1[C@@H](Cl)[C@@H](Cl)[C@H](Cl)[C@@H](Cl)[C@@H]1Cl</chem>	90	25	Cali Set	FALSE
Molecule 14	12789-03-6	<chem>C1C2C(C(C1Cl)Cl)C3(C=C(C2(C3(CI)Cl)Cl)Cl)Cl</chem>	0	137	Cali Set	TRUE
Molecule 15	7421-93-4	<chem>O=CC1CC2C3C1C4(Cl)C(Cl)(Cl)C3(CI)C5(CI)C2C45Cl</chem>	90	24	Cali Set	TRUE
Molecule 16	72-54-8	<chem>ClC1=CC=C(C=C1)C(C2=CC=C(Cl)C=C2)C(Cl)Cl</chem>	90	24	Cali Set	TRUE
Molecule 17	72-55-9	<chem>ClC(Cl)=C(C1=CC=C(Cl)C=C1)C2=CC=C(Cl)C=C2</chem>	90	24	Cali Set	TRUE
Molecule 18	50-29-3	<chem>ClC1=CC=C(C=C1)C(C2=CC=C(Cl)C=C2)C(Cl)(Cl)Cl</chem>	90	24	Cali Set	TRUE
Molecule 19	60-57-1	<chem>Cl[C@@]12[C@]3([C@]([C@@](Cl)([C@]1(CI)Cl)C(CI)=C2Cl)([C@]4(C[C@]3([C@H]5[C@@H]4O5)[H])([H])([H])[H])</chem>	90	24	Cali Set	FALSE
Molecule 20	74-83-9	BrC	80	25	Cali Set	TRUE
Molecule 21	107-13-1	N#CC=C	0	37	Cali Set	TRUE
Molecule 22	84-66-2	<chem>O=C(OCC)C=1C=CC=CC1C(=O)OCC</chem>	83.75	64	Cali Set & Lit Set	TRUE
Molecule 23	53-70-3	<chem>C1=CC=C2C(=C1)C=CC3=CC4=C(C=CC5=CC=CC=C54)C=C32</chem>	90	33	Cali Set & Exp Set	TRUE
Molecule 24	127-18-4	<chem>C(=C(Cl)Cl)(Cl)Cl</chem>	58.49	49	Cali Set	TRUE
Molecule 25	122-66-7	<chem>C=1C=CC(=CC1)NNC=2C=CC=CC2</chem>	80	42	Cali Set	TRUE
Molecule 26	59-50-7	<chem>ClC1=CC=C(O)C=C1C</chem>	80	19	Cali Set	TRUE
Molecule 27	106-46-7	<chem>C1=CC(=CC=C1Cl)Cl</chem>	80	49	Cali Set	TRUE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 28	88-75-5	<chem>O=N(O)C=1C=CC=CC1O</chem>	80	36	Cali Set	TRUE
Molecule 29	83-32-9	<chem>C=1C=C2C=CC=C3C2=C(C1)CC3</chem>	94	40	Cali Set & Exp Set	TRUE
Molecule 30	75-25-2	<chem>BrC(Br)Br</chem>	39.66	44	Cali Set	TRUE
Molecule 31	193-39-5	<chem>C1=CC=C2C(=C1)C3=CC=C4C=CC5=CC=CC=6C=C2C3=C4C56</chem>	96.77	42	Cali Set & Exp Set	TRUE
Molecule 32	75-69-4	<chem>FC(Cl)(Cl)Cl</chem>	0	9	Cali Set	TRUE
Molecule 33	71-43-2	<chem>C1=CC=CC=C1</chem>	0	33	Cali Set	TRUE
Molecule 34	206-44-0	<chem>C1=CC=C2C(=C1)C3=CC=CC4=CC=CC2=C43</chem>	95.47	57	Cali Set & Exp Set	TRUE
Molecule 35	50-32-8	<chem>C1=CC=C2C3=C4C(=CC2=C1)C=CC5=C4C(=CC=C5)C=C3</chem>	94	42	Cali Set & Exp Set	TRUE
Molecule 36	78-59-1	<chem>O=C1C=C(C)CC(C)(C)C1</chem>	80	42	Cali Set	TRUE
Molecule 37	534-52-1	<chem>O=N(O)C=1C=C(C(O)=C(C1)C)N(O)=O</chem>	80	37	Cali Set	TRUE
Molecule 38	606-20-2	<chem>O=N(O)C1=CC=CC(=C1C)N(O)=O</chem>	80	34	Cali Set	TRUE
Molecule 39	12674-11-2	<chem>C1=CC(=CC(=C1)Cl)C2=CC(=CC(=C2)Cl)Cl</chem>	90	24	Cali Set	TRUE
Molecule 40	120-12-7	<chem>C=1C=CC=2C=C3C=CC=CC3=CC2C1</chem>	90	45	Cali Set & Exp Set	TRUE
Molecule 41	71-55-6	<chem>CC(Cl)(Cl)Cl</chem>	0	32	Cali Set	TRUE
Molecule 42	105-67-9	<chem>OC1=CC=C(C=C1)C</chem>	80	37	Cali Set	TRUE
Molecule 43	86-30-6	<chem>O=NN(C=1C=CC=CC1)C=2C=CC=CC2</chem>	80	42	Cali Set	TRUE
Molecule 44	75-01-4	<chem>ClC=C</chem>	0	32	Cali Set	TRUE
Molecule 45	218-01-9	<chem>C=1C=CC2=C(C1)C=CC=3C=4C=CC=CC4C=CC23</chem>	94	39	Cali Set & Exp Set	TRUE
Molecule 46	60-29-7	<chem>O(CC)CC</chem>	80	25	Cali Set	FALSE
Molecule 47	108-95-2	<chem>C1=CC=C(C=C1)O</chem>	97.19	57	Cali Set	TRUE
Molecule 48	53469-21-9	<chem>C1=CC(=C(C(=C1)Cl)Cl)C2=C(C=C(C(=C2)Cl)Cl</chem>	90	24	Cali Set	TRUE
Molecule 49	57-12-5	<chem>[C-]#N</chem>	11.96	6	Cali Set	TRUE
Molecule 50	91-20-3	<chem>C=1C=CC=2C=CC=CC2C1</chem>	96	43	Cali Set & Exp Set	TRUE
Molecule 51	107-06-2	<chem>ClCCCCl</chem>	80	23	Cali Set	FALSE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 52	79-00-5	<chem>ClCC(Cl)Cl</chem>	0	32	Cali Set	FALSE
Molecule 53	78-87-5	<chem>ClCC(Cl)C</chem>	80	24	Cali Set	FALSE
Molecule 54	121-14-2	<chem>O=N(=O)C1=CC=C(C(=C1)N(=O)=O)C</chem>	80	42	Cali Set	TRUE
Molecule 55	92-87-5	<chem>NC=1C=CC(=CC1)C2=CC=C(N)C=C2</chem>	80	34	Cali Set	TRUE
Molecule 56	51-28-5	<chem>O=N(=O)C1=CC=C(O)C(=C1)N(=O)=O</chem>	80	43	Cali Set	TRUE
Molecule 57	191-24-2	<chem>C1=CC2=C3C(=C1)C4=CC=CC5=C4C6=C(C(=C5)C=CC(=C36)C=C2</chem>	90	32	Cali Set & Exp Set	TRUE
Molecule 58	124-48-1	<chem>ClC(Br)Br</chem>	0	95	Cali Set	TRUE
Molecule 59	207-08-9	<chem>C1=CC=C2C=C3C4=CC=CC5=C4C(=CC=C5)C3=CC2=C1</chem>	90	33	Cali Set & Exp Set	TRUE
Molecule 60	101-55-3	<chem>BrC1=CC=C(OC=2C=CC=CC2)C=C1</chem>	80	34	Cali Set	TRUE
Molecule 61	84-74-2	<chem>O=C(OCCCC)C=1C=CC=CC1C(=O)OCCCC</chem>	77.27	69	Cali Set & Lit Set	TRUE
Molecule 62	621-64-7	<chem>O=NN(CCC)CCC</chem>	80	43	Cali Set	TRUE
Molecule 63	33213-65-9	<chem>Cl[C@@]12[C@]3([C@@]([C@@](Cl)([C@@]1(Cl)Cl)C(Cl)=C2Cl)(CO[S@](=O)OC3)[H])[H]</chem>	90	24	Cali Set	FALSE
Molecule 64	95-50-1	<chem>ClC=1C=CC=CC1Cl</chem>	80	35	Cali Set	TRUE
Molecule 65	100-02-7	<chem>O=N(=O)C1=CC=C(O)C=C1</chem>	80	36	Cali Set	TRUE
Molecule 66	85-68-7	<chem>O=C(OCC=1C=CC=CC1)C=2C=CC=CC2C(=O)OCCCC</chem>	80	42	Cali Set & Lit Set	TRUE
Molecule 67	91-94-1	<chem>C1=CC(=C(C=C1C2=CC(=C(C=C2)N)Cl)Cl)N</chem>	80	36	Cali Set	TRUE
Molecule 68	75-09-2	<chem>ClCCI</chem>	38.89	141	Cali Set	TRUE
Molecule 69	542-75-6	<chem>C(C=CCl)Cl</chem>	0	15	Cali Set	FALSE
Molecule 70	87-68-3	<chem>ClC(Cl)=C(Cl)C(Cl)=C(Cl)Cl</chem>	80	42	Cali Set	TRUE
Molecule 71	100-41-4	<chem>CCC1=CC=CC=C1</chem>	23.89	38	Cali Set	TRUE
Molecule 72	52663-62-4	<chem>C1=CC(=C(C(=C1)Cl)Cl)C2=C(C(=C(C=C2)Cl)Cl)Cl</chem>	90	24	Cali Set	TRUE
Molecule 73	98-95-3	<chem>O=N(=O)C=1C=CC=CC1</chem>	80	43	Cali Set	TRUE
Molecule 74	959-98-8	<chem>Cl[C@@]12[C@]3([C@@]([C@@](Cl)([C@@]1(Cl)Cl)C(Cl)=C2Cl)(CO[S@@](=O)OC3)[H])[H]</chem>	80	24	Cali Set	TRUE
Molecule 75	106-93-4	<chem>BrCCBr</chem>	0	9	Cali Set	TRUE
Molecule 76	120-82-1	<chem>ClC1=CC=C(Cl)C(Cl)=C1</chem>	80	35	Cali Set	TRUE
Molecule 77	120-83-2	<chem>ClC1=CC=C(O)C(Cl)=C1</chem>	80	34	Cali Set	TRUE
Molecule 78	14797-55-8	<chem>O=N(=O)[O-]</chem>	0	6	Cali Set	FALSE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 79	75-00-3	ClCC	54.66	30	Cali Set	TRUE
Molecule 80	117-81-7	CCCCC(CC)COC(=O)C1=CC=CC=C1C(=O)OCC(CC)CCCC	84.89	142	Cali Set & Lit Set	TRUE
Molecule 81	77-47-4	ClC=1C(Cl)=C(Cl)C(Cl)(Cl)C1Cl	80	45	Cali Set	TRUE
Molecule 82	108-88-3	CC1=CC=CC=C1	85.01	173	Cali Set	TRUE
Molecule 83	7005-72-3	ClC1=CC=C(OC=2C=CC=CC2)C=C1	80	35	Cali Set	TRUE
Molecule 84	79-34-5	ClC(Cl)C(Cl)Cl	0	32	Cali Set	FALSE
Molecule 85	12672-29-6	C1=C(C=C(C=C1Cl)Cl)C2=CC(=CC(=C2)Cl)Cl	90	24	Cali Set	TRUE
Molecule 86	85-01-8	C=1C=CC2=C(C1)C=CC=3C=CC=CC32	90	31	Cali Set	TRUE
Molecule 87	118-74-1	ClC=1C(Cl)=C(Cl)C(Cl)=C(Cl)C1Cl	80	44	Cali Set	TRUE
Molecule 88	117-84-0	O=C(OCCCCCCCC)C=1C=CC=CC1C(=O)OCCCCCCCC	80	39	Cali Set & Lit Set	TRUE
Molecule 89	131-11-3	O=C(OC)C=1C=CC=CC1C(=O)OC	80	44	Cali Set & Lit Set	TRUE
Molecule 90	86-73-7	C=1C=CC2=C(C1)C=3C=CC=CC3C2	96.77	42	Cali Set & Exp Set	TRUE
Molecule 91	129-00-0	C=1C=C2C=CC3=CC=CC=4C=CC(C1)=C2C34	92.67	53	Cali Set & Exp Set	TRUE
Molecule 92	1634-04-4	O(C)C(C)(C)C	30	27	Cali Set	TRUE
Molecule 93	95-57-8	ClC=1C=CC=CC1O	80	36	Cali Set	TRUE
Molecule 94	107-02-8	O=CC=C	37.5	26	Cali Set	TRUE
Molecule 95	11104-28-2	ClC1=CC=CC(=C1)C1=CC=CC=C1	90	24	Cali Set	TRUE
Molecule 96	62-75-9	O=NN(C)C	75	48	Cali Set	TRUE
Molecule 97	156-60-5	C(=CCl)Cl	80	24	Cali Set	FALSE
Molecule 98	74-87-3	ClC	80	52	Cali Set	TRUE
Molecule 99	11096-82-5	C1=CC(=C(C(=C1C2=C(C(=C(C=C2)Cl)Cl)Cl)Cl)Cl)Cl	90	24	Cali Set	TRUE
Molecule 100	88-06-2	ClC=1C=C(Cl)C(O)=C(Cl)C1	80	59	Cali Set	TRUE
Molecule 101	87-86-5	C1(=C(C(=C(C(=C1Cl)Cl)Cl)Cl)Cl)O	80	37	Cali Set	TRUE
Molecule 102	75-34-3	ClC(Cl)C	80	24	Cali Set	TRUE
Molecule 103	111-91-1	C(CCl)OCOCCCl	80	36	Cali Set	TRUE
Molecule 104	111-44-4	C(CCl)OCCCl	80	43	Cali Set	TRUE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 105	205-99-2	<chem>C=1C=CC2=C(C1)C=C3C4=CC=CC=C4C5=CC=CC2=C53</chem>	90	31	Cali Set	TRUE
Molecule 106	79-01-6	<chem>ClC=C(Cl)Cl</chem>	0	33	Cali Set	FALSE
Molecule 107	67-72-1	<chem>ClC(Cl)(Cl)C(Cl)(Cl)Cl</chem>	80	41	Cali Set	TRUE
Molecule 108	14808-79-8	<chem>O=S(=O)([O-])[O-]</chem>	0	66	Cali Set	FALSE
Molecule 109	11141-16-5	<chem>ClC1=CC=CC=C1C1=CC=CC=C1</chem>	90	24	Cali Set	TRUE
Molecule 110	56-55-3	<chem>C1=CC=C2C(=C1)C=CC3=CC4=CC=CC=C4C=C32</chem>	94	43	Cali Set & Exp Set	TRUE
Molecule 111	56-23-5	<chem>ClC(Cl)(Cl)Cl</chem>	0	32	Cali Set	FALSE
Molecule 112	75-35-4	<chem>ClC(Cl)=C</chem>	0	32	Cali Set	TRUE
Molecule 113	208-96-8	<chem>C=1C=C2C=CC=C3C=CC(C1)=C23</chem>	90	26	Cali Set & Exp Set	TRUE
Molecule 114	91-58-7	<chem>ClC=1C=CC=2C=CC=CC2C1</chem>	80	34	Cali Set	TRUE
Molecule 115	67-66-3	<chem>C(Cl)(Cl)Cl</chem>	4.12	403	Cali Set	TRUE
Molecule 116	39638-32-9	<chem>CC(C)(OC(C)(C)Cl)Cl</chem>	79.75	43	Cali Set	TRUE
Molecule 117	541-73-1	<chem>ClC1=CC=CC(Cl)=C1</chem>	80	34	Cali Set	TRUE
Molecule 118	108-90-7	<chem>ClC=1C=CC=CC1</chem>	0	32	Cali Set	FALSE
Molecule 119	3424-82-6	<chem>ClC(Cl)=C(C1=CC=C(Cl)C=C1)C=2C=CC=CC2Cl</chem>	95	9	Cali Set	TRUE
Molecule 120	36643-28-4	<chem>[Sn+](CCCC)(CCCC)CCCC</chem>	55.48	18	Cali Set	FALSE
Molecule 121	53-19-0	<chem>ClC1=CC=C(C=C1)C(C=2C=CC=CC2Cl)C(Cl)Cl</chem>	95	8	Cali Set	TRUE
Molecule 122	789-02-6	<chem>ClC1=CC=C(C=C1)C(C=2C=CC=CC2Cl)C(Cl)(Cl)Cl</chem>	95	9	Cali Set	TRUE
Molecule 123	10061-01-5	<chem>C(\CCl)=C\Cl</chem>	77.5	23	Cali Set	FALSE
Molecule 124	22967-92-6	<chem>[Hg+2][CH3-]</chem>	82.95	49	Cali Set	FALSE
Molecule 125	8065-48-3	<chem>O=P(OCC)(OCC)SCCSCC.S=P(OCC)(OCC)OCCSCC</chem>	90	8	Cali Set	FALSE
Molecule 126	56-38-2	<chem>O=N(=O)C1=CC=C(OP(=S)(OCC)OCC)C=C1</chem>	90	7	Cali Set	TRUE
Molecule 127	121-75-5	<chem>O=C(OCC)CC(SP(=S)(OC)OC)C(=O)OCC</chem>	88.06	8	Cali Set	TRUE
Molecule 128	86-50-0	<chem>O=C1C=2C=CC=CC2N=NN1CSP(=S)(OC)OC</chem>	85	8	Cali Set	TRUE
Molecule 129	14265-44-2	<chem>O=P([O-])([O-])[O-]</chem>	91.89	11	Cali Set	FALSE
Molecule 130	16984-48-8	<chem>[F-]</chem>	27.77	16	Cali Set	FALSE
Molecule 131	96-12-8	<chem>ClCC(Br)CBr</chem>	68.18	1	Cali Set	FALSE
Molecule 132	10061-02-6	<chem>C(\CCl)=C/Cl</chem>	0	15	Cali Set	FALSE
Molecule 133	67-64-1	<chem>O=C(C)C</chem>	78.52	1	Cali Set	FALSE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 134	2385-85-5	<chem>ClC1(Cl)C2(Cl)C3(Cl)C4(Cl)C(Cl)C(Cl)C5(Cl)C3(Cl)C1(Cl)C5(Cl)C24Cl</chem>	90	16	Cali Set	TRUE
Molecule 135	75-99-0	<chem>O=C(O)C(Cl)(Cl)C</chem>	0	9	Cali Set	TRUE
Molecule 136	14265-45-3	<chem>O=S([O-])[O-]</chem>	47.73	6	Cali Set	FALSE
Molecule 137	1071-83-6	<chem>O=C(O)CNCP(=O)(O)O</chem>	41.67	2	Cali Set	FALSE
Molecule 138	88-85-7	<chem>O=N(=O)C=1C=C(C(O)=C(C1)C(C)CC)N(=O)=O</chem>	0	1	Cali Set	FALSE
Molecule 139	3268-87-9	<chem>ClC=1C(Cl)=C(Cl)C=2OC=3C(Cl)=C(Cl)C(Cl)=C(Cl)C3OC2C1Cl</chem>	95.36	46	Cali Set	TRUE
Molecule 140	57117-41-6	<chem>ClC1=CC=2OC3=CC(Cl)=C(Cl)C(Cl)=C3C2C=C1Cl</chem>	0	3	Cali Set	FALSE
Molecule 141	57117-44-9	<chem>ClC=1C=C2OC=3C(Cl)=C(Cl)C(Cl)=CC3C2=C(Cl)C1Cl</chem>	0	3	Cali Set	FALSE
Molecule 142	57117-31-4	<chem>ClC=1C=C2OC=3C(Cl)=C(Cl)C(Cl)=CC3C2=CC1Cl</chem>	0	3	Cali Set	FALSE
Molecule 143	72918-21-9	<chem>ClC1=CC=2OC3=CC(Cl)=C(Cl)C(Cl)=C3C2C(Cl)=C1Cl</chem>	35	2	Cali Set	FALSE
Molecule 144	35822-46-9	<chem>ClC=1C=C2OC=3C(Cl)=C(Cl)C(Cl)=C(Cl)C3OC2=C(Cl)C1Cl</chem>	92.04	18	Cali Set	TRUE
Molecule 145	51207-31-9	<chem>ClC=1C=C2OC=3C=C(Cl)C(Cl)=CC3C2=CC1Cl</chem>	52.15	4	Cali Set	TRUE
Molecule 146	19408-74-3	<chem>ClC=1C=C2OC3=CC(Cl)=C(Cl)C(Cl)=C3OC2=C(Cl)C1Cl</chem>	35.48	2	Cali Set	FALSE
Molecule 147	60851-34-5	<chem>ClC=1C=C2C(OC=3C(Cl)=C(Cl)C(Cl)=CC32)=C(Cl)C1Cl</chem>	6.67	2	Cali Set	FALSE
Molecule 148	39001-02-0	<chem>ClC=1C(Cl)=C(Cl)C2=C(OC=3C(Cl)=C(Cl)C(Cl)=C(Cl)C32)C1Cl</chem>	84.09	17	Cali Set	TRUE
Molecule 149	70648-26-9	<chem>ClC1=CC=2OC3=C(Cl)C(Cl)=C(Cl)C(Cl)=C3C2C=C1Cl</chem>	0	7	Cali Set	TRUE
Molecule 150	57653-85-7	<chem>ClC=1C=C2OC3=C(Cl)C(Cl)=C(Cl)C=C3OC2=C(Cl)C1Cl</chem>	0	1	Cali Set	FALSE
Molecule 151	39227-28-6	<chem>ClC=1C=C2OC=3C(Cl)=C(Cl)C(Cl)=C(Cl)C3OC2=CC1Cl</chem>	3.33	2	Cali Set	FALSE
Molecule 152	67562-39-4	<chem>ClC=1C=C2C(OC=3C(Cl)=C(Cl)C(Cl)=C(Cl)C32)=C(Cl)C1Cl</chem>	73.68	17	Cali Set	TRUE
Molecule 153	55673-89-7	<chem>ClC1=CC=2OC3=C(Cl)C(Cl)=C(Cl)C(Cl)=C3C2C(Cl)=C1Cl</chem>	8	1	Cali Set	FALSE
Molecule 154	14488-53-0	<chem>[Sn+4]([CH2-]CCC)[CH2-]CCC</chem>	100	1	Cali Set	FALSE
Molecule 155	1868-53-7	<chem>FC(Br)Br</chem>	26.37	1	Cali Set	FALSE
Molecule 156	91-57-6	<chem>CC1=CC2=CC=CC=C2C=C1</chem>	100	11	Exp Set	TRUE
Molecule 157	260-94-6	<chem>C1=CC=C2C(=C1)C=C3C=CC=CC3=N2</chem>	100	4	Exp Set	TRUE
Molecule 158	205-82-3	<chem>C1=CC=C2C(=C1)C=CC3=C2C4=CC=CC5=C4C3=CC=C5</chem>	100	11	Exp Set	TRUE
Molecule 159	192-97-2	<chem>C1=CC=C2C(=C1)C3=CC=CC4=C3C5=C(C=CC=C25)C=C4</chem>	100	2	Exp Set	FALSE
Molecule 160	198-55-0	<chem>C1=CC2=C3C(=C1)C4=CC=CC5=C4C(=CC=C5)C3=CC=C2</chem>	100	2	Exp Set	FALSE
Molecule 161	91-22-5	<chem>C1=CC=C2C(=C1)C=CC=N2</chem>	100	6	Exp Set	TRUE
Molecule 162	122-39-4	<chem>C1=CC=C(C=C1)NC2=CC=CC=C2</chem>	87.91	22	Exp Set	TRUE
Molecule 163	375-22-4	<chem>C(=O)(C(C(C(F)(F)F)(F)F)(F)F)O</chem>	0	46	Exp Set	TRUE
Molecule 164	375-73-5	<chem>C(C(C(F)(F)S(=O)(=O)O)(F)F)(C(F)(F)F)(F)F</chem>	6.39	21	Exp Set	TRUE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 165	335-76-2	<chem>C(=O)(C(C(C(C(C(C(C(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)O</chem>	0	21	Exp Set	TRUE
Molecule 166	375-85-9	<chem>C(=O)(C(C(C(C(C(C(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)O</chem>	0	56	Exp Set	TRUE
Molecule 167	307-24-4	<chem>C(=O)(C(C(C(C(C(F)(F)(F)(F)(F)(F)(F)(F)(F)O</chem>	0	60	Exp Set	TRUE
Molecule 168	355-46-4	<chem>C(C(C(C(F)S(=O)(=O)O)(F)(F)(C(C(F)(F)(F)(F)F</chem>	0	39	Exp Set	TRUE
Molecule 169	375-95-1	<chem>C(=O)(C(C(C(C(C(C(C(F)(F)(F)(F)(F)(F)(F)(F)(F)(F)O</chem>	0	56	Exp Set	TRUE
Molecule 170	335-67-1	<chem>C(=O)(C(C(C(C(C(C(C(F)(F)(F)(F)(F)(F)(F)(F)(F)O</chem>	0	60	Exp Set	TRUE
Molecule 171	1763-23-1	<chem>C(C(C(C(C(F)S(=O)(=O)O)(F)(F)(F)(C(C(C(F)(F)(F)(F)F</chem>	0	46	Exp Set	TRUE
Molecule 172	2706-90-3	<chem>C(=O)(C(C(C(C(F)(F)(F)(F)(F)(F)O</chem>	0	47	Exp Set	TRUE
Molecule 173	99-76-3	<chem>COC(=O)C1=CC=C(C=C1)O</chem>	100	17	Exp Set	TRUE
Molecule 174	120-47-8	<chem>CCOC(=O)C1=CC=C(C=C1)O</chem>	100	21	Exp Set	TRUE
Molecule 175	17696-62-7	<chem>C1=CC=C(C=C1)OC(=O)C2=CC=C(C=C2)O</chem>	100	21	Exp Set	TRUE
Molecule 176	94-13-3	<chem>CCCOC(=O)C1=CC=C(C=C1)O</chem>	100	21	Exp Set	TRUE
Molecule 177	94-26-8	<chem>CCCCOC(=O)C1=CC=C(C=C1)O</chem>	100	21	Exp Set	TRUE
Molecule 178	74-31-7	<chem>C1=CC=C(C=C1)NC2=CC=C(C=C2)NC3=CC=CC=C3</chem>	100	9	Exp Set	TRUE
Molecule 179	15017-02-4	<chem>CC1=CC=CC=C1NC2=CC=C(C=C2)NC3=CC=CC=C3C</chem>	100	9	Exp Set	TRUE
Molecule 180	84-69-5	<chem>CC(C)COC(=O)C1=CC=CC=C1C(=O)OCC(C)C</chem>	75	1	Lit Set	FALSE
Molecule 181	131-18-0	<chem>CCCCCOC(=O)c1ccccc1C(=O)OCCCCC</chem>	88.57	1	Lit Set	FALSE
Molecule 182	15307-86-5	<chem>OC(=O)Cc1ccccc1Nc2c(Cl)cccc2Cl</chem>	67.5	6	Lit Set	TRUE
Molecule 183	882-09-7	<chem>CC(C)(Oc1ccc(Cl)cc1)C(O)=O</chem>	52.92	2	Lit Set	FALSE
Molecule 184	21145-77-7	<chem>CC1CC(C2=C(C1(C)C)C=C(C(=C2)C(=O)C)C)C</chem>	77.18	2	Lit Set	FALSE
Molecule 185	3380-34-5	<chem>C1=CC(=C(C=C1Cl)O)OC2=C(C=C(C=C2)Cl)Cl</chem>	99	32	Lit Set	TRUE
Molecule 186	22071-15-4	<chem>CC(C1=CC(=CC=C1)C(=O)C2=CC=CC=C2)C(=O)O</chem>	84.62	5	Lit Set	TRUE
Molecule 187	56980-93-9	<chem>CCN(CC)C(=O)NC1=CC(=C(C=C1)OCC(CNC(C)(C)C)O)C(=O)C</chem>	36.36	1	Lit Set	FALSE
Molecule 188	81103-11-9	<chem>CCC1C(C(C(=O)C(C(C(C(C(C(=O)O1)C)OC2CC(C(C(O2)C)O)(C)OC)C)OC3C(C(C(O3)C)N(C)C)O)(C)OC)C)O)(C)O</chem>	0	13	Lit Set	TRUE
Molecule 189	1222-05-5	<chem>CC1COCC2=CC3=C(C=C12)C(C(C3(C)C)C)C</chem>	43.85	3	Lit Set	FALSE
Molecule 190	57-63-6	<chem>CC12CCC3C(C1CCC2(C#C)O)CCC4=C3C=CC(=C4)O</chem>	66.67	3	Lit Set	FALSE
Molecule 191	50-28-2	<chem>CC12CCC3C(C1CCC2O)CCC4=C3C=CC(=C4)O</chem>	83.25	3	Lit Set	FALSE
Molecule 192	50-50-0	<chem>CC12CCC3C(C1CCC2O)CCC4=C3C=CC(=C4)OC(=O)C5=CC=CC=C5</chem>	0	1	Lit Set	FALSE
Molecule 193	41318-75-6	<chem>C1=CC(=CC=C1OC2=C(C=C(C=C2)Br)Br)Br</chem>	69.97	29	Lit Set	TRUE
Molecule 194	86-13-5	<chem>CN1C2CCC1CC(C2)OC(C3=CC=CC=C3)C4=CC=CC=C4</chem>	75.14	5	Lit Set	TRUE
Molecule 195	469-62-5	<chem>CCC(=O)OC(CC1=CC=CC=C1)(C2=CC=CC=C2)C(C)CN(C)C</chem>	25.18	1	Lit Set	FALSE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 196	207122-16-5	<chem>C1=CC(=C(C=C1OC2=C(C(=C(C(=C2Br)Br)Br)Br)Br)Br)Br</chem>	91.86	37	Lit Set	TRUE
Molecule 197	60166-93-0	<chem>CC(C(=O)NC1=C(C(=C(C(=C1)C(=O)NC(CO)CO)I)C(=O)NC(CO)CO)I)O</chem>	17.39	1	Lit Set	FALSE
Molecule 198	128-37-0	<chem>CC1=CC(=C(C(=C1)C(C)(C)C)O)C(C)(C)C</chem>	75.89	1	Lit Set	FALSE
Molecule 199	13674-84-5	<chem>CC(CI)COP(=O)(OCC(CI)C)OCC(CI)C</chem>	19	1	Lit Set	FALSE
Molecule 200	117-96-4	<chem>CC(=O)NC1=C(C(=C(C(=C1)C(=O)[O-])I)NC(=O)C)I</chem>	0	1	Lit Set	FALSE
Molecule 201	41859-67-0	<chem>CC(C)(C(=O)O)OC1=CC=C(C(=C1)CCNC(=O)C2=CC=C(C(=C2)Cl</chem>	53.28	2	Lit Set	FALSE
Molecule 202	6533-00-2	<chem>CCC12CCC3C(C1CCC2(C#C)O)CCC4=CC(=O)CCC34</chem>	0	1	Lit Set	FALSE
Molecule 203	52-53-9	<chem>CC(C)C(CCCN(C)CCC1=CC(=C(C(=C1)OC)OC)(C#N)C2=CC=C(C(=C2)OC)OC</chem>	11.23	6	Lit Set	TRUE
Molecule 204	207122-15-4	<chem>C1=C(C(=CC(=C1Br)Br)Br)OC2=CC(=C(C(=C2Br)Br)Br</chem>	90.18	37	Lit Set	TRUE
Molecule 205	66108-95-0	<chem>CC(=O)N(CC(CO)O)C1=C(C(=C(C(=C1)C(=O)NCC(CO)O)I)C(=O)NCC(CO)O)I</chem>	88.89	1	Lit Set	FALSE
Molecule 206	53-16-7	<chem>CC12CCC3C(C1CCC2=O)CCC4=C3C=CC(=C4)O</chem>	34.85	4	Lit Set	TRUE
Molecule 207	78649-41-9	<chem>CN(C1=C(C(=C(C(=C1)C(=O)NCC(CO)O)I)C(=O)NCC(CO)O)I)C(=O)CO</chem>	89	1	Lit Set	FALSE
Molecule 208	79617-96-2	<chem>CNC1CCC(C2=CC=CC=C12)C3=CC(=C(C(=C3)Cl)Cl</chem>	2.56	6	Lit Set	TRUE
Molecule 209	859-18-7	<chem>CCCC1CC(N(C1)C)C(=O)NC(C2C(C(C(C(O2)SC)O)O)O)C(C)O</chem>	5.35	8	Lit Set	TRUE
Molecule 210	54910-89-3	<chem>CNCCC(C1=CC=CC=C1)OC2=CC=C(C(=C2)C(F)(F)F</chem>	24.06	10	Lit Set	TRUE
Molecule 211	114-07-8	<chem>CCC1C(C(C(C(=O)C(CC(C(C(C(C(=O)O1)C)OC2CC(C(C(O2)C)O)(C)OC)C)OC3C(C(C(O3)C)N(C)C)O)(C)O)C)O)(C)O</chem>	0	12	Lit Set	TRUE
Molecule 212	67035-22-7	<chem>CC1=C(C(=C(C(=N1)C)C(=O)OC)C2=CC=CC=C2[N+](=O)[O-])C(=O)OC</chem>	32.49	12	Lit Set	TRUE
Molecule 213	396-01-0	<chem>C1=CC=C(C(=C1)C2=NC3=C(N=C(N(=C3N(=C2N)N)N</chem>	2.29	9	Lit Set	TRUE
Molecule 214	10238-21-8	<chem>COC1=C(C(=C(C(=C1)Cl)C(=O)NCCC2=CC=C(C(=C2)S(=O)(=O)NC(=O)NC3CCCCC3</chem>	74.46	1	Lit Set	FALSE
Molecule 215	73334-07-3	<chem>CN(CC(CO)O)C(=O)C1=C(C(=C(C(=C1)C(=O)NCC(CO)O)I)NC(=O)COC)I</chem>	83.33	1	Lit Set	FALSE
Molecule 216	58-22-0	<chem>CC12CCC3C(C1CCC2O)CCC4=CC(=O)CCC34C</chem>	100	3	Lit Set	FALSE
Molecule 217	26538-44-3	<chem>CC1CCCC(CCCCCC2=C(C(=CC(=C2)O)O)C(=O)O1)O</chem>	100	1	Lit Set	FALSE
Molecule 218	57-68-1	<chem>CC1=CC(=NC(=N1)NS(=O)(=O)C2=CC=C(C(=C2)N)C</chem>	52	1	Lit Set	FALSE
Molecule 219	68-22-4	<chem>CC12CCC3C(C1CCC2(C#C)O)CCC4=CC(=O)CCC34</chem>	100	3	Lit Set	FALSE
Molecule 220	57-10-3	<chem>CCCCCCCCCCCCCCCC(=O)O</chem>	98.02	1	Lit Set	FALSE
Molecule 221	9002-93-1	<chem>CCCCCCCCC1=CC=C(C(=C1)O</chem>	100	4	Lit Set	TRUE
Molecule 222	57-11-4	<chem>CCCCCCCCCCCCCCCC(=O)O</chem>	98.05	1	Lit Set	FALSE
Molecule 223	53-86-1	<chem>CC1=C(C2=C(N1C(=O)C3=CC=C(C(=C3)Cl)C=CC(=C2)OC)CC(=O)O</chem>	11.9	1	Lit Set	FALSE
Molecule 224	88150-42-9	<chem>CCOC(=O)C1=C(NC(=C(C1C2=CC=CC=C2Cl)C(=O)OC)C)COCCN</chem>	73.03	6	Lit Set	TRUE
Molecule 225	50-36-2	<chem>CN1C2CCC1C(C(C2)OC(=O)C3=CC=CC=C3)C(=O)OC</chem>	91.76	6	Lit Set	TRUE
Molecule 226	511-15-9	<chem>CC(C)C1=C(C(=CC2=C1CCC3C2(CCCC3(C)C)C)O</chem>	85.19	1	Lit Set	FALSE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 227	300-62-9	<chem>CC(CC1=CC=CC=C1)N</chem>	95.9	9	Lit Set	TRUE
Molecule 228	443-48-1	<chem>CC1=NC=C(N1CCO)[N+](=O)[O-]</chem>	38.89	1	Lit Set	FALSE
Molecule 229	644-62-2	<chem>CC1=C(C(=C(C=C1)Cl)NC2=CC=CC=C2C(=O)O)Cl</chem>	5.79	3	Lit Set	FALSE
Molecule 230	61-68-7	<chem>CC1=C(C(=CC=C1)NC2=CC=CC=C2C(=O)O)C</chem>	2.13	1	Lit Set	FALSE
Molecule 231	1175-06-0	<chem>CC(C)CCCC(C)C1CCC2C1(CCC3C2CC(=O)C4C3(CCC(C4)O)C)C</chem>	100	3	Lit Set	FALSE
Molecule 232	189084-62-6	<chem>C1=CC=C(C=C1)OC2=C(C(=C(C=C2Br)Br)Br)Br</chem>	71.87	11	Lit Set	TRUE
Molecule 233	519-09-5	<chem>CN1C2CCC1C(C(C2)OC(=O)C3=CC=CC=C3)C(=O)O</chem>	81.5	6	Lit Set	TRUE
Molecule 234	243982-82-3	<chem>C1=CC(=C(C=C1Br)Br)OC2=C(C=C(C(=C2)Br)O)Br</chem>	80.45	36	Lit Set	TRUE
Molecule 235	134-62-3	<chem>CCN(CC)C(=O)C1=CC=CC(=C1)C</chem>	91.55	6	Lit Set	TRUE
Molecule 236	73590-58-6	<chem>CC1=CN=C(C(=C1OC)C)CS(=O)C2=NC3=C(N2)C=C(C=C3)OC</chem>	8.49	1	Lit Set	FALSE
Molecule 237	37853-59-1	<chem>C1=C(C=C(C(=C1Br)OCCOC2=C(C(=C(C=C2Br)Br)Br)Br)Br)Br</chem>	94.58	35	Lit Set	TRUE
Molecule 238	131-57-7	<chem>COC1=CC(=C(C=C1)C(=O)C2=CC=CC=C2)O</chem>	78.12	1	Lit Set	FALSE
Molecule 239	83-48-7	<chem>CCC(C=CC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C(C)C</chem>	66	3	Lit Set	FALSE
Molecule 240	74103-06-3	<chem>C1CN2C(=CC=C2C(=O)C3=CC=CC=C3)C1C(=O)O</chem>	43.98	1	Lit Set	FALSE
Molecule 241	182346-21-0	<chem>C1=CC(=C(C=C1Br)Br)OC2=C(C(=C(C=C2)Br)Br)Br</chem>	90.76	36	Lit Set	TRUE
Molecule 242	474-62-4	<chem>CC(C)C(C)CCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C</chem>	80.84	3	Lit Set	FALSE
Molecule 243	57-87-4	<chem>CC(C)C(C)C=CC(C)C1CCC2C1(CCC3C2=CC=C4C3(CCC(C4)O)C)C</chem>	71.64	3	Lit Set	FALSE
Molecule 244	519-98-2	<chem>CC1=C(C(=O)N(N1C)C2=CC=CC=C2)NC</chem>	66.93	1	Lit Set	FALSE
Molecule 245	566-28-9	<chem>CC(C)CCCC(C)C1CCC2C1(CCC3C2C(=O)C=C4C3(CCC(C4)O)C)C</chem>	98.22	3	Lit Set	FALSE
Molecule 246	137862-53-4	<chem>CCCCC(=O)N(CC1=CC=C(C=C1)C2=CC=CC=C2C3=NNN=N3)C(C(C)C)C(=O)O</chem>	98.18	6	Lit Set	TRUE
Molecule 247	83-07-8	<chem>CC1=C(C(=O)N(N1C)C2=CC=CC=C2)N</chem>	55.44	1	Lit Set	FALSE
Molecule 248	360-68-9	<chem>CC(C)CCCC(C)C1CCC2C1(CCC3C2CCC4C3(CCC(C4)O)C)C</chem>	84.02	3	Lit Set	FALSE
Molecule 249	601-53-6	<chem>CC(C)CCCC(C)C1CCC2C1(CCC3C2CCC4C3(CCC(=O)C4)C)C</chem>	90.6	3	Lit Set	FALSE
Molecule 250	313-04-2	<chem>CC(CCC=C(C)C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C</chem>	92.52	3	Lit Set	FALSE
Molecule 251	54-11-5	<chem>CN1CCCC1C2=CN=CC=C2</chem>	98.15	1	Lit Set	FALSE
Molecule 252	83-46-5	<chem>CCC(CCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C(C)C</chem>	81.54	3	Lit Set	FALSE
Molecule 253	83-15-8	<chem>CC1=C(C(=O)N(N1C)C2=CC=CC=C2)NC(=O)C</chem>	46.13	1	Lit Set	FALSE
Molecule 254	17605-67-3	<chem>CC=C(CCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C(C)C</chem>	78.8	3	Lit Set	FALSE
Molecule 255	611-59-6	<chem>CN1C=NC2=C1C(=O)N(C(=O)N2)C</chem>	100	13	Lit Set	TRUE
Molecule 256	58-55-9	<chem>CN1C2=C(C(=O)N(C1=O)C)NC=N2</chem>	100	6	Lit Set	TRUE
Molecule 257	1163-19-5	<chem>C1(=C(C(=C(C(=C1Br)Br)Br)Br)Br)OC2=C(C(=C(C(=C2Br)Br)Br)Br)Br</chem>	94.14	36	Lit Set	TRUE

Table S1. List of all substances in the experimental set, including the CAS RN, SMILES, median percent removal value used in fittings, the number of measurements, the source of those measurements, and whether or not those substances were used in the training of the CAIRRE model suite.

Molecule 258	1672-58-8	<chem>CC1=C(C(=O)N(N1C)C2=CC=CC=C2)NC=O</chem>	68.18	1	Lit Set	FALSE
Molecule 259	57-88-5	<chem>CC(C)CCCC(C)C1CCC2C1(CCC3C2CC=C4C3(CCC(C4)O)C)C</chem>	85.42	3	Lit Set	FALSE
Molecule 260	103-90-2	<chem>CC(=O)NC1=CC=C(C=C1)O</chem>	100	13	Lit Set	TRUE