Electronic Supplementary Information for Unveiling CO_2 reactivity with data-driven methods

Maike Eckhoff, Kerstin L. Bublitz, and Jonny Proppe*

TU Braunschweig

Institute of Physical and Theoretical Chemistry

Gauss Str 17, 38106 Braunschweig, Germany

E-mail: j.proppe@tu-braunschweig.de

More supplementary information can be accessed through the project-related GitLab repository: `https://git.rz.tu-bs.de/proppe-group/co2_electrophilicity`.

Section S1: Basis set test

For a preliminary study on an appropriate basis set size, four different settings were tested: All structures (nucleophiles, CO_2, CO_2–nucleophile transition states) were optimised using B3LYP-D3(BJ)[1–4] in conjunction with either the def2-SVPD basis set or the def2-TZVPD basis set.[5–7] The basis sets have been generated with the Basis Set Exchange program[8] and added manually for the Gaussian (version 16, revision C01)[9] calculations. Single-point calculations introducing implicit solvation with DMSO as solvent were then carried out on both optimised structures using both basis set combinations. In Table S1, abbreviations are introduced for the different computational settings applied for this test.

The results (Fig. S1) suggest that both basis sets provide comparable structures, but differ in energy. Therefore, we decided to perform efficient structure optimisations with the smaller def2-SVPD basis set and to calculate the molar Gibbs free energies of the optimised structures with the more comprehensive def2-TZVPD basis set.

Table S1: Computational settings selected for free-energy calculations and their abbreviations used in Fig. S1. For all calculations, B3LYP-D3(BJ) is applied as exchange–correlation functional.

Abbreviation	Computational setting
B3LYP_1a	Structure: def2-SVPD, Energy: def2-SVPD/SMD(DMSO)
B3LYP_1b	Structure: def2-TZVPD, Energy: def2-SVPD/SMD(DMSO)
B3LYP_2a	Structure: def2-SVPD, Energy: def2-TZVPD/SMD(DMSO)
B3LYP_2b	Structure: def2-TZVPD, Energy: def2-TZVPD/SMD(DMSO)

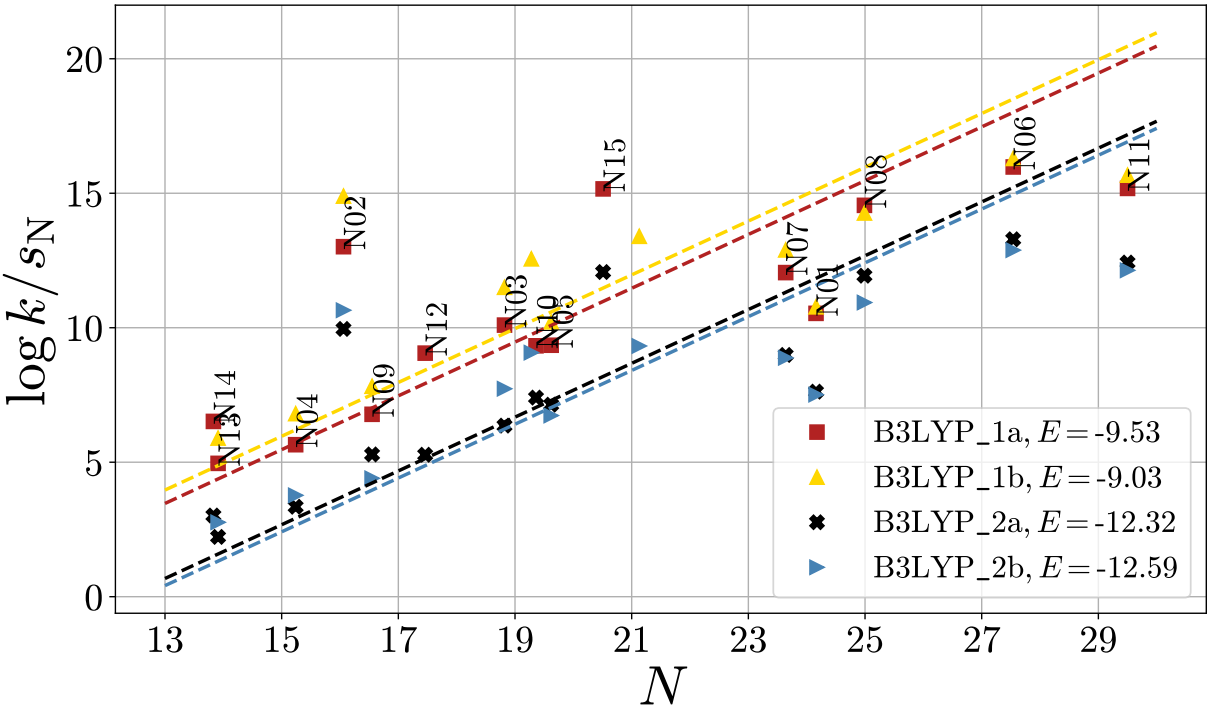

Figure S1: Calibration of $E(CO_2)$ against B3LYP results based on different basis sets (cf. Table S1).

Section S2: Isolated reactants versus pre-reaction complexes

Because van-der-Waals complexes can be more stable than the corresponding isolated reactants, which would increase the free energy of activation and decrease the bimolecular rate constant, we investigated this issue. For a subset of five carbanions (**N01**, **N04**, **N11**, **N13**,

N15) we calculated the B3LYP molar Gibbs free energy of their pre-reaction complexes formed with CO_2 (Table S2). We find that the isolated reactants are more stable (by 3–12 kcal/mol) than their pre-reaction complexes except for nucleophile **N15**. However, the difference is close to zero, leading us to the conclusion that the description of isolated reactants is sufficient for the calculation of activation energies.

Table S2: Difference in the molar Gibbs free energy in solution (kcal mol^{-1}) between isolated reactants and their corresponding van-der-Waals complexes for five systems. $\Delta G_{sol} = G_{sol}(CO_2) + G_{sol}(\text{carbanion}) - G_{sol}(\text{van-der-Waals complex})$.

Nucleophile	ΔG_{sol}
N01	−2.13
N04	−9.24
N11	−4.52
N13	−12.17
N15	0.16

Section S3: Information on conformer ensembles

We performed conformer sampling with CREST (version 2.12) in conjunction with GFN2-xTB.[10,11] All conformers were subjected to DFT calculations as described in the main text to obtain Gibbs free energies in solution (G_{sol}). The rank of the lowest-energy GFN2-xTB conformer with respect to the G_{sol} ranking (#LowE) along with the conformer ensemble size are list in Tables S3 and S4 for reactants and transition states of several heteroallene–carbanion reactions. XYZ coordinates of the conformer ensembles and of the lowest-energy conformers can be accessed through the project-related GitLab repository.[12]

Section S4: Additional information on the ARD regression ML model

The selection of the final parameters for the multivariate linear (ML) model (see main text Section 3.2) was conducted using the following procedure: First, all individual contributions to the Gibbs free activation energy (summarised in Table S8) were taken into account to approximate the experimental rate constant. After analysing the results and reviewing the values associated with each contribution, all components related to electronic energies were removed, as these values for CO_2 differ significantly from those of the other heteroallenes.

Table S3: Overview of conformer ensemble sizes and ranks of lowest-energy GFN2-xTB conformers with respect to their corresponding DFT energy rankings (#LowE) for heteroallene–carbanion reactions.[13]

	$\sum$	#LowE		$\sum$	#LowE
E1	1	1	**E1-N16**	9	3
E2	1	1	**E1-N17**	4	2
E3	1	1	**E1-N18**	14	2
E4	5	3	**E1-N19**	1	1
CO_2	1	1	**E2-N03**	43	40
N01	1	1	**E2-N10**	6	4
N03	13	7	**E2-N17**	36	14
N07	4	2	**E2-N21**	14	2
N10	1	1	**E2-N24**	11	7
N16	4	1	**E3-N03**	115	32
N17	2	2	**E3-N10**	9	9
N18	4	4	**E3-N20**	235	218
N19	1	1	**E3-N21**	14	2
N20	31	11	**E3-N22**	35	9
N21	6	2	**E4-N03**	151	16
N22	2	1	**E4-N07**	285	left out
N23	1	1	**E4-N17**	58	17
N24	1	1	**E4-N20**	499	left out
E1-N01	1	1	**E4-N23**	30	8
E1-N07	8	3	**CO_2-N01**	2	2

Next, various molecular descriptors from conceptual DFT[14–20] were tested. As these descriptors are calculated from the frontier molecular orbitals of the electrophile, the correlation between different molecular descriptors and Mayr's electrophilicity parameter was assessed (summarised in Table S6). All linearly independent terms of molecular conceptual DFT descriptors applied in a former study[15] were tested. The most promising descriptor, the electronic chemical potential μ, was added to the remaining contributions from the Gibbs free activation energy. Last, to describe the nucleophiles, Mayr's parameters N and s_N were included. With these preselected parameters in hand, the ML models were trained using the leave-one-out approach. As the parameters are standardised, the magnitude of the optimised parameters can be analysed to determine their importance. Based on this analysis, as well as performance metrics (RSME and max. AE), all contributions from the Gibbs free activation energy were eliminated step by step, except for the temperature-scaled overall entropy. Mayr's sensitivity parameter s_N was removed as well, leaving three remaining terms, one from each parameter group.

In Table S7, $\log k_\mathrm{QC}$, $\log k_\mathrm{ML}$, and $\log k_\mathrm{exp}$ values are provided for several heteroallene–

Table S4: Overview of conformer ensemble sizes and ranks of lowest-energy GFN2-xTB conformers with respect to their corresponding DFT energy rankings (#LowE) for reactions with CO_2 in this study.

	$\sum$	#LowE		$\sum$	#LowE
CO_2	1	1	**CO_2-N01**	2	2
N01	1	1	**CO_2-N02**	3	3
N02	2	2	**CO_2-N03**	9	1
N03	13	7	**CO_2-N04**	4	2
N04	4	2	**CO_2-N05**	7	2
N05	6	2	**CO_2-N06**	5	3
N06	5	2	**CO_2-N07**	10	5
N07	4	2	**CO_2-N08**	3	1
N08	3	1	**CO_2-N09**	5	2
N09	1	1	**CO_2-N10**	1	1
N10	1	1	**CO_2-N11**	4	4
N11	3	1	**CO_2-N12**	3	1
N12	3	1	**CO_2-N13**	3	3
N13	2	2	**CO_2-N14**	27	19
N14	52	20	**CO_2-N15**	39	5
N15	7	7			

carbanion reactions. Figs. S2 to S21 show $\log k_{QC}$ (unfilled circles) and $\log k_{ML}$ (black circles) values versus their experimental analogues for the heteroallene–carbanion reactions listed in Table 3 of the main text. The reaction shown in red has been left out for training and is predicted by the respective leave-one-out model. The final ML model is shown in Fig. S22. Detailed information about the regression method is given in Sections 2.2.1 and 3.2 of the main text.

Table S5: Dictionary to translate IDs of molecules used in this work to IDs used in previous work.[13]

elec.	nuc.	ID[13]
E1	**N01**	2j
	N07	2f
	N16	2g
	N17	2a
	N18	2b
	N19	2i
E2	**N03**	2o
	N10	2n
	N17	2a
	N21	2p
	N24	2l
E3	**N03**	2o
	N10	2n
	N20	2m
	N21	2p
	N22	2c
E4	**N03**	2o
	N17	2a
	N23	2h
CO_2	**N01**	2j

Table S6: Correlations (measured as Pearson's r^2) between molecular conceptual DFT descriptors,[14,16–20] which are based on $\varepsilon_{\mathrm{LUMO}}$ and $\varepsilon_{\mathrm{HOMO}}$ energies of the electrophile, and Mayr's electrophilicity parameter for the five heteroallenes ($E(\mathrm{CO_2}) = 16.3$, from experiment[13]).

molecular descriptor	r^2
$\mu^{+}_{\mathrm{FMO}} = \varepsilon_{\mathrm{LUMO}}$[16]	0.1013
$\mu^{-}_{\mathrm{FMO}} = \varepsilon_{\mathrm{HOMO}}$[16]	0.3551
$\varepsilon^2_{\mathrm{LUMO}}$	0.2379
$\varepsilon^2_{\mathrm{HOMO}}$	0.3231
$\varepsilon_{\mathrm{LUMO}} \cdot \varepsilon_{\mathrm{HOMO}}$	0.1669
$(\varepsilon_{\mathrm{L}} - \varepsilon_{\mathrm{H}})^{-1}$	0.0027
$\mu_{\mathrm{FMO}} = \frac{\varepsilon_{\mathrm{LUMO}}+\varepsilon_{\mathrm{HOMO}}}{2}$[17,18]	0.8358
$\eta_{\mathrm{FMO}} = \varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}}$[19]	0.0725
$\omega_{\mathrm{FMO}} = \frac{\mu^2_{\mathrm{FMO}}}{2\eta_{\mathrm{FMO}}}$[20]	0.3751

Table S7: Logarithmic rate constants obtained from quantum chemical calculations ($\log k_{\mathrm{QC}}$) and our ML model ($\log k_{\mathrm{ML}}$) compared to experimental reference values[13] ($\log k_{\mathrm{exp}}$) for several heteroallene–carbanion reactions.

elec.	nuc.	$\log k_{\mathrm{QC}}$	$\log k_{\mathrm{ML}}$	$\log k_{\mathrm{exp}}$[13]
E1	**N01**	5.14	4.41	4.99
	N07	5.09	3.30	3.25
	N16	3.63	1.55	1.34
	N17	3.24	1.06	1.00
	N18	2.04	−0.05	−0.38
	N19	1.98	1.88	1.42
E2	**N03**	−0.16	0.06	0.06
	N10	−0.80	1.07	1.43
	N17	1.38	−0.30	0.31
	N21	−0.93	−0.17	−0.19
	N24	3.26	2.00	1.72
E3	**N03**	1.84	1.82	1.96
	N10	1.48	2.82	3.09
	N20	3.44	2.17	2.43
	N21	1.77	1.58	0.88
	N22	1.45	0.13	0.39
E4	**N03**	−0.62	−0.36	−1.46
	N17	0.71	−0.72	−0.55
	N23	2.30	2.70	2.23
CO_2	**N01**	5.18	5.14	5.32

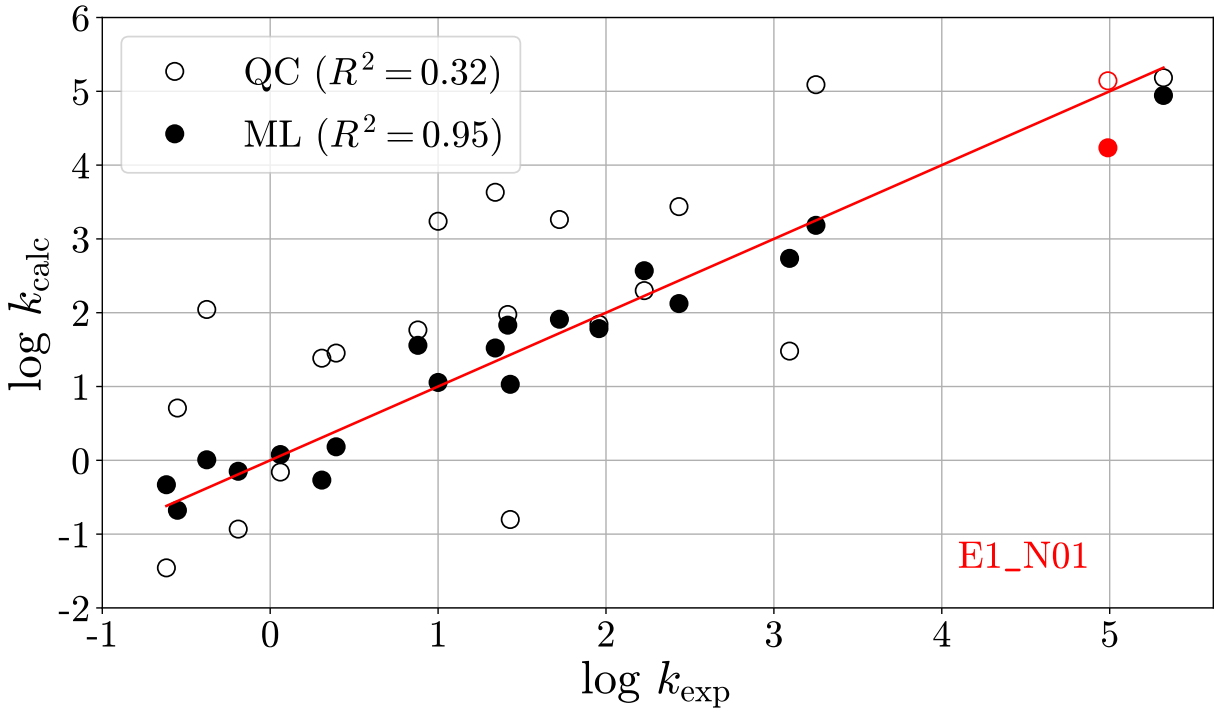

Figure S2: Leave-one-out prediction for **E1-N01**.

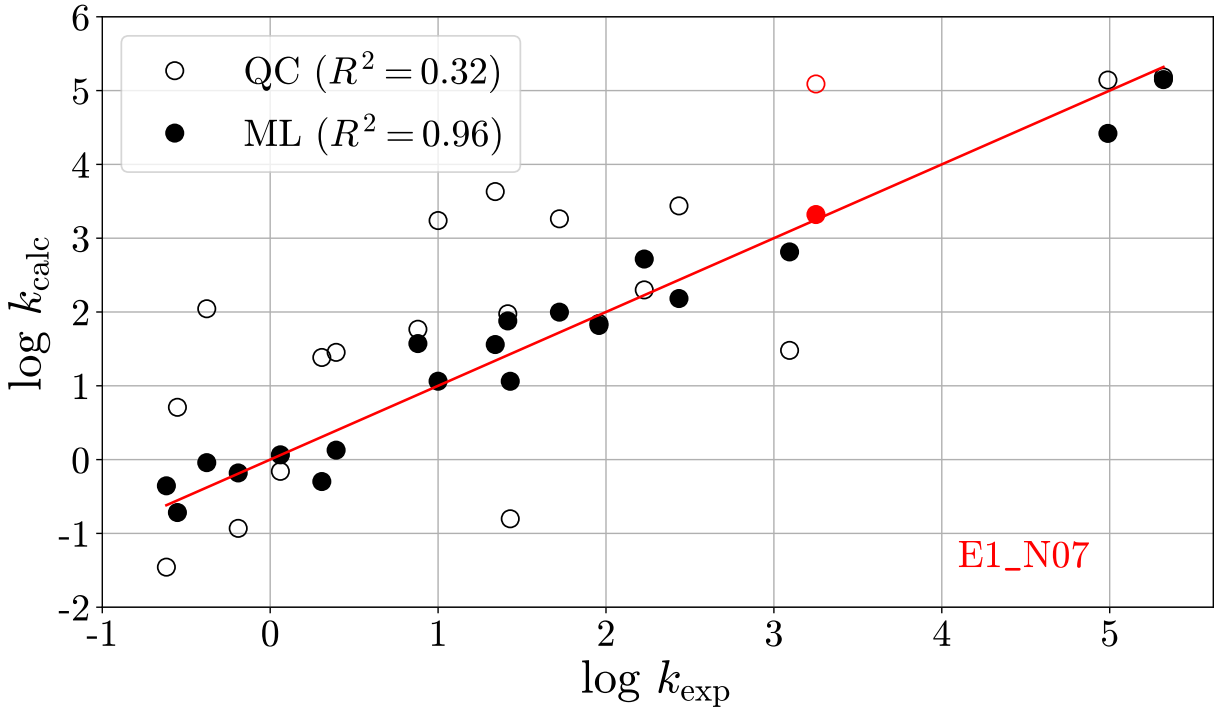

Figure S3: Leave-one-out prediction for **E1-N07**.

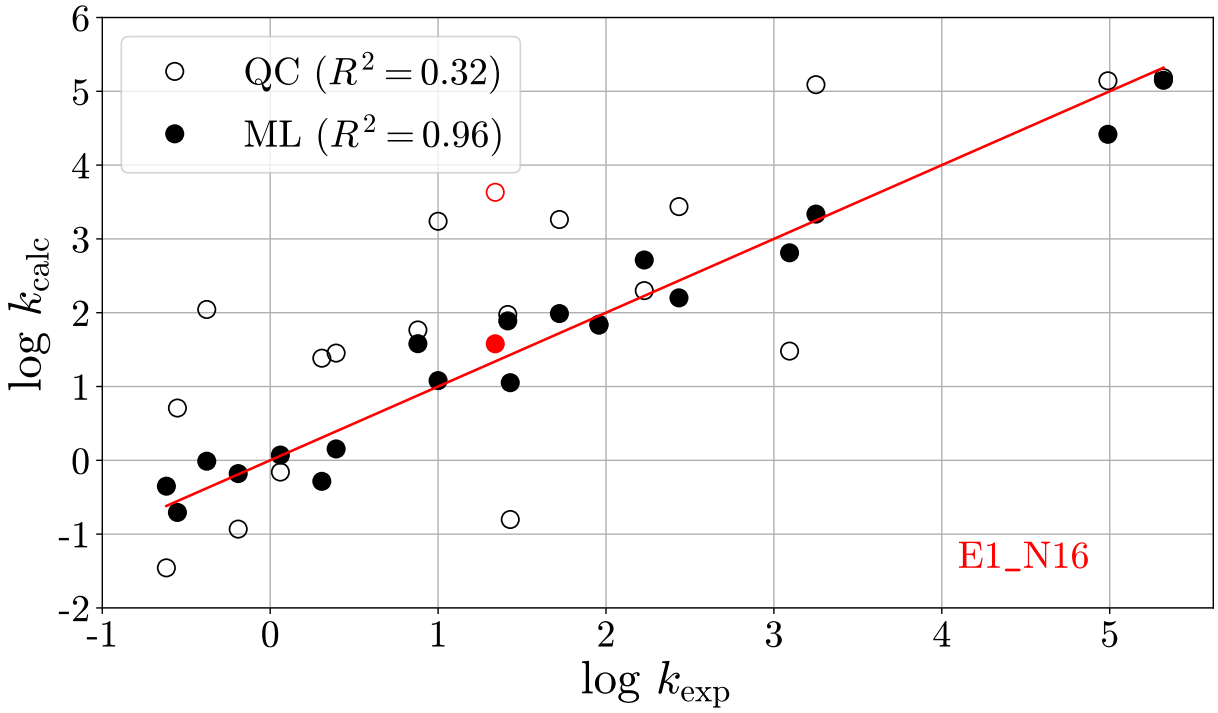

Figure S4: Leave-one-out prediction for **E1-N16**.

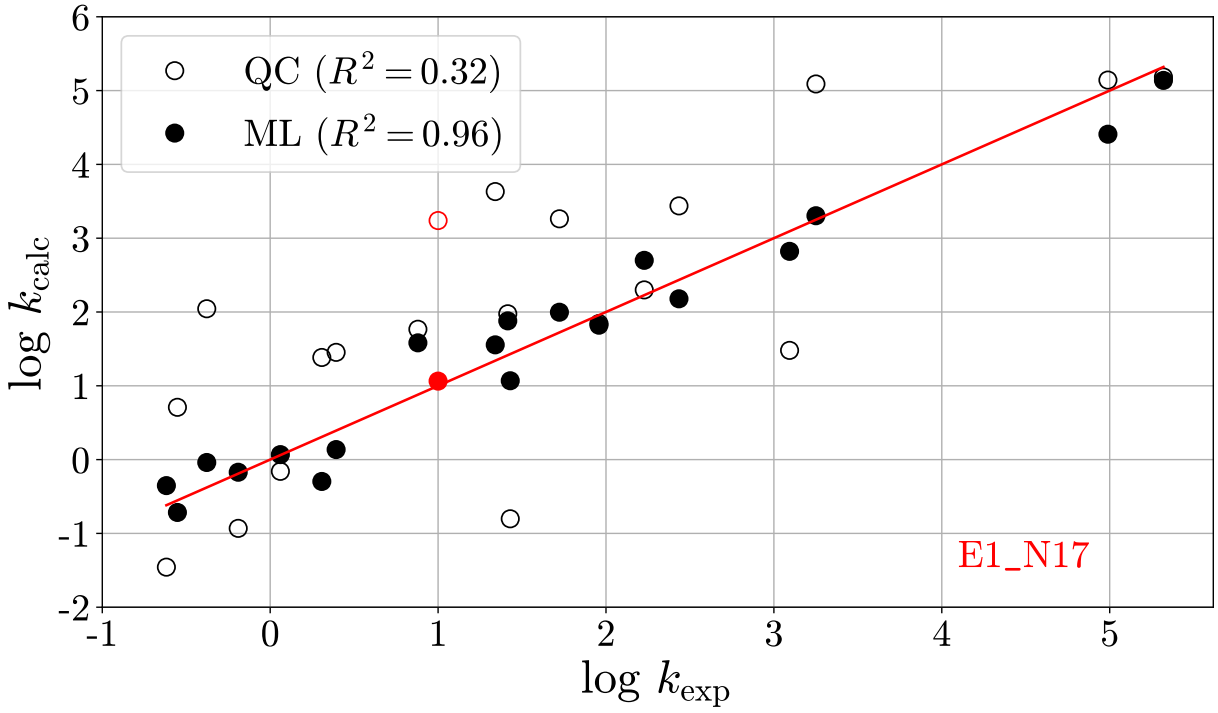

Figure S5: Leave-one-out prediction for **E1-N17**.

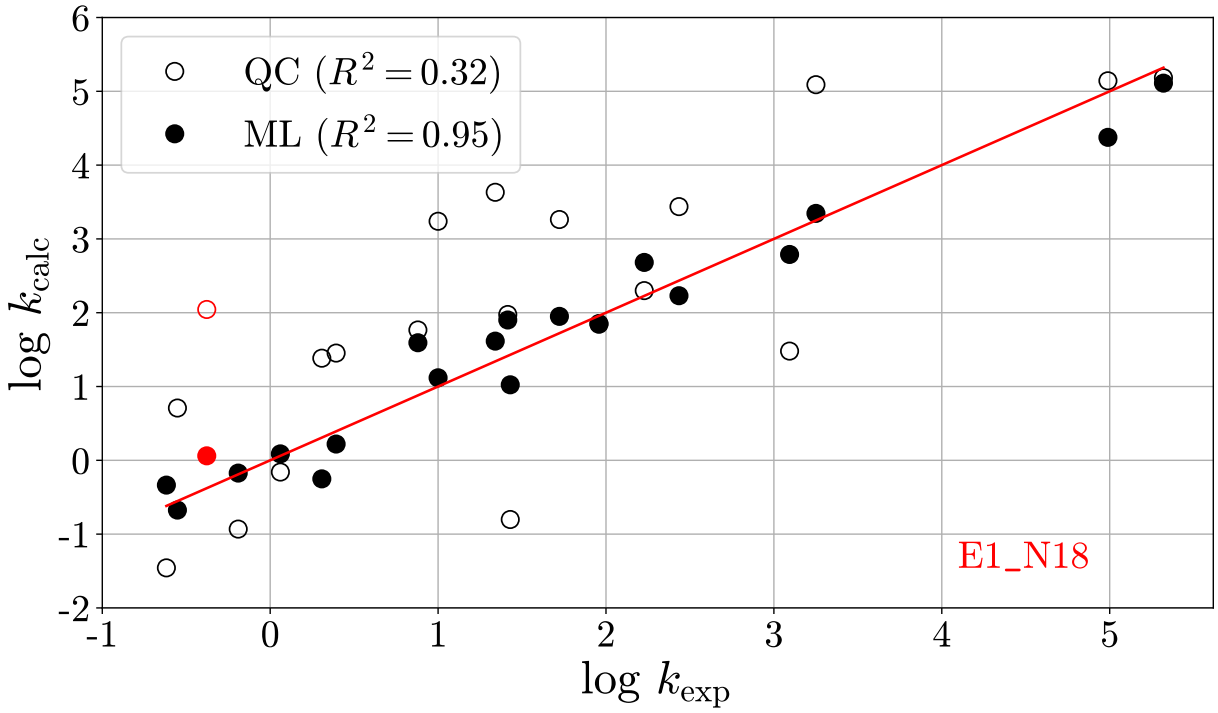

Figure S6: Leave-one-out prediction for **E1-N18**.

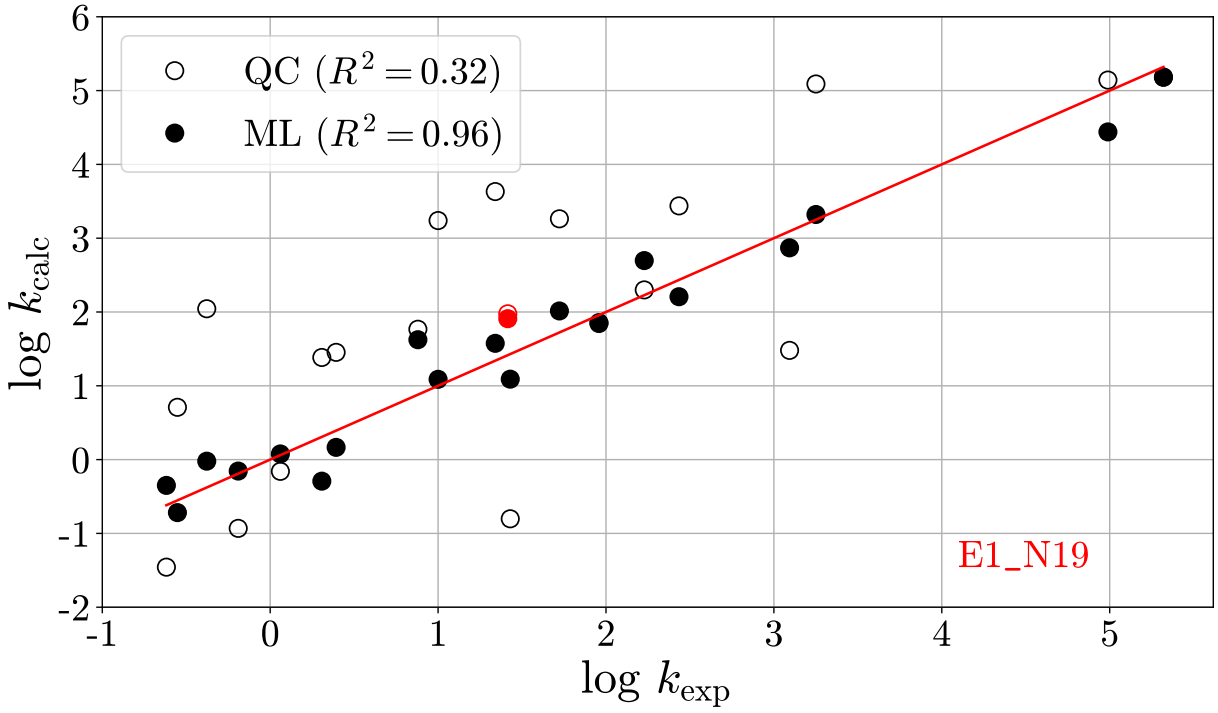

Figure S7: Leave-one-out prediction for **E1-N19**.

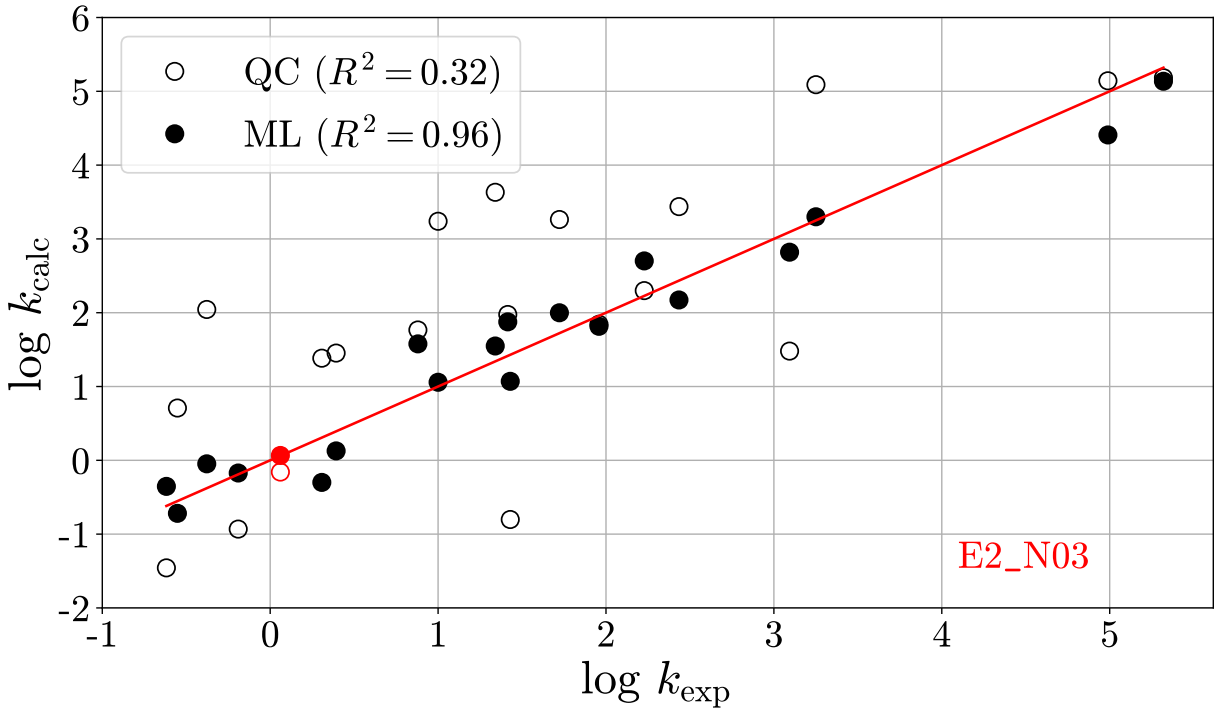

Figure S8: Leave-one-out prediction for **E2-N03**.

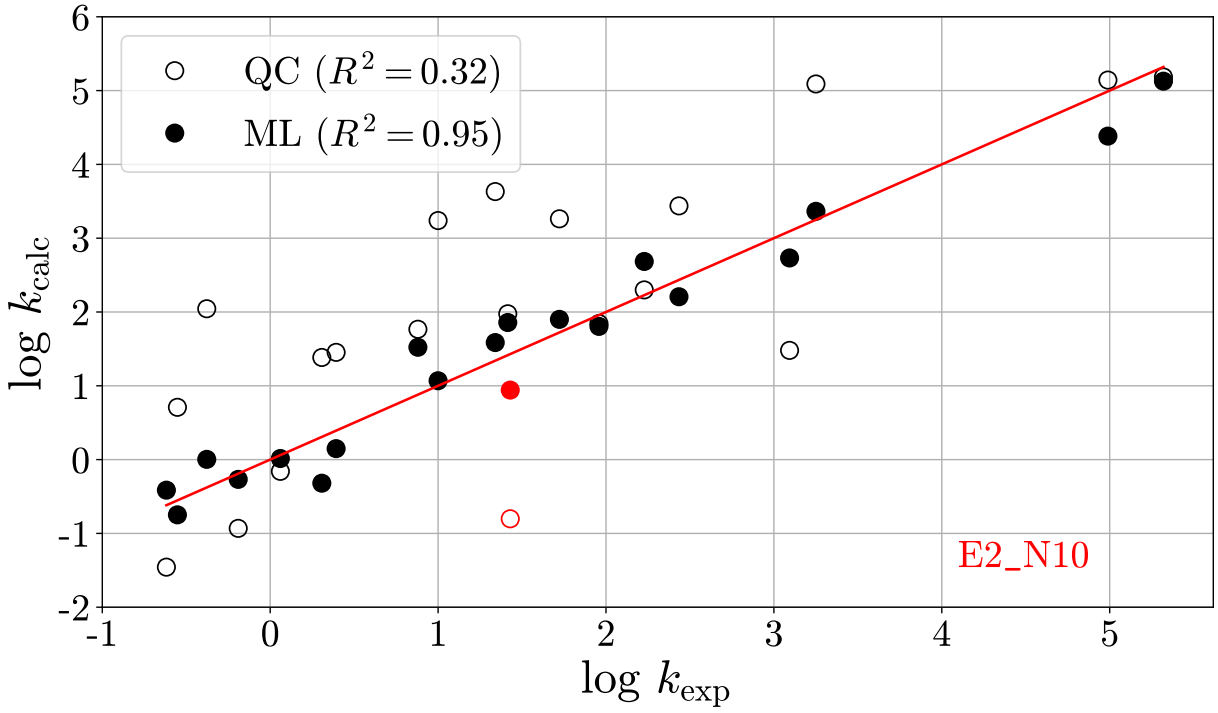

Figure S9: Leave-one-out prediction for **E2-N10**.

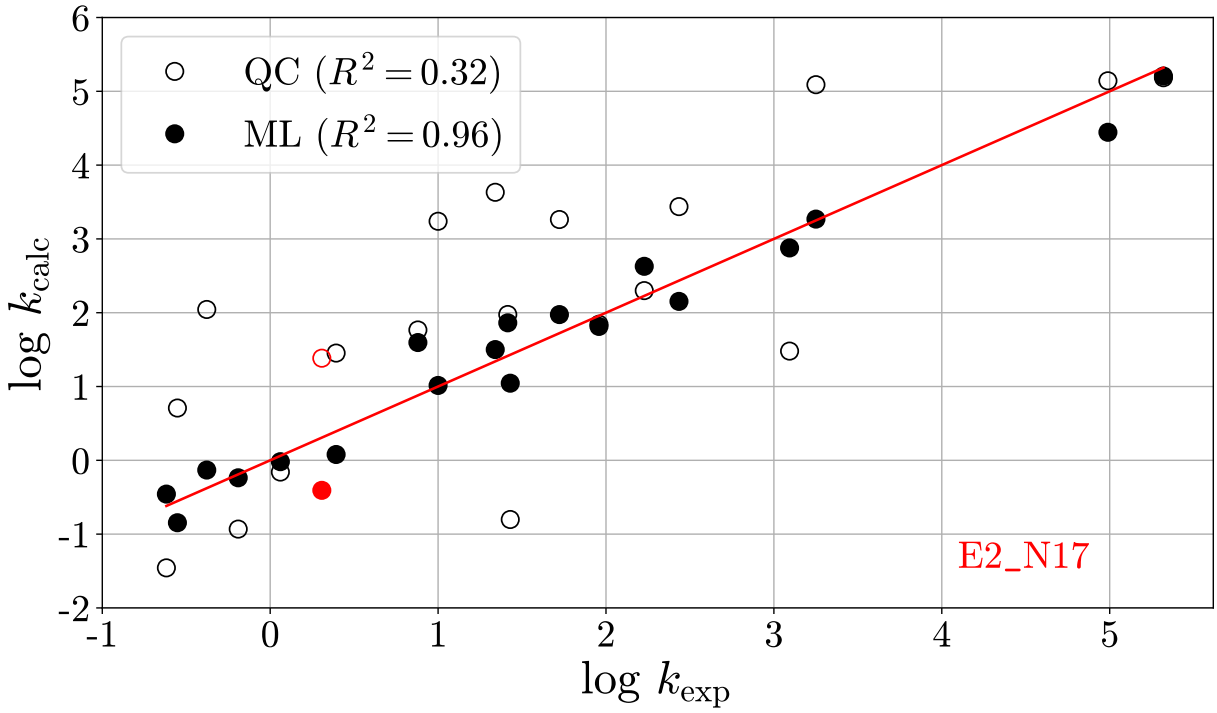

Figure S10: Leave-one-out prediction for **E2-N17**.

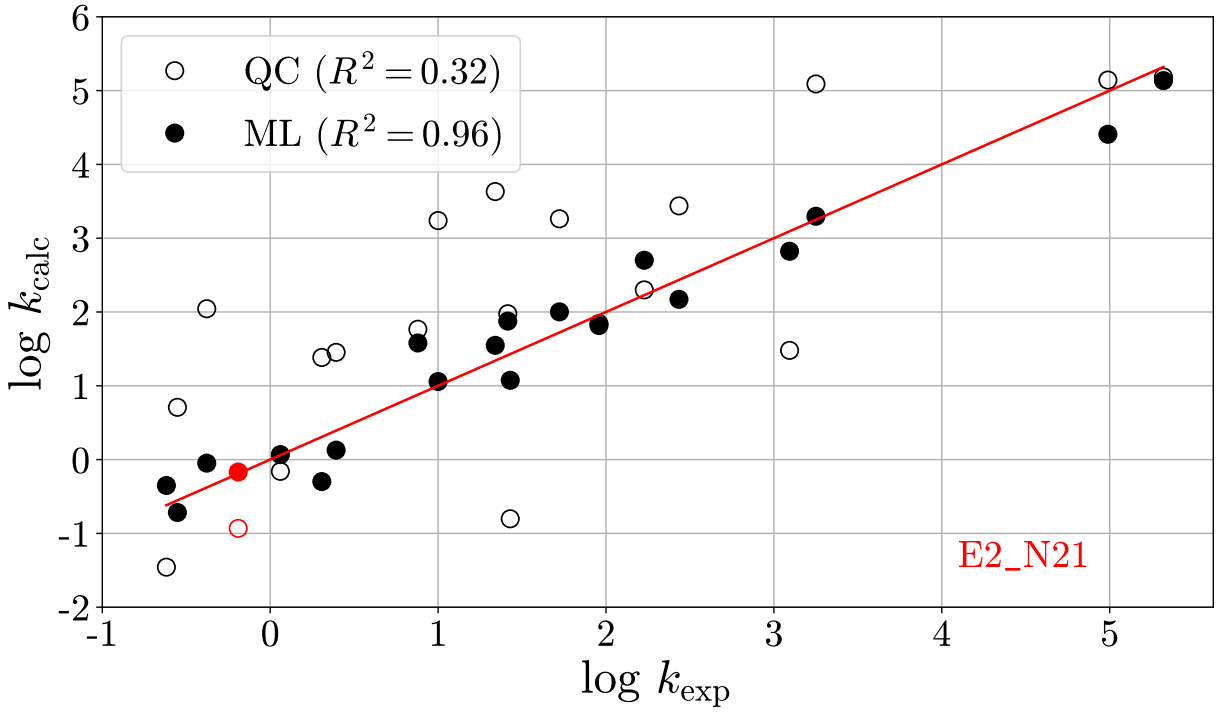

Figure S11: Leave-one-out prediction for **E2-N21**.

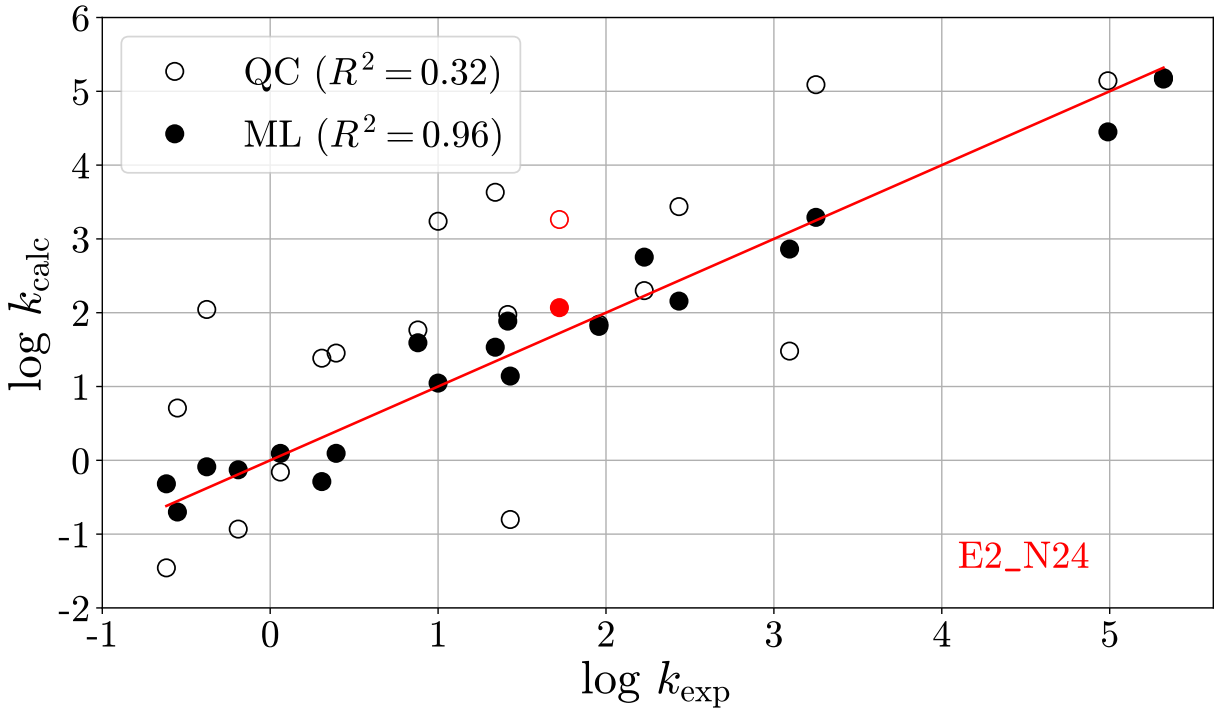

Figure S12: Leave-one-out prediction for **E2-N24**.

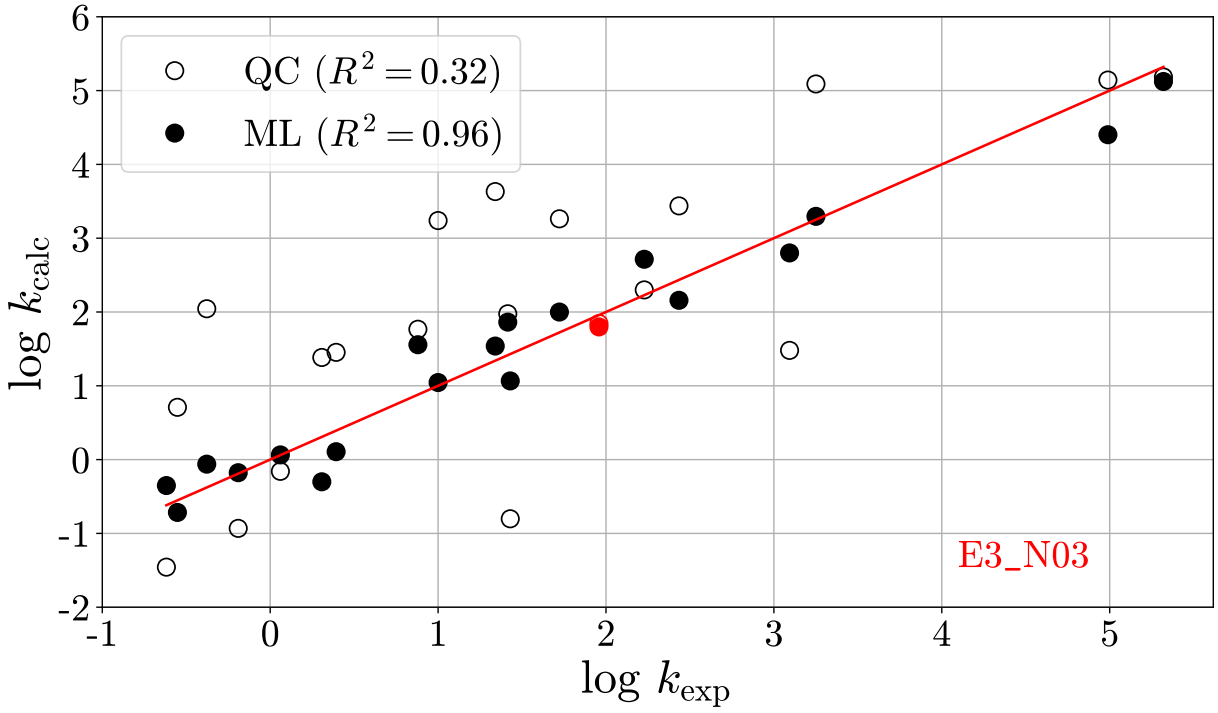

Figure S13: Leave-one-out prediction for **E3-N03**.

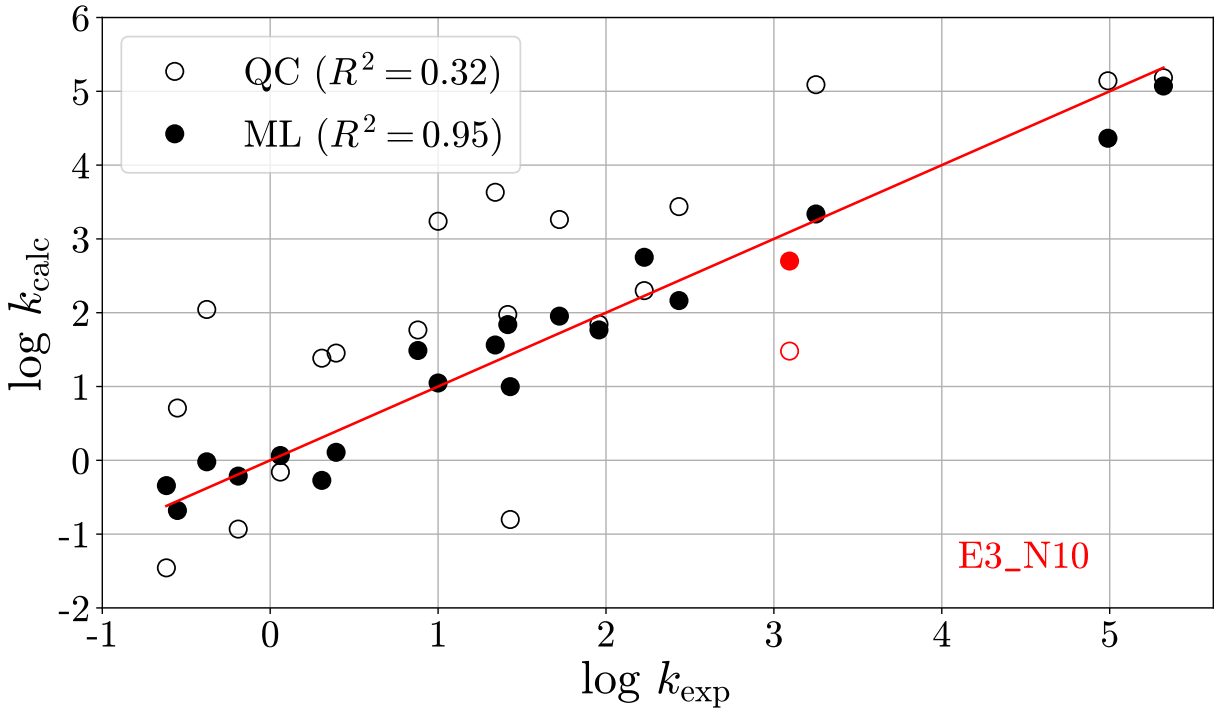

Figure S14: Leave-one-out prediction for **E3-N10**.

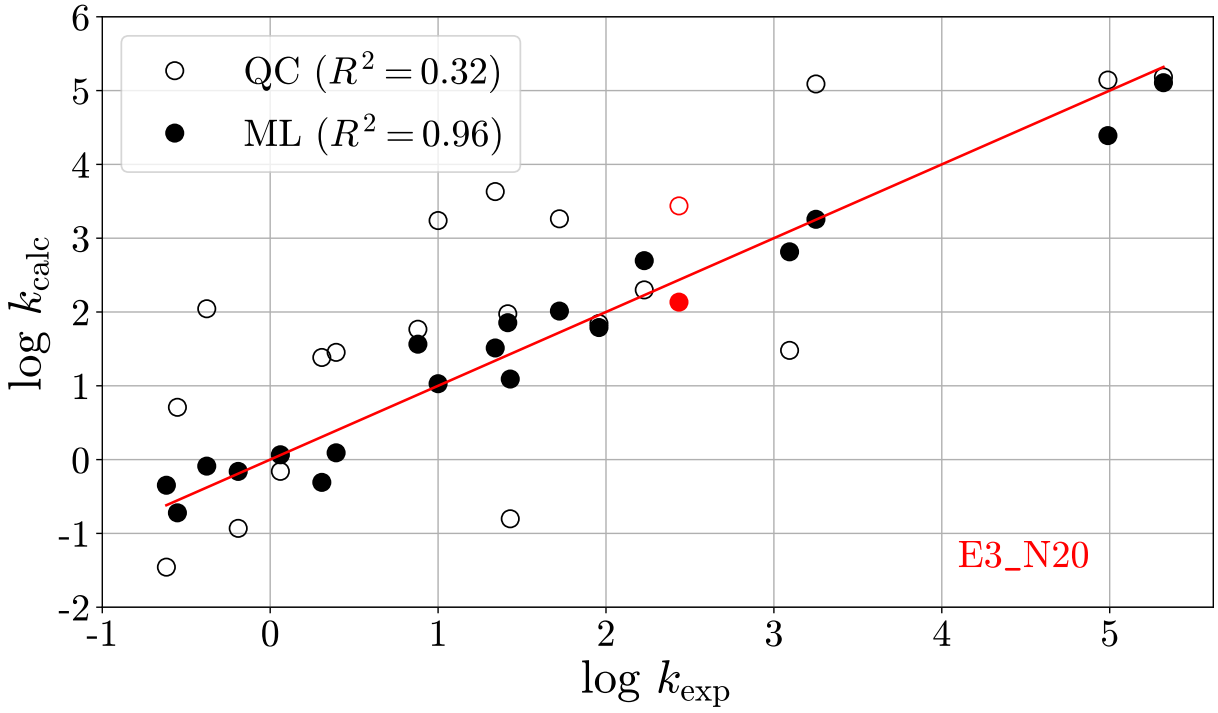

Figure S15: Leave-one-out prediction for **E3-N20**.

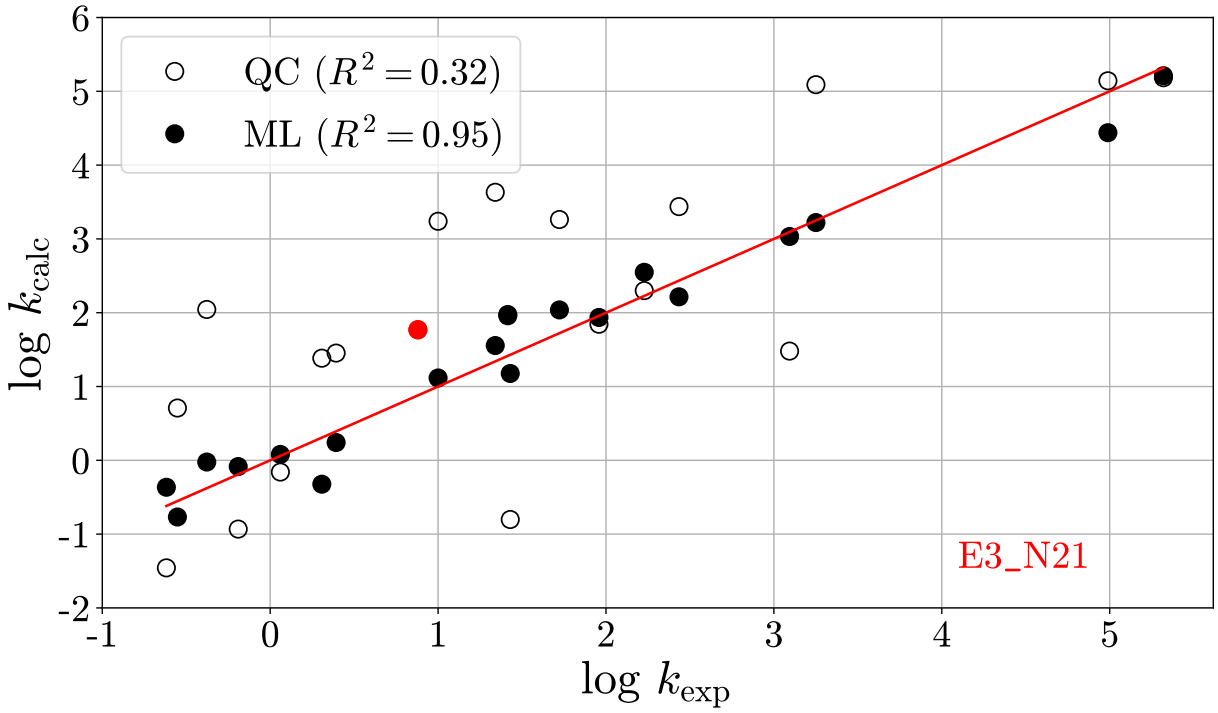

Figure S16: Leave-one-out prediction for **E3-N21**.

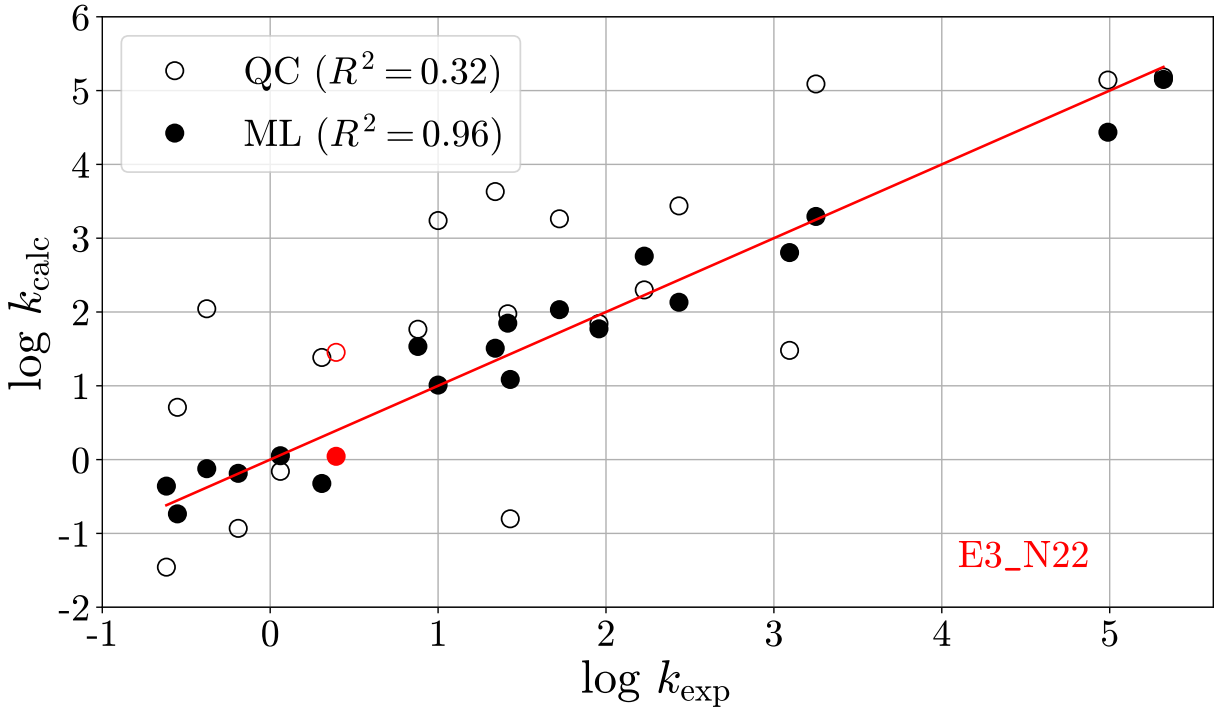

Figure S17: Leave-one-out prediction for **E3-N22**.

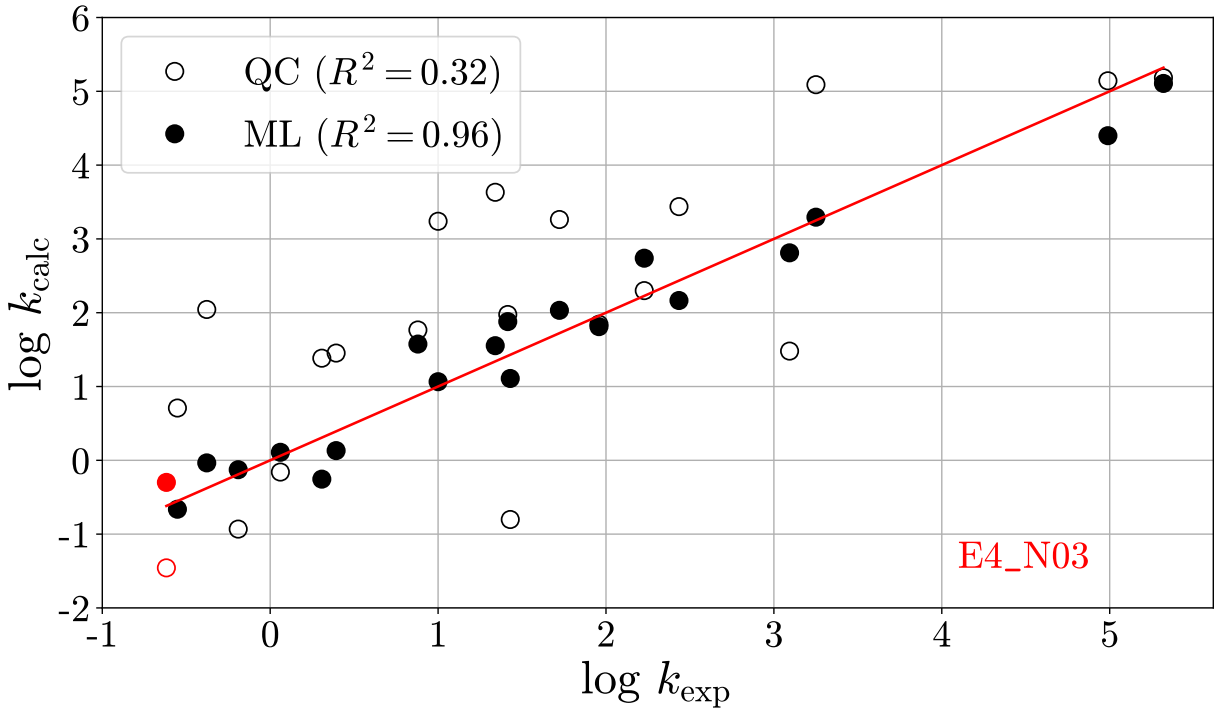

Figure S18: Leave-one-out prediction for **E4-N03**.

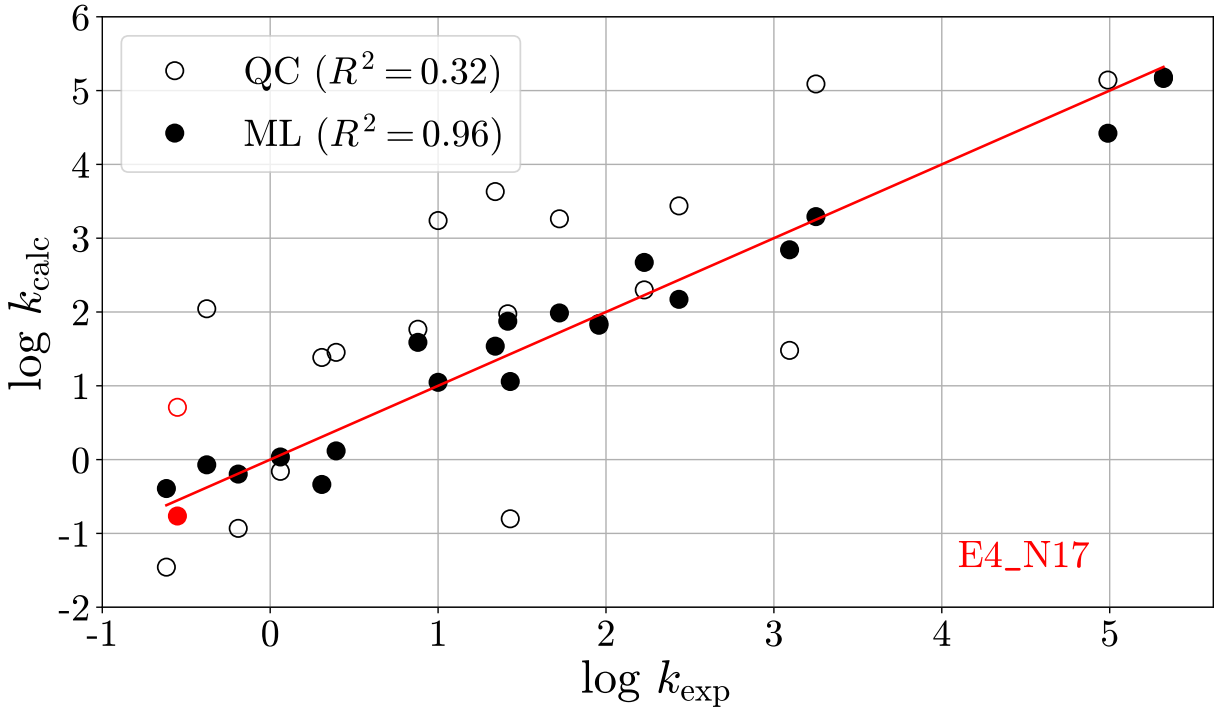

Figure S19: Leave-one-out prediction for **E4-N17**.

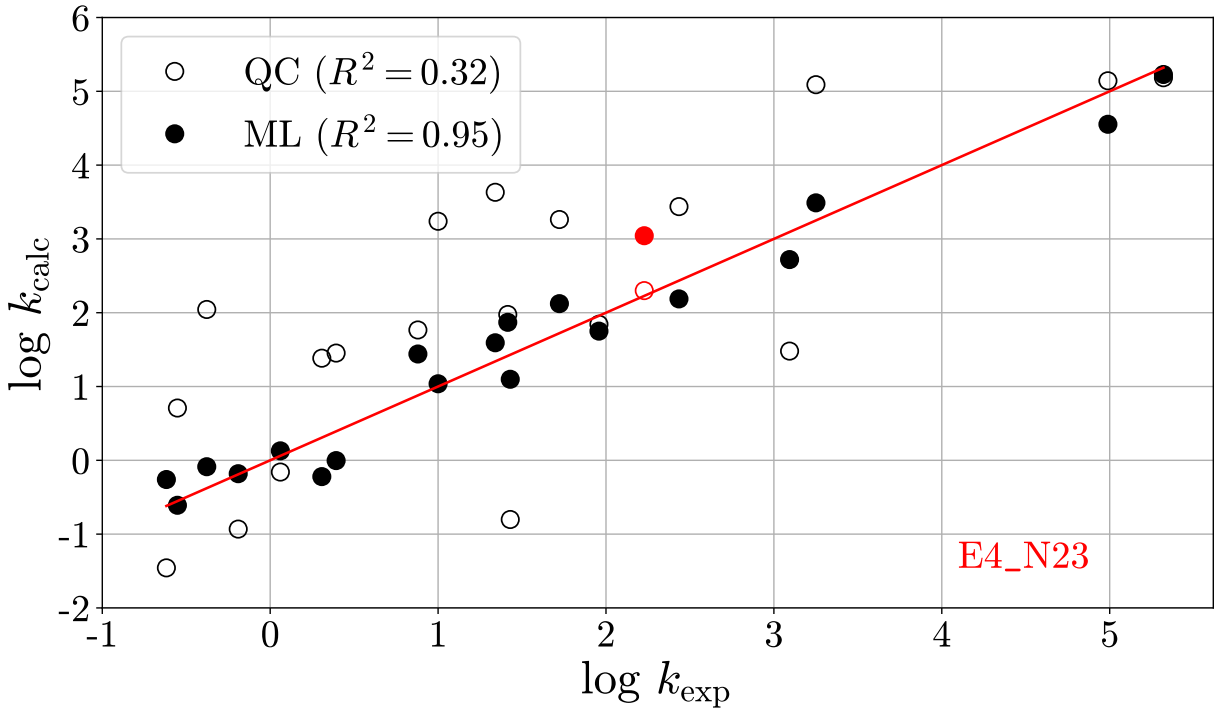

Figure S20: Leave-one-out prediction for **E4-N23**.

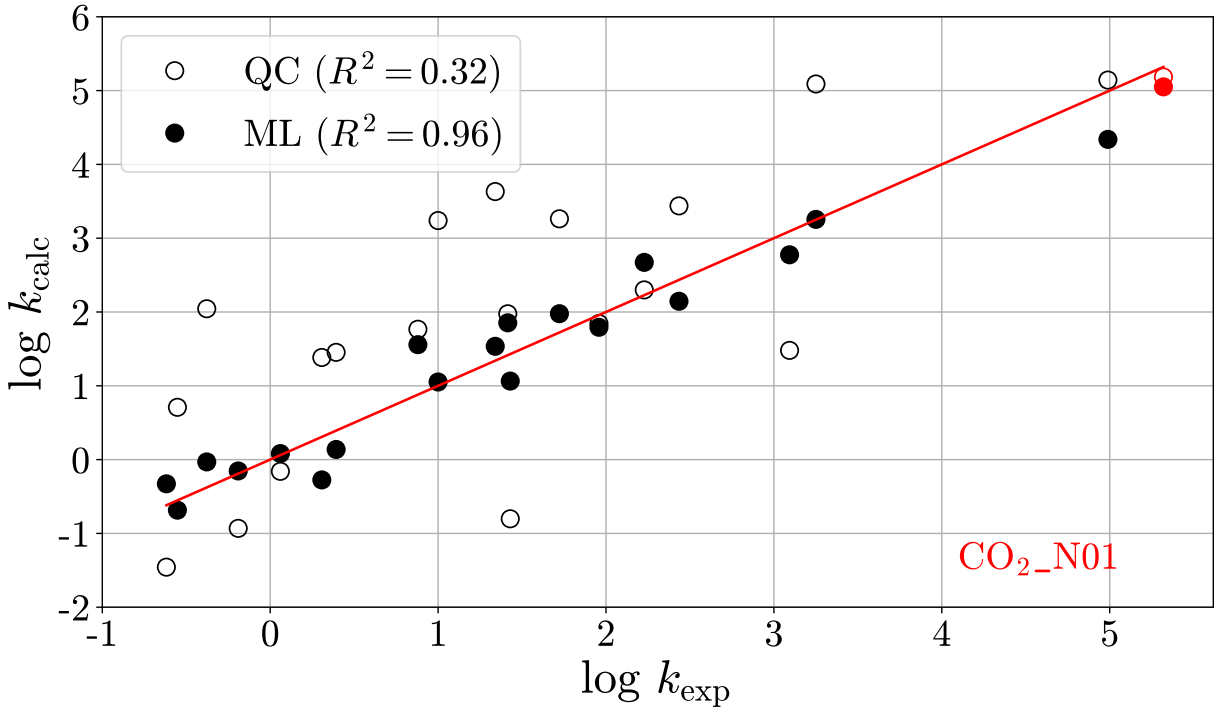

Figure S21: Leave-one-out prediction for **CO_2-N01**.

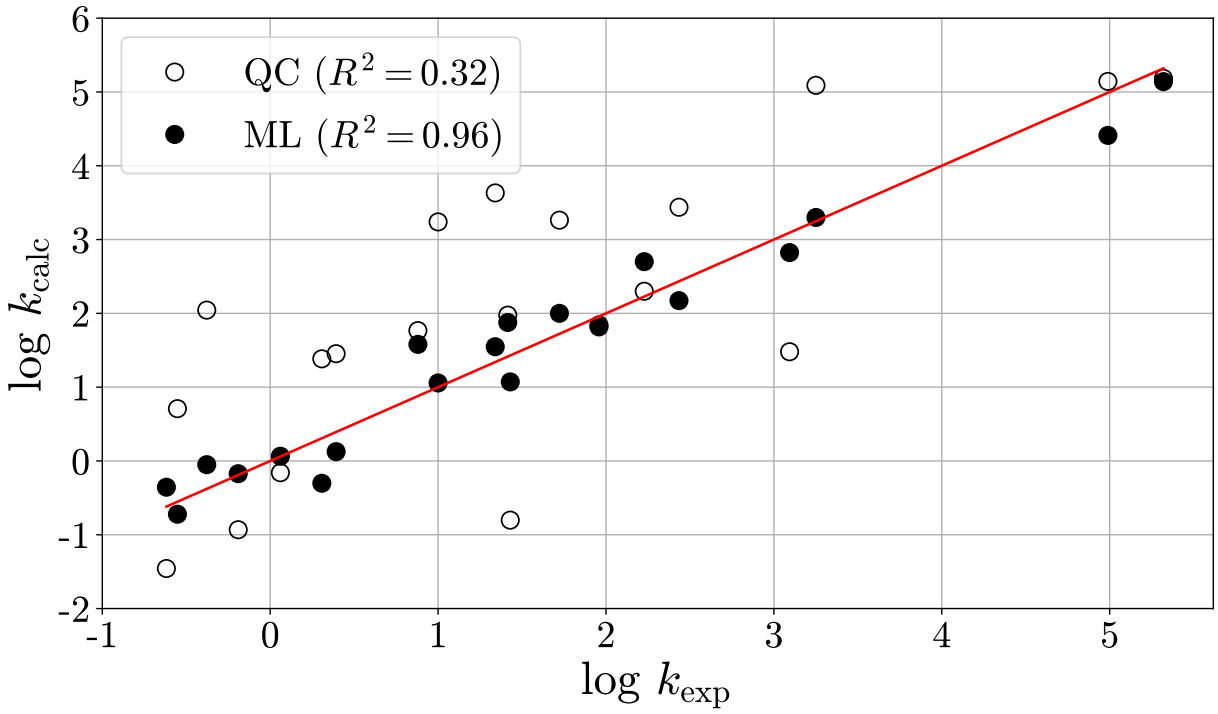

Figure S22: Final ML model.

Table S8: Parameters of heteroallene–carbanion reactions tested for the development of a multivariate linear (ML) model approximating $\log k$, including Mayr's nucleophilicity parameters (N, s_N), $\varepsilon_{elec}^{HOMO}$ and $\varepsilon_{elec}^{LUMO}$ of the electrophile, the electronic chemical potential of the electrophile (μ_{elec}), the single point energy in DMSO (E_{DMSO}), the negative temperature-scaled ($T = 293.15$) vibrational, translational, and rotational entropies ($-T \cdot S_{vib}$,$-T \cdot S_{trans}$,$-T \cdot S_{rot}$), the electronic energy (E_{elec}), the zero point vibrational energy (ZPVE), the enthalpy (H), the temperature-scaled overall entropy ($T \cdot S$), and the Gibbs free energy in solution ($G_{sol}(T)$) for all molecules (nucleophiles, electrophiles and transition states), respectively. The green columns show the parameters included in the final ML model.

molecule	N	s_N	$\varepsilon_{elec}^{HOMO}$ [E_h]	$\varepsilon_{elec}^{LUMO}$ [E_h]	μ_{elec} [E_h]	E_{DMSO} [E_h]	$-T \cdot S_{vib}$ [E_h]	$-T \cdot S_{trans}$ [E_h]	$-T \cdot S_{rot}$ [E_h]	E_{elec} [E_h]	ZPVE [E_h]	H [E_h]	$T \cdot S$ [E_h]	$G_{sol}(T)$ [E_h]
E1	–	–	−0.28300	−0.08468	−0.18384	−834.572244	−0.001042	−0.018133	−0.007287	−834.260305	0.006906	−834.249429	0.026464	−834.275893
E2	–	–	−0.24159	−0.05973	−0.15066	−722.880062	−0.008091	−0.018934	−0.013786	−722.400903	0.102008	−722.290842	0.039896	−722.330738
E3	–	–	−0.26667	−0.12023	−0.19345	−927.479329	−0.013205	−0.019334	−0.014629	−926.769802	0.104675	−926.654683	0.045573	−926.700257
E4	–	–	−0.23763	−0.04319	−0.14041	−611.180231	−0.018090	−0.019439	−0.015130	−610.531787	0.197065	−610.322389	0.049828	−610.372217
CO_2	–	–	−0.38805	−0.01522	−0.20164	−188.674154	−0.000313	−0.017372	−0.006096	−188.459565	0.011810	−188.444264	0.023782	−188.468045
N01	24.16	0.68	–	–	–	−347.433396	−0.004609	−0.018711	−0.013155	−346.990065	0.126264	−346.856899	0.036483	−346.893382
N03	18.82	0.69	–	–	–	−460.090669	−0.013947	−0.018870	−0.013713	−459.514771	0.141838	−459.362049	0.045423	−459.407472
N07	23.64	0.65	–	–	–	−630.816700	−0.018917	−0.019395	−0.014882	−630.077320	0.185407	−629.878482	0.051413	−629.929896
N10	19.36	0.67	–	–	–	−224.619879	−0.002992	−0.017916	−0.011542	−224.289634	0.031245	−224.252802	0.032445	−224.285247
N16	20.00	0.71	–	–	–	−743.138352	−0.019807	−0.019536	−0.015152	−742.277803	0.190077	−742.073827	0.052621	−742.126448
N17	18.67	0.68	–	–	–	−1157.11814	−0.017804	−0.019632	−0.014912	−1156.033499	0.129534	−1155.89096	0.050510	−1155.941469
N18	17.33	0.74	–	–	–	−1494.32533	−0.026667	−0.020003	−0.015849	−1492.865465	0.134219	−1492.714689	0.058872	−1492.773561
N19	19.67	0.68	–	–	–	−568.104858	−0.011476	−0.019179	−0.014344	−567.430153	0.117753	−567.302289	0.044379	−567.346668
N20	20.22	0.65	–	–	–	−574.685791	−0.018914	−0.019162	−0.014411	−573.985194	0.175229	−573.796913	0.050401	−573.847314
N21	17.64	0.73	–	–	–	−345.490695	−0.009124	−0.018502	−0.012909	−345.038309	0.108420	−344.921219	0.040147	−344.961366
N22	16.28	0.75	–	–	–	−1249.41298	−0.021335	−0.019780	−0.015389	−1248.234501	0.128844	−1248.090984	0.054281	−1248.145265
N23	25.11	0.54	–	–	–	−455.782938	−0.010276	−0.018994	−0.014050	−455.228805	0.113214	−455.106055	0.042834	−455.148889
N24	21.54	0.62	–	–	–	−283.974703	−0.003986	−0.018097	−0.011722	−283.577002	0.063698	−283.507224	0.033781	−283.541005
E1-N01	24.16	0.68	−0.28300	−0.08468	−0.18384	−1182.00443	−0.013117	−0.019417	−0.014584	−1181.267389	0.134363	−1181.122389	0.045826	−1181.168216
E1-N07	23.64	0.65	−0.28300	−0.08468	−0.18384	−1465.38682	−0.029558	−0.019868	−0.015781	−1464.344915	0.192760	−1464.134798	0.060998	−1464.195796
E1-N16	20.00	0.71	−0.28300	−0.08468	−0.18384	−1577.70525	−0.030514	−0.019969	−0.015973	−1576.540306	0.197196	−1576.325167	0.062202	−1576.387369
E1-N17	18.67	0.68	−0.28300	−0.08468	−0.18384	−1991.68502	−0.027352	−0.020040	−0.015707	−1990.297424	0.136977	−1990.143588	0.059407	−1990.202996
E1-N18	17.33	0.74	−0.28300	−0.08468	−0.18384	−2328.89002	−0.036220	−0.020326	−0.016467	−2327.125192	0.141796	−2326.963026	0.067487	−2327.030514
E1-N19	19.67	0.68	−0.28300	−0.08468	−0.18384	−1402.66826	−0.021342	−0.019717	−0.015449	−1401.691109	0.124937	−1401.552193	0.053802	−1401.605995
E2-N21	17.64	0.73	−0.24159	−0.05973	−0.15066	−1068.36106	−0.025709	−0.019700	−0.015320	−1067.450869	0.211246	−1067.223062	0.058120	−1067.281182
E2-N03	18.82	0.69	−0.24159	−0.05973	−0.15066	−1182.96294	−0.031482	−0.019868	−0.015682	−1181.925908	0.244592	−1181.662439	0.063169	−1181.725608
E2-N10	19.36	0.67	−0.24159	−0.05973	−0.15066	−947.488859	−0.019686	−0.019481	−0.015019	−946.6993580	0.133838	−946.552142	0.051748	−946.603890
E2-N17	18.67	0.68	−0.24159	−0.05973	−0.15066	−1879.99271	−0.037855	−0.020292	−0.016277	−1878.441993	0.231212	−1878.189425	0.068549	−1878.257974
E2-N24	21.54	0.62	−0.24159	−0.05973	−0.15066	−1006.85201	−0.021766	−0.019542	−0.015004	−1005.993141	0.165983	−1005.812939	0.053486	−1005.866424
E3-N03	18.82	0.69	−0.26667	−0.12023	−0.19345	−1387.56738	−0.036548	−0.020087	−0.016215	−1386.312522	0.247621	−1386.043662	0.068297	−1386.111959
E3-N10	19.36	0.67	−0.26667	−0.12023	−0.19345	−1152.09271	−0.025830	−0.019763	−0.015616	−1151.082127	0.136265	−1150.929902	0.057675	−1150.987576
E3-N20	20.22	0.65	−0.26667	−0.12023	−0.19345	−1502.16532	−0.042760	−0.020216	−0.016433	−1500.781347	0.280458	−1500.477254	0.073538	−1500.550792
E3-N21	17.64	0.73	−0.26667	−0.12023	−0.19345	−1272.96552	−0.032656	−0.019945	−0.015913	−1271.833198	0.213584	−1271.600345	0.064361	−1271.664706
E3-N22	16.28	0.75	−0.26667	−0.12023	−0.19345	−2176.88802	−0.046347	−0.020540	−0.017023	−2175.019900	0.233421	−2174.761091	0.077144	−2174.838235
E4-N03	18.82	0.69	−0.23763	−0.04319	−0.14041	−1071.26124	−0.040752	−0.020149	−0.016208	−1070.062933	0.339101	−1069.700718	0.071609	−1070.96768
E4-N17	18.67	0.68	−0.23763	−0.04319	−0.14041	−1768.29421	−0.043859	−0.020504	−0.016650	−1766.583807	0.326359	−1766.232321	0.075286	−1766.307607
E4-N23	25.11	0.54	−0.23763	−0.04319	−0.14041	−1066.96053	−0.036967	−0.020200	−0.016411	−1065.779140	0.309350	−1065.447925	0.068996	−1065.516921
CO_2-N01	24.16	0.68	−0.38805	−0.01522	−0.20164	−536.105174	−0.011755	−0.019162	−0.014149	−535.4620970	0.138931	−535.313209	0.043858	−535.357067

Table S9: Parameter values used for making predictions of $\log k$ with the ML model for CO_2–carbanion reactions. Further information on the parameters is provided in Table S8.

molecule	N	μ_{elec} $[E_\mathrm{h}]$	$T \cdot S_{\mathrm{nuc}}$ $[E_\mathrm{h}]$
CO_2	–	−0.201635	–
N01	24.16	–	0.036483
N02	16.06	–	0.047033
N03	18.82	–	0.045423
N04	15.24	–	0.043472
N05	19.62	–	0.041834
N06	27.54	–	0.047759
N07	23.64	–	0.051413
N08	24.99	–	0.042411
N09	16.55	–	0.043146
N10	19.36	–	0.032445
N11	29.50	–	0.053159
N12	17.46	–	0.044986
N13	13.91	–	0.042508
N14	13.83	–	0.057818
N15	20.51	–	0.057284

Section S5: Additional information on the determination of E(E1), E(E2), and E(E3)

Tables S11–S13 and Figure S23 present the findings related to the determination of electrophilicity parameters for heteroallenes **E1**–**E3**. To further evaluate our protocol, we simulated a scenario analogous to that of CO_2, conducting two separate simulations (ML_A and ML_B) for each heteroallene. In each simulation, the ML model was retrained using data from only one reaction of the selected heteroallene with a nucleophile that is also involved in another reaction. This setup mirrors the conditions present in the CO_2 reaction study. Each scenario is denoted as ML_A/ML_B(**NX**), where **NX** refers to the nucleophile included in the training dataset. The electrophilicity parameters were calibrated using all reactions of the electrophile, and the results are summarized in Tables S11–S13. The resulting distribution of E in each scenario incorporates the experimental result, thereby successfully validating our protocol.

Table S10: Experimental and calculated rate constants in the form of $\log k$ (in DMSO) for reactions of carbanions with **E1** (CS_2) and of the indenide anion (**N01**) with CO_2 as measured by Li et al.[13]

nuc.	Exp.	B3LYP	DLPNO-B2PLYP	DLPNO-CCSD(T)
N01	4.99	5.14	5.06	3.03
N07	3.25	5.09	5.27	4.16
N16	1.34	3.63	4.11	4.41
N17	1.00	3.24	3.02	1.22
N18	−0.38	2.04	1.97	0.47
N19	1.42	1.98	2.50	2.67
N01 + CO_2	5.18	5.42	4.39	2.54

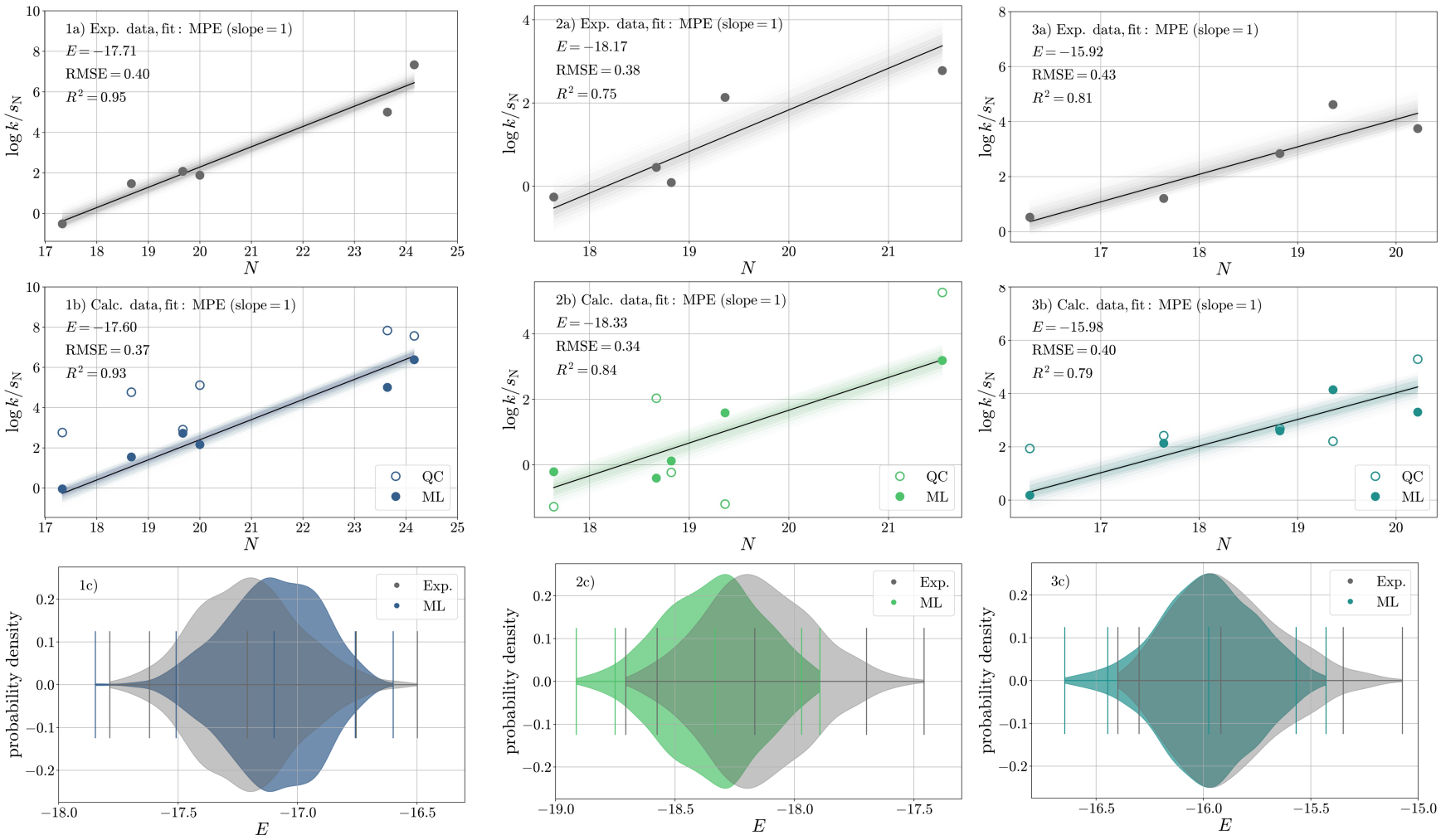

Figure S23: Calibration of E against a) experimental and b) ML-derived $\log k$ values for reactions of carbanions with heteroallenes **E1** (1), **E2** (2), **E3** (3) in DMSO based on bootstrapped least-squares optimisation with respect to the Mayr–Patz equation (MPE, fixed slope). c) Bootstrapped distributions of E for ML results (coloured) and experimental data (grey), respectively. See Section 3.3 in the main text for a more detailed analysis.

Table S11: Mayr's electrophilicity E for **E1** as well as performance statistics (RMSE, R^2) derived from the results shown in Fig. 5 and Fig. S23-1. Minimum and maximum values of the bootstrap distributions are denoted E_{min} and E_{max}, respectively. Test scenario results with only **NX** included as reaction partner for **E1** for retraining purposes are denoted ML_A(**NX**) and ML_B(**NX**), respectively. See Table 7 in the main text for more details.

	Exp.	ML	QC	ML_A(**N01**)	ML_B(**N07**)
E_{min}	−18.29	−18.35	−16.49	−17.89	−18.21
E_{lower}	−18.12	−18.01	−16.16	−17.64	−17.91
E	**−17.71**	**−17.60**	**−15.39**	−17.22	−17.39
E_{upper}	−17.25	−17.26	−14.64	−16.92	−17.05
E_{max}	−17.00	−17.10	−14.32	−16.82	−16.95
$u_{95}(E)$	0.87	0.75	1.51	0.72	0.86
RMSE($\log k$)	0.40	0.37	0.72	0.34	0.41
$R^2(\log k)$	0.95	0.93	0.68	0.94	0.90

Table S12: Mayr's electrophilicity E for **E2** as well as performance statistics (RMSE, R^2) derived from the results shown in Fig. S23-2. Minimum and maximum values of the bootstrap distributions are denoted E_{min} and E_{max}, respectively. Test scenario results with only **NX** included as reaction partner for **E2** for retraining purposes are denoted ML_A(**NX**) and ML_B(**NX**), respectively. See Table 7 in the main text for more details.

	Exp.	ML	QC	ML_A(**N17**)	ML_B(**N21**)
E_{min}	−18.71	−18.91	−20.04	−19.01	−19.17
E_{lower}	−18.57	−18.75	−19.49	−18.84	−18.99
E	**−18.17**	**−18.33**	**−18.36**	−18.43	−18.55
E_{upper}	−17.70	−17.97	−17.14	−18.07	−18.17
E_{max}	−17.46	−17.89	−16.56	−17.94	−18.02
$u_{95}(E)$	0.88	0.78	2.35	0.77	0.82
RMSE($\log k$)	0.38	0.34	1.06	0.32	0.35
$R^2(\log k)$	0.75	0.84	0.55	0.86	0.85

Table S13: Mayr's electrophilicity E for **E3** as well as performance statistics (RMSE, R^2) derived from the results shown in Fig. S23-3. Minimum and maximum values of the bootstrap distributions are denoted E_{min} and E_{max}, respectively. Test scenario results with only **NX** included as reaction partner for **E3** for retraining purposes are denoted ML_A(**NX**) and ML_B(**NX**), respectively. See Table 7 in the main text for more details.

	Exp.	ML	QC	ML_A(**N10**)	ML_B(**N21**)
E_{min}	−16.40	−16.65	−16.73	−16.71	−16.93
E_{lower}	−16.30	−16.45	−16.34	−16.44	−16.72
E	**−15.92**	**−15.98**	**−15.47**	−15.88	−16.26
E_{upper}	−15.35	−15.57	−14.81	−15.35	−15.84
E_{max}	−15.07	−15.43	−14.46	−15.02	−15.58
$u_{95}(E)$	0.95	0.88	1.53	1.10	0.88
RMSE($\log k$)	0.43	0.40	0.69	0.47	0.37
$R^2(\log k)$	0.81	0.79	0.13	0.74	0.83

Section S6: Additional information on the determination of $E(\text{CO}_2)$

Table S14: Gibbs free energies of activation in solution ($\Delta G^\ddagger_{\text{sol}}$) and logarithmic rate constants derived from quantum chemical calculations ($\log k_{\text{QC}}$) and our ML model ($\log k_{\text{ML}}$) for each CO_2–nucleophile reaction.

nuc.	$\Delta G^\ddagger_{\text{sol}}$ [kcal/mol]	$\log k_{\text{QC}}$	$\log k_{\text{ML}}$
N01	10.20	5.18	5.14
N02	7.94	6.87	0.78
N03	11.27	4.39	2.15
N04	13.83	2.47	0.61
N05	10.73	4.78	2.73
N06	6.99	7.57	6.04
N07	9.31	5.84	4.03
N08	7.55	7.16	5.18
N09	11.62	4.13	1.24
N10	10.49	4.97	3.16
N11	8.82	6.21	6.63
N12	12.05	3.80	1.55
N13	14.59	1.91	0.05
N14	13.75	2.54	−0.88
N15	6.78	7.73	2.24

Reaction specific uncertainty in $\log k$ (95 % confidence) can be estimated analytically by

Table S15: This is a more comprehensive version of Table 7 in the main text. Minimum and maximum values of the bootstrap distribution are denoted $E_{\mathrm{min}}/s_{\mathrm{E,min}}$ and $E_{\mathrm{max}}/s_{\mathrm{E,max}}$, respectively.

	MPE (eqn (1))	GE (eqn (6))
E_{min}	−16.27	−15.39
E_{lower}	−16.00	−15.05
E	**−15.45**	**−14.62**
E_{upper}	−14.97	−14.18
E_{max}	−14.52	−13.96
$u_{95}(E)$	1.03	0.87
$s_{\mathrm{E,min}}$	–	0.69
$s_{\mathrm{E,lower}}$	–	0.73
s_{E}	**1.00**	**0.81**
$s_{\mathrm{E,upper}}$	–	0.87
$s_{\mathrm{E,max}}$	–	0.89
$u_{95}(s_{\mathrm{E}})$	–	0.13
RMSE($\log k$)	0.72	0.40
$R^2(\log k)$	0.89	0.97

assuming normally distributed errors,[21]

$$u_{.95}(s_{\mathrm{N}}, N) = 1.96 \cdot \sqrt{\frac{R}{R-M-1}\mathrm{MSE} + \begin{pmatrix} s_{\mathrm{N}} & s_{\mathrm{N}}N \end{pmatrix} \mathbf{\Sigma} \begin{pmatrix} s_{\mathrm{N}} \\ s_{\mathrm{N}}N \end{pmatrix}} \tag{1}$$

Mayr's nucleophile-specific parameters (s_{N}, N) serve as input, $R = 15$ is the number of reactions and M denotes the number of input variables, here $M = 1$ (GE). The MSE is the mean squared error in log k, here MSE=0.16, and $\mathbf{\Sigma}$ is the covariance matrix of the model parameters, which reads

$$\mathbf{\Sigma} = \begin{pmatrix} \mathrm{var}(s_{\mathrm{E}}E) & \mathrm{cov}(s_{\mathrm{E}}E, s_{\mathrm{E}}) \\ \mathrm{cov}(s_{\mathrm{E}}, s_{\mathrm{E}}E) & \mathrm{var}(s_{\mathrm{E}}) \end{pmatrix} = \begin{pmatrix} +0.3771 & -0.0196 \\ -0.0196 & +0.0011 \end{pmatrix} \tag{2}$$

The reaction-specific uncertainty values corresponding to the carboxylation reactions considered in this work are summarized in Table S16. In general, $u_{.95}(s_{\mathrm{N}}, N)$ exhibits a slight increase near the extremes of the nucleophilicity scale under investigation, a behaviour characteristic of such systems.

Table S16: Reaction-specific uncertainty in log k_{GE} (95 % confidence) as a function of s_N and N.

Nucleophile	s_N	N	$u_{.95}(s_N, N)$
N14	0.84	13.83	0.9045
N13	0.86	13.91	0.9062
N04	0.74	15.24	0.8775
N02	0.69	16.06	0.8681
N09	0.78	16.55	0.8725
N12	0.72	17.46	0.8653
N03	0.69	18.82	0.8631
N11	0.67	19.36	0.8628
N05	0.67	19.62	0.8635
N15	0.64	20.51	0.8648
N07	0.65	23.64	0.8893
N01	0.68	24.16	0.8999
N08	0.60	24.99	0.8961
N06	0.57	27.54	0.9212
N10	0.50	29.50	0.9263

Section S7: Comparison of the determination of E(E1), E(E2), and E(E3) based on both MPE and GE

We investigated the scenario involving CS_2 using the same nucleophiles as in our study on CO_2 to determine the parameter E. Based on the same set of 15 nucleophiles whose reactions with CO_2 were examined, our findings indicate that the $s_E = 0.78$ parameter should also be applied to CS_2. We examined the ML-predicted results for six reactions involving CS_2 for which experimental data are available. When fitting the generalised equation to ML-predicted log k values of these reactions, we find $s_E = 0.85$ (see Table S17, column ML (GE)). However, when we use the corresponding experimental log k values instead, we obtain $s_E = 1.01$ (see Table S17, column Exp. (GE)), despite a strong correlation (coefficient of determination of $R^2 = 0.96$) between ML-predicted and experimental log k values. A comparison of Figs. 5a and 5b in the main text, along with the strong correlation just mentioned, suggests that the significant difference between the two results is mainly due to the far-right data point (see Fig. S23-1a vs. Fig. S23-1b). In the ML-predicted data, this point is shifted downwards relative to the experimental data, causing a decrease in the slope (= s_E) of the fitted line. Regression coefficients obtained from larger datasets are less sensitive to such changes, which is why we conclude that an s_E parameter significantly smaller than 1 is more likely to reflect the actual situation that an s_E parameter close to 1. The same trends in the comparison of ML-predicted and experimental log k values can

be found for electrophiles **E2** and **E3** (see Tables S18 and Tables S19). However, for the extended set of 15 nucleophiles, our calculations suggest that s_{E} is consistently smaller than one (Table S20).

Table S17: This is a more comprehensive version of Table S11. Experimental and ML derived values for **E1** are fitted with MPE and GE (see Table S15).

	Exp. (MPE)	Exp. (GE)	ML (MPE)	ML (GE)
E_{min}	−18.45	−18.35	−18.39	−17.54
E_{lower}	−18.15	−18.04	−18.07	−17.42
E	**−17.69**	**−17.72**	**−17.58**	**17.10**
E_{upper}	−17.28	−17.12	−17.26	−16.37
E_{max}	−17.12	−16.73	−17.10	−15.37
$u_{95}(E)$	0.87	1.05	0.80	1.05
$s_{\mathrm{E,min}}$	–	0.73	–	0.61
$s_{\mathrm{E,lower}}$	–	0.80	–	0.72
s_{E}	**1.00**	**1.01**	**1.00**	**0.85**
$s_{\mathrm{E,upper}}$	–	1.15	–	0.93
$s_{\mathrm{E,max}}$	–	1.25	–	1.01
$u_{95}(s_{\mathrm{E}})$	–	0.35	–	0.21
RMSE($\log k$)	0.40	0.40	0.37	0.27
$R^2(\log k)$	0.95	0.95	0.93	0.97

Table S18: This is a more comprehensive version of Table S12. Experimental and ML derived values for **E2** are fitted with MPE and GE (see Table S15).

	Exp. (MPE)	Exp. (GE)	ML (MPE)	ML (GE)
E_{min}	−18.74	−18.35	−18.62	−19.76
E_{lower}	−18.57	−18.04	−18.46	−18.82
E	**−18.16**	**−18.02**	**−18.35**	**−18.35**
E_{upper}	−17.71	−17.24	−17.96	−17.87
E_{max}	−17.36	−16.14	−17.81	−17.66
$u_{95}(E)$	0.86	1.21	0.81	0.95
$s_{\mathrm{E,min}}$	–	0.43	–	0.22
$s_{\mathrm{E,lower}}$	–	0.70	–	0.66
s_{E}	**1.00**	**0.84**	**1.00**	**0.97**
$s_{\mathrm{E,upper}}$	–	1.29	–	1.26
$s_{\mathrm{E,max}}$	–	2.27	–	2.18
$u_{95}(s_{\mathrm{E}})$	–	0.59	–	0.60
RMSE($\log k$)	0.38	0.35	0.34	0.34
$R^2(\log k)$	0.75	0.79	0.84	0.85

Table S19: This is a more comprehensive version of Table S13. Experimental and ML derived values for **E3** are fitted with MPE and GE (see Table S15).

	Exp. (MPE)	Exp. (GE)	ML (MPE)	ML (GE)
E_{min}	−16.42	−17.01	−16.71	−16.53
E_{lower}	−16.28	−16.63	−16.48	−16.11
E	**−15.91**	**−15.90**	**−15.98**	**15.73**
E_{upper}	−15.34	−15.28	−15.55	−13.75
E_{max}	−14.84	−9.55	−15.29	−5.78
$u_{95}(E)$	0.94	1.35	0.93	2.37
$s_{\text{E,min}}$	–	0.40	–	0.26
$s_{\text{E,lower}}$	–	0.81	–	0.55
s_{E}	**1.00**	**1.01**	**1.00**	**0.89**
$s_{\text{E,upper}}$	–	1.46	–	1.20
$s_{\text{E,max}}$	–	1.88	–	1.45
$u_{95}(s_{\text{E}})$	–	0.65	–	0.65
RMSE($\log k$)	0.43	0.42	0.40	0.37
$R^2(\log k)$	0.81	0.82	0.79	0.82

Table S20: ML-derived values for 15 nucleophiles (cf. Table S14) used to obtain E and s_{E} for CO_2 and **E1**–**E3** fitted with GE.

	CO_2	**E1**	**E2**	**E3**
E_{min}	−15.39	−16.47	−18.90	−15.83
E_{lower}	−15.05	−16.20	−18.66	−15.56
E	**−14.62**	**−15.79**	**−18.26**	**−15.14**
E_{upper}	−14.18	−15.38	−17.90	−14.72
E_{max}	−13.96	−15.12	−17.61	−14.48
$u_{95}(E)$	0.87	0.82	0.76	0.84
$s_{\text{E,min}}$	0.69	0.67	0.60	0.69
$s_{\text{E,lower}}$	0.73	0.70	0.65	0.72
s_{E}	**0.81**	**0.78**	**0.73**	**0.79**
$s_{\text{E,upper}}$	0.87	0.83	0.77	0.85
$s_{\text{E,max}}$	0.89	0.85	0.80	0.87
$u_{95}(s_{\text{E}})$	0.13	0.13	0.12	0.13
RMSE($\log k$)	0.40	0.40	0.40	0.40
$R^2(\log k)$	0.97	0.97	0.97	0.97

References

(1) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(2) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(3) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(4) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(5) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(6) Weigend, F. Accurate Coulomb-fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

(7) Rappoport, D.; Furche, F. Property-Optimized Gaussian Basis Sets for Molecular Response Calculations. *J. Chem. Phys.* **2010**, *133*, 134105.

(8) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.

(9) Frisch, M. J. et al. *Gaussian 16 Revision C.01*; Gaussian Inc.: Wallingford (CT), United States, 2016.

(10) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

(11) Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

(12) Eckhoff, M.; Bublitz, K. L.; Proppe, J. Determination of the electrophilicity of CO2. `https://git.rz.tu-bs.de/proppe-group/co2_electrophilicity`, last accessed on 18 November 2024.

(13) Li, Z.; Mayer, R. J.; Ofial, A. R.; Mayr, H. From Carbodiimides to Carbon Dioxide: Quantification of the Electrophilic Reactivities of Heteroallenes. *J. Am. Chem. Soc.* **2020**, *142*, 8383–8402.

(14) Hoffmann, G.; Balcilar, M.; Tognetti, V.; Héroux, P.; Gaüzère, B.; Adam, S.; Joubert, L. Predicting Experimental Electrophilicities from Quantum and Topological Descriptors: A Machine Learning Approach. *J. Comput. Chem.* **2020**, *41*, 2124–2136.

(15) Eckhoff, M.; Diedrich, J. V.; Mücke, M.; Proppe, J. Quantitative Structure–Reactivity Relationships for Synthesis Planning: The Benzhydrylium Case. *J. Phys. Chem. A* **2024**, *128*, 343–354.

(16) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49*, 1691–1694.

(17) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. Electronegativity: The Density Functional Viewpoint. *J. Chem. Phys.* **1978**, *68*, 3801–3807.

(18) Mulliken, R. S. A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1934**, *2*, 782–793.

(19) Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.

(20) Parr, R. G.; v. Szentpály, L.; Liu, S. Electrophilicity Index. *J. Am. Chem. Soc.* **1999**, *121*, 1922–1924.

(21) Proppe, J.; Reiher, M. Reliable Estimation of Prediction Uncertainty for Physicochemical Property Models. *J. Chem. Theory Comput.* **2017**, *13*, 3297–3317.