Supporting Information

Revisiting Mayr's Reactivity Database: Expansion, Sensitivity Analysis, and Uncertainty Quantification

Moritz K.-E. Wolff[1], Armin R. Ofial[2], Jonny Proppe[1,*]

[1]TU Braunschweig, Germany, j.proppe@tu-braunschweig.de
[2]Ludwig-Maximilians-Universität München, Germany

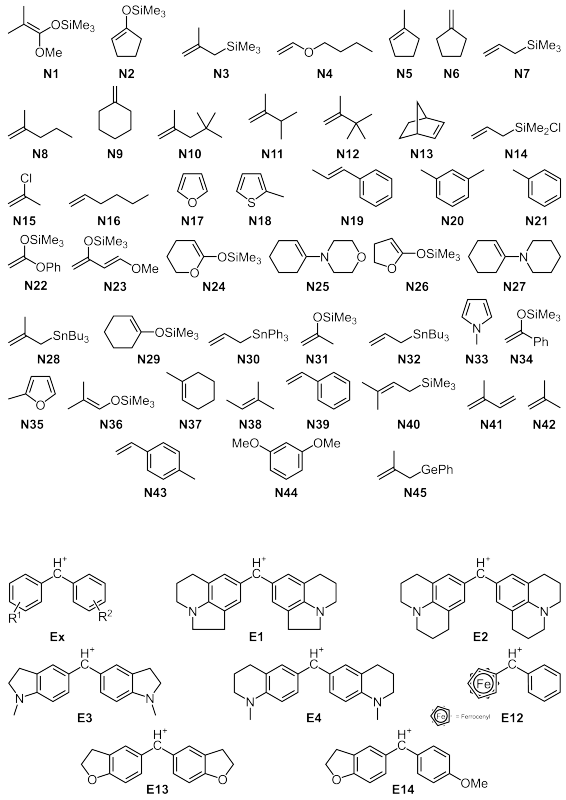

Figure S1: Nucleophiles and electrophiles of the established (i.e., non-expanded) reference set. IDs were adopted from the study by Proppe and Kircher.[12]

S1 Reduced Set for Bayesian Bootstrapping

To reduce the computational cost of the Bayesian bootstrapping technique utilised in the analysis of the "least connective reaction" algorithm, a smaller, well connected subset was chosen including **N1**, **N2**, **N22**–**N26**, **N28**, as well as **E1**–**E10**. **N24** and **E8** were chosen as the anchor species, with the fixed parameters assigned to the values for the optimised parameters of the full expanded reference set. This ensures that the optimised parameters for the small set are generally comparable to those of the full reference set.

S2 Least Connected Species Algorithm

Similarly to the "least connective reaction" algorithm, the "least connected species" algorithm was employed to determine a sequence of 67 species that can be removed from the reference set incrementally. At first, it narrows the pool down to species that do not lead to other species being removed by the 3N2E rule, then the species with the lowest number of valid reactions and finally to species without reactions involving the anchor species. As the removal of species with less valid reactions and low effects on other species are prioritised, the resulting subsets should have higher connectivity between species. A shift in optimised parameters was observed, e.g., in Fig. S2. The model

Table S1: Substituents for benzhydrylium ions **Ex** in Fig. S1.

	R^1	R^2		R^1	R^2
E5	4-(*N*-pyrrolidino)	R^1	**E21**	4-Me	H
E6	4-$N(Me)_2$	R^1	**E22**	4-F	R^1
E7	4-N(Me)(Ph)	R^1	**E23**	4-F	H
E8	4-(*N*-morpholino)	R^1	**E24**	3-F, 4-Me	R^1
E9	4-$N(Ph)_2$	R^1	**E25**	H	R^1
E10	4-N(Me)(CH_2CF_3)	R^1	**E26**	4-Cl	R^1
E11	4-N(Ph)(CH_2CF_3)	R^1	**E27**	3-F	H
E15	4-MeO	R^1	**E28**	4-(CF_3)	H
E16	4-Meo	4-PhO	**E29**	3,5-F_2	H
E17	4-Meo	4-Me	**E30**	3-F	R^1
E18	4-MeO	H	**E31**	3,5-F_2	3-F
E19	4-PhO	H	**E32**	4-(CF_3)	R^1
E20	4-Me	R^1	**E33**	3,5-F_2	R^1

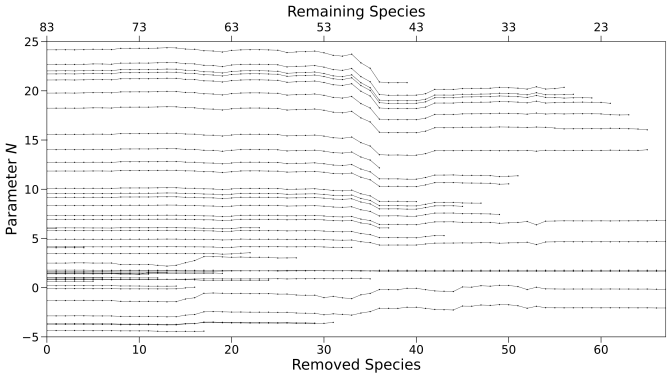

Figure S2: *N* parameters for subsets created according to the "least connected species" algorithm. Each line corresponds to one nucleophile and ends when the corresponding species was removed.

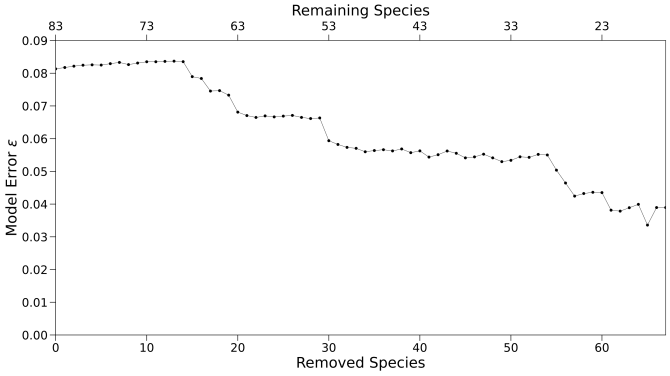

Figure S3: Model error ε for subsets created by removing species with low connectivity ("least connected species" algorithm).

error shown in Fig. S3 shows a general trend towards lower values, further indicating that a well-connected network of species has a positive impact on the model error. The parameter shift may have been the result of a local optimisation minimum with noticeably different parameters but similar model error becoming the new global minimum with only a small change in the data set.

S3 Parameters and Full Comparison

Table S2: Parameters calculated for the expanded reference set. Also shown are the corresponding original parameters[13] denoted with "DB". Omitted are species left out of calculations due to data selection criteria, as well as anchor electrophile **E15**. The s_N parameters of anchor nucleophile **N7** are denoted in italics to reflect their fixed status.

Nucleophiles	s_N	$s_{N,DB}$	N	N_{DB}	Electrophiles	E	E_{DB}
N1	0.85	0.98	10.09	9.00	**E1**	−11.20	−10.04
N2	0.84	0.93	7.33	6.57	**E2**	−10.56	−9.45
N3	0.87	0.96	4.92	4.41	**E3**	−9.81	−8.76
N4	0.87	0.91	4.15	3.76	**E4**	−9.17	−8.22
N5	0.98	1.17	2.50	1.18	**E5**	−8.53	−7.69
N7	*1.00*	*1.00*	1.78	1.68	**E6**	−7.82	−7.02
N8	1.07	1.06	0.86	0.84	**E7**	−6.57	−5.89
N9	1.02	1.04	1.40	1.16	**E8**	−6.16	−5.53
N10	1.05	1.07	1.02	0.79	**E9**	−5.25	−4.72
N11	0.91	1.00	1.47	0.65	**E10**	−4.30	−3.85
N12	1.07	1.07	0.20	0.06	**E11**	−3.48	−3.14
N13	1.09	1.09	−0.11	−0.25	**E12**	−2.96	−2.64
N14	1.25	1.06	−1.31	−0.57	**E13**	−1.50	−1.36
N15	2.13	1.97	−3.72	−3.65	**E14**	−0.87	−0.81
N16	1.54	1.41	−2.87	−2.77	**E16**	0.55	0.61
N17	1.15	1.29	1.48	1.33	**E17**	1.41	1.48
N18	0.88	0.99	1.50	1.35	**E18**	1.94	2.11
N20	2.28	2.08	−3.66	−3.57	**E19**	2.81	2.90
N21	1.88	1.77	−4.36	−4.36	**E20**	3.73	3.63
N22	0.73	0.81	9.18	8.23	**E21**	4.42	4.43
N23	0.75	0.84	9.56	8.57	**E22**	4.96	5.01
N24	0.77	0.86	11.85	10.61	**E23**	5.17	5.20
N25	0.74	0.83	12.73	11.40	**E24**	5.18	5.24
N26	0.63	0.70	14.02	12.56	**E25**	5.40	5.47
N27	0.71	0.81	15.01	13.36	**E26**	5.41	5.48
N28	0.79	0.89	8.35	7.48	**E27**	6.13	6.23
N29	0.89	1.00	5.80	5.21	**E28**	6.61	6.70
N30	0.82	0.90	3.46	3.09	**E29**	6.64	6.74
N31	0.82	0.91	6.04	5.41	**E30**	6.75	6.87
N32	0.82	0.89	6.07	5.46	**E31**	7.31	7.52
N34	0.85	0.96	6.93	6.22	**E32**	7.60	7.96
N35	0.99	1.11	4.05	3.61	**E34**	−20.06	−17.90
N38	1.11	1.17	0.81	0.65	**E35**	−19.39	−17.29
N40	1.38	1.17	0.66	0.90	**E36**	−18.05	−16.11
N42	1.09	0.98	0.98	1.11	**E37**	−17.74	−15.83
N43	1.11	1.06	1.63	1.70	**E38**	−14.98	−13.39
N46	0.55	0.62	24.16	21.54	**E39**	−13.63	−12.18
N47	0.58	0.65	22.68	20.22	**E40**	−9.63	−8.63
N48	0.59	0.67	22.02	19.62			
N49	0.60	0.67	21.71	19.36			
N50	0.61	0.69	21.10	18.82			
N51	0.65	0.73	19.77	17.64			
N52	0.68	0.77	18.24	16.27			
N53	0.76	0.86	15.57	13.91			

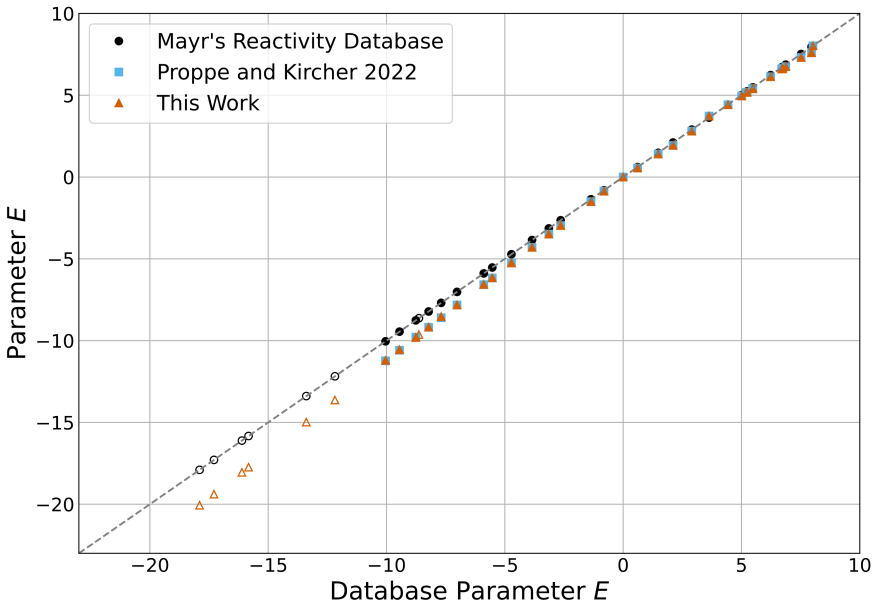

Figure S4: E parameters reported in Mayr's Reactivity Database,[13] Proppe and Kircher,[12] and this work. Hollow markers indicate quinone methides.

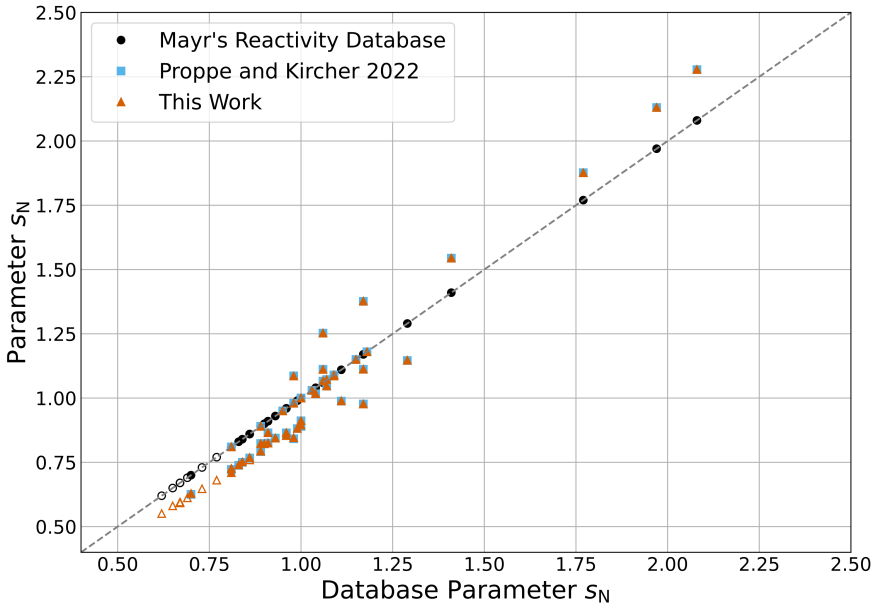

Figure S5: s_N parameters reported in Mayr's Reactivity Database,[13] Proppe and Kircher,[12] and this work. Hollow markers indicate carbanions.

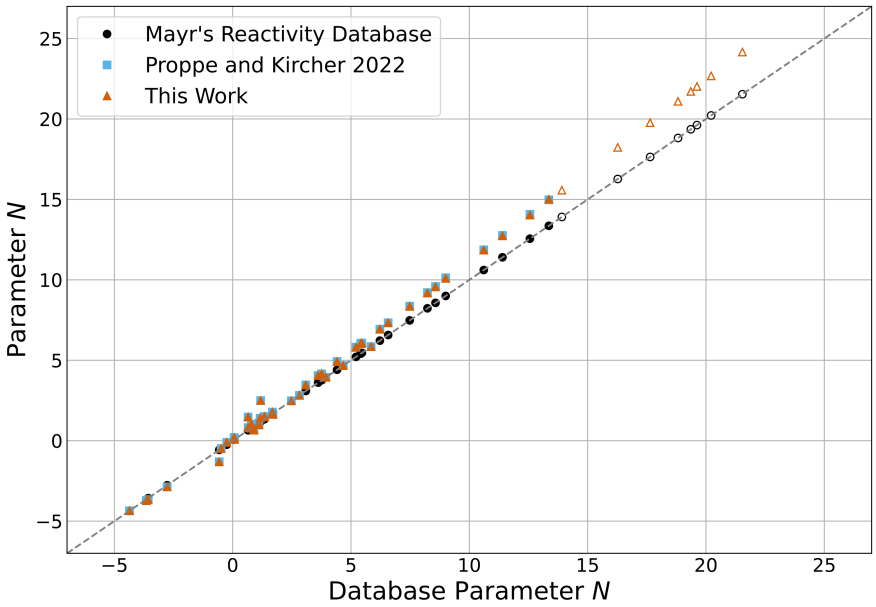

Figure S6: N parameters reported in Mayr's Reactivity Database,[13] Proppe and Kircher,[12] and this work. Hollow markers indicate carbanions.