Electronic Supplementary Information

Theoretical and machine learning models for reaction-barrier predictions: acrylate and methacrylate radical reactions

Makito Takagi,[a*] Tomomi Shimazaki,[a*] Osamu Kobayashi,[a] Takayoshi Ishimoto[a,b,c] and Masanori Tachikawa[a*]

[a] Quantum Chemistry Division, Yokohama City University, Seto 22-2, Kanazawa-ku, Yokohama 236-0027, Kanagawa, Japan.

[b] Smart Innovation Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan

[c] Division of Materials Model-Based Research, Digital Monozukuri (Manufacturing) Education and Research Center, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan

*Authors to whom correspondence should be addressed.

E-mail:

(Makito Takagi) mtakagi@yokohama-cu.ac.jp

(Tomomi Shimazaki) tshima@yokohama-cu.ac.jp

(Masanori Tachikawa) tachi@yokohama-cu.ac.jp

Electronic Supplementary Information

S1. ΔE_{TS} and ΔE_{prod} calculated by DFT and descriptors

The calculated ΔE_{TS} and ΔE_{prod} for all combinations of the 10 monomers, i.e., 100 reactions, are listed in **Table S1**, showing the descriptors for the random forest models (**Fig. 4(c)** in the main manuscript). For convenience, we also showed sequential numbers of reactants (see also **Fig. S1**).

Table S1. ΔE_{TS} [kcal/mol] and ΔE_{prod} [kcal/mol] calculated by DFT and descriptors for the random forest model; molecular weight [g/mol] and dummy parameter DP(*m*) to represent ACR or MA with *m* for specifying X or Y.

Combination of reactants	ΔE_{TS}	ΔE_{prod}	DP(X)	Molecular weight (X)	DP(Y)	Molecular weight(Y)
01-01	4.94	-17.30	0	72.0627	0	72.0627
01-02	4.26	-19.46	0	72.0627	1	86.0892
01-03	5.19	-17.00	0	72.0627	0	86.0892
01-04	4.56	-18.71	0	72.0627	1	100.1158
01-05	5.58	-16.54	0	72.0627	0	128.169
01-06	4.64	-18.48	0	72.0627	1	142.1956
01-07	5.16	-17.31	0	72.0627	0	156.136
01-08	4.25	-19.69	0	72.0627	1	170.1626
01-09	5.58	-16.38	0	72.0627	0	182.2594
01-10	4.35	-18.56	0	72.0627	1	196.286
02-01	6.17	-12.60	1	86.0892	0	72.0627
02-02	6.76	-13.11	1	86.0892	1	86.0892
02-03	7.94	-10.84	1	86.0892	0	86.0892
02-04	7.19	-12.50	1	86.0892	1	100.1158
02-05	7.15	-11.55	1	86.0892	0	128.169
02-06	7.01	-12.54	1	86.0892	1	142.1956

02-07	6.36	-12.53	1	86.0892	0	156.136
02-08	6.45	-13.57	1	86.0892	1	170.1626
02-09	6.95	-11.70	1	86.0892	0	182.2594
02-10	6.46	-13.03	1	86.0892	1	196.286
03-01	4.77	-17.55	0	86.0892	0	72.0627
03-02	4.61	-19.28	0	86.0892	1	86.0892
03-03	5.08	-17.22	0	86.0892	0	86.0892
03-04	5.01	-18.64	0	86.0892	1	100.1158
03-05	5.51	-16.74	0	86.0892	0	128.169
03-06	5.34	-18.04	0	86.0892	1	142.1956
03-07	5.05	-17.49	0	86.0892	0	156.136
03-08	4.99	-19.14	0	86.0892	1	170.1626
03-09	5.42	-16.65	0	86.0892	0	182.2594
03-10	4.40	-18.74	0	86.0892	1	196.286
04-01	5.80	-13.11	1	100.1158	0	72.0627
04-02	6.29	-13.79	1	100.1158	1	86.0892
04-03	8.28	-10.66	1	100.1158	0	86.0892
04-04	7.38	-12.52	1	100.1158	1	100.1158
04-05	7.08	-11.80	1	100.1158	0	128.169
04-06	7.57	-12.23	1	100.1158	1	142.1956
04-07	6.17	-12.86	1	100.1158	0	156.136
04-08	6.55	-13.68	1	100.1158	1	170.1626
04-09	6.85	-11.98	1	100.1158	0	182.2594
04-10	6.60	-13.13	1	100.1158	1	196.286
05-01	4.57	-17.74	0	128.169	0	72.0627
05-02	3.91	-19.98	0	128.169	1	86.0892
05-03	4.90	-17.38	0	128.169	0	86.0892

05-04	4.81	-18.84	0	128.169	1	100.1158
05-05	5.36	-16.26	0	128.169	0	128.169
05-06	4.88	-18.67	0	128.169	1	142.1956
05-07	4.88	-18.17	0	128.169	0	156.136
05-08	4.25	-20.41	0	128.169	1	170.1626
05-09	5.36	-16.07	0	128.169	0	182.2594
05-10	4.37	-18.09	0	128.169	1	196.286
06-01	5.46	-13.45	1	142.1956	0	72.0627
06-02	5.95	-14.12	1	142.1956	1	86.0892
06-03	5.80	-13.14	1	142.1956	0	86.0892
06-04	6.74	-13.16	1	142.1956	1	100.1158
06-05	6.48	-12.18	1	142.1956	0	128.169
06-06	7.25	-12.54	1	142.1956	1	142.1956
06-07	8.85	-10.64	1	142.1956	0	156.136
06-08	8.07	-12.57	1	142.1956	1	170.1626
06-09	6.45	-12.14	1	142.1956	0	182.2594
06-10	6.12	-13.33	1	142.1956	1	196.286
07-01	6.14	-15.08	0	156.136	0	72.0627
07-02	4.38	-17.62	0	156.136	1	86.0892
07-03	6.26	-14.95	0	156.136	0	86.0892
07-04	4.40	-17.28	0	156.136	1	100.1158
07-05	6.52	-13.77	0	156.136	0	128.169
07-06	4.48	-14.74	0	156.136	1	142.1956
07-07	6.38	-14.82	0	156.136	0	156.136
07-08	4.37	-15.68	0	156.136	1	170.1626
07-09	6.27	-13.81	0	156.136	0	182.2594
07-10	4.92	-13.74	0	156.136	1	196.286

08-01	8.54	-9.57	1	170.1626	0	72.0627
08-02	8.89	-10.34	1	170.1626	1	86.0892
08-03	8.89	-9.25	1	170.1626	0	86.0892
08-04	9.20	-9.87	1	170.1626	1	100.1158
08-05	8.78	-9.21	1	170.1626	0	128.169
08-06	9.18	-9.42	1	170.1626	1	142.1956
08-07	10.06	-8.44	1	170.1626	0	156.136
08-08	8.99	-11.09	1	170.1626	1	170.1626
08-09	8.17	-9.76	1	170.1626	0	182.2594
08-10	7.90	-10.84	1	170.1626	1	196.286
09-01	4.56	-17.82	0	182.2594	0	72.0627
09-02	3.90	-20.10	0	182.2594	1	86.0892
09-03	4.89	-17.49	0	182.2594	0	86.0892
09-04	4.86	-18.77	0	182.2594	1	100.1158
09-05	5.35	-16.56	0	182.2594	0	128.169
09-06	4.91	-19.16	0	182.2594	1	142.1956
09-07	4.89	-18.53	0	182.2594	0	156.136
09-08	4.26	-20.86	0	182.2594	1	170.1626
09-09	5.77	-15.91	0	182.2594	0	182.2594
09-10	4.40	-18.39	0	182.2594	1	196.286
10-01	5.36	-13.74	1	196.286	0	72.0627
10-02	6.46	-13.86	1	196.286	1	86.0892
10-03	5.71	-13.44	1	196.286	0	86.0892
10-04	6.83	-13.32	1	196.286	1	100.1158
10-05	6.45	-12.36	1	196.286	0	128.169
10-06	7.58	-12.70	1	196.286	1	142.1956
10-07	9.02	-10.76	1	196.286	0	156.136

10-08	8.01	-13.03	1	196.286	1	170.1626
10-09	6.36	-12.40	1	196.286	0	182.2594
10-10	6.03	-13.65	1	196.286	1	196.286

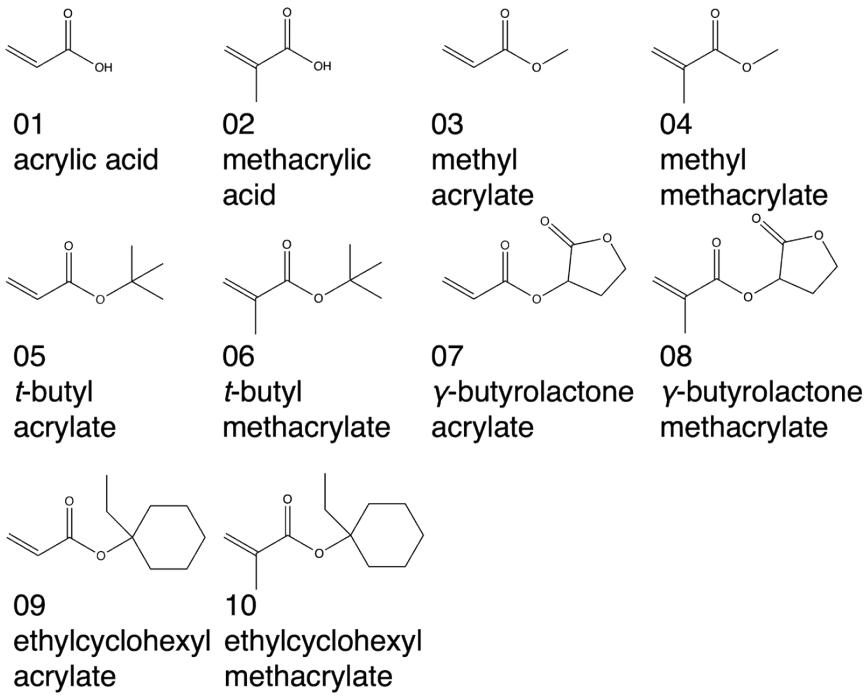

Fig. S1. The target reactants and sequential numbers.

S2. The BEP relationship between ΔE_{TS} and ΔE_{prod}

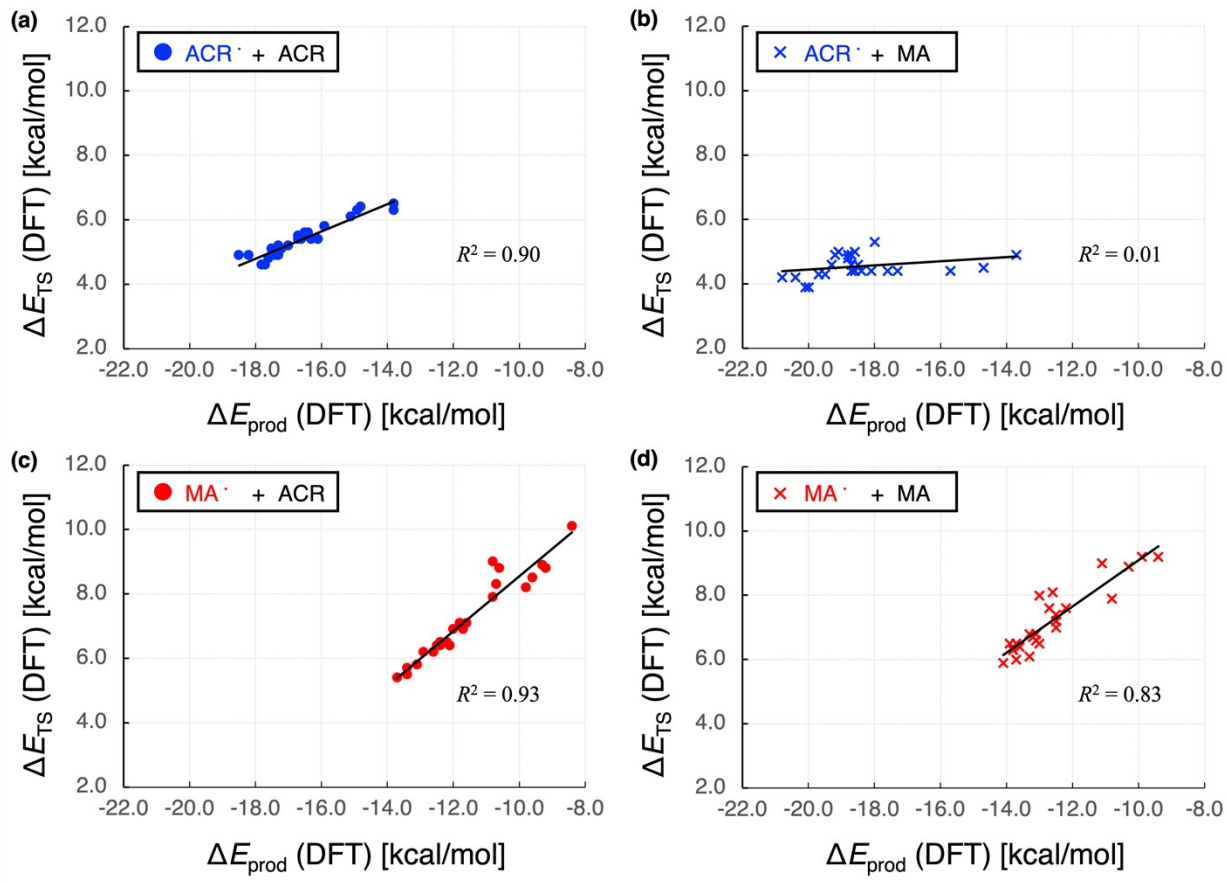

Fig. S2. Scatter plot of the ΔE_{TS} with respect to the DFT-obtained ΔE_{prod} for each categoly (a)ACR · + ACR (b)ACR · + MA (c)MA · + ACR (d)MA · + MA.

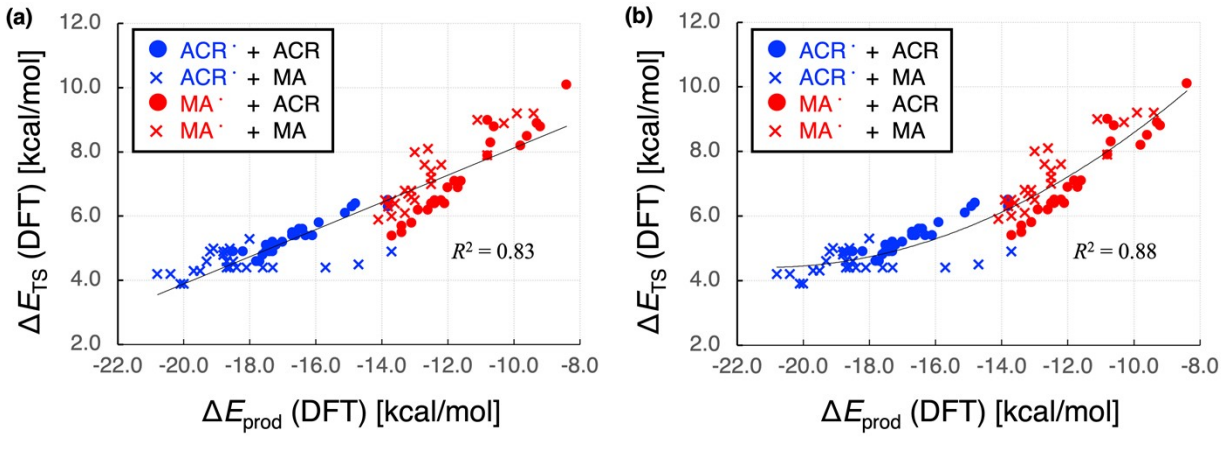

Fig. S3. Scatter plot of the ΔE_{TS} with respect to the DFT-obtained ΔE_{prod} using (a) the BEP approximation line and (b) the quadratic approximation curve. The blue and red colors represent ACR · and MA · for the reactant radical (X ·), respectively. Conversely, ● and × represent ACR and MA for the reactant monomer (Y), respectively.

S3. Random forest model of ΔE_{TS} (all descriptors from Open Babel and RDkit, DP(m))

We first constructed a machine-learning model with $R^2 = 0.84$ to predict ΔE_{TS} based on the random forest algorithm with a total of 50 descriptors, using several group contribution methods implemented in Open Babel and RDkit libralies, in addition to the dummy parameter DP(m) to represent ACR or MA. For each reactant species, 24 descriptors from group contribution methods are employed: molecular weight, exact mass, number of hydrogen bond donors (HBD), number of hydrogen acceptors (HBA), topological polar surface area (TPSA), partition coefficient (LogP), Molecular refractivity, Melting point, number of single bonds, number of double bonds, number of triple bonds, number of aromatic bonds, number of Rotatable Bonds (NumRotatableBonds), number of aromatic rings (NumAromaticRings), number of saturated rings (NumSaturatedRings), number of rings (RingCount), fraction of sp^3 hybridized carbons (FractionCSP3), quantitative estimate of drug-likeness (qed), Volume, Hansen solubility parameters (HSP-D, HSP-P, HSP-H), and DP(m). We showed the predicted ΔE_{TS} from the random forest model with respect to the calculated ΔE_{TS} by DFT in **Fig. S4**. We also constructed the simpler model described in the main manuscript selecting descriptors above (see also **Fig. 4(c)** in the main manuscript).

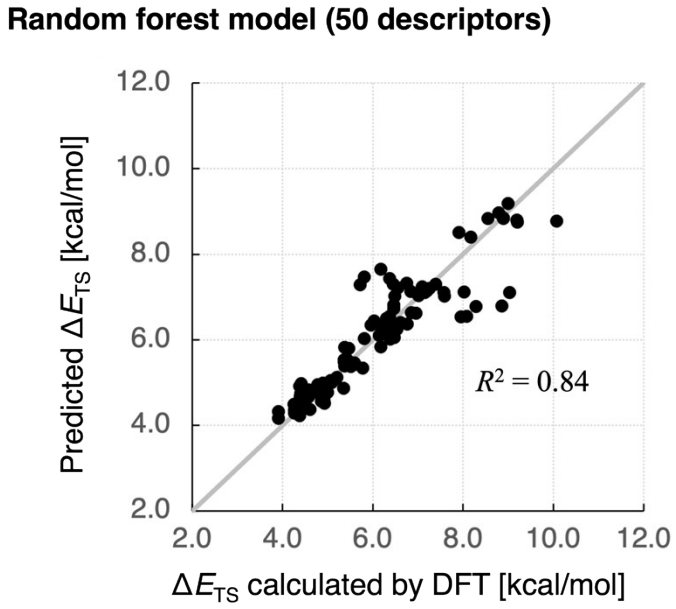

Fig. S4. Comparison between machine-leaning prediction (the random forest model with 50 descriptors) and DFT calculations on the energy barrier (ΔE_{TS}).

S4. Linear regression model for ΔE_{TS} (molecular weight, DP(m))

We constructed the simpler model (R^2 = 0.80) with only four descriptors, where the molecular weight and the DP(m) for each reactant species are employed (see also **Fig. 4(c)** in the main manuscript). In this section, we examined a linear regression model with the same four descriptors (**Fig. S5**). The R^2 value of this linear regression model is 0.63, which is not accurate compared with the random forest model, but roughly reproduces DFT results. Conversely, this linear regression model underestimates DFT results in the high ΔE_{TS} region. These suggest that nonlinear effects exist in the high ΔE_{TS} region. To construct better models over the linear model, we need to deal with nonlinear effects using some machine-leaning algorithm.

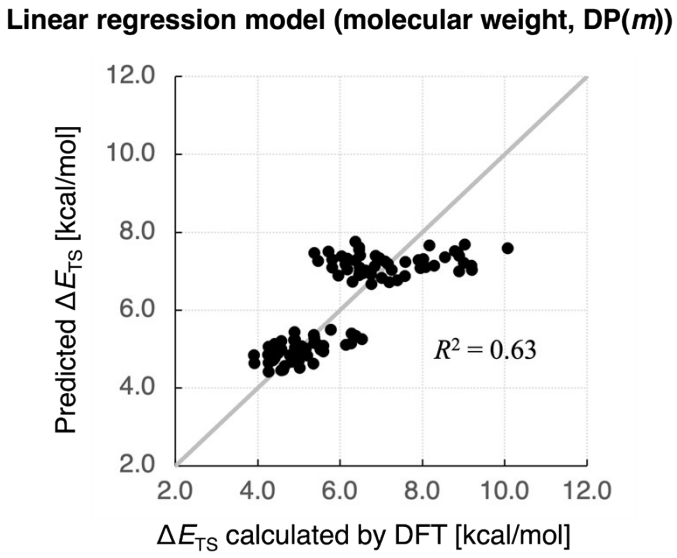

Fig. S5. Comparison between the linear regression model and DFT calculations. Here**,** only molecular weight and dummy parameters are used as descriptors.

S5. The relationship between ΔE_{TS} and molecular weight for each categoly

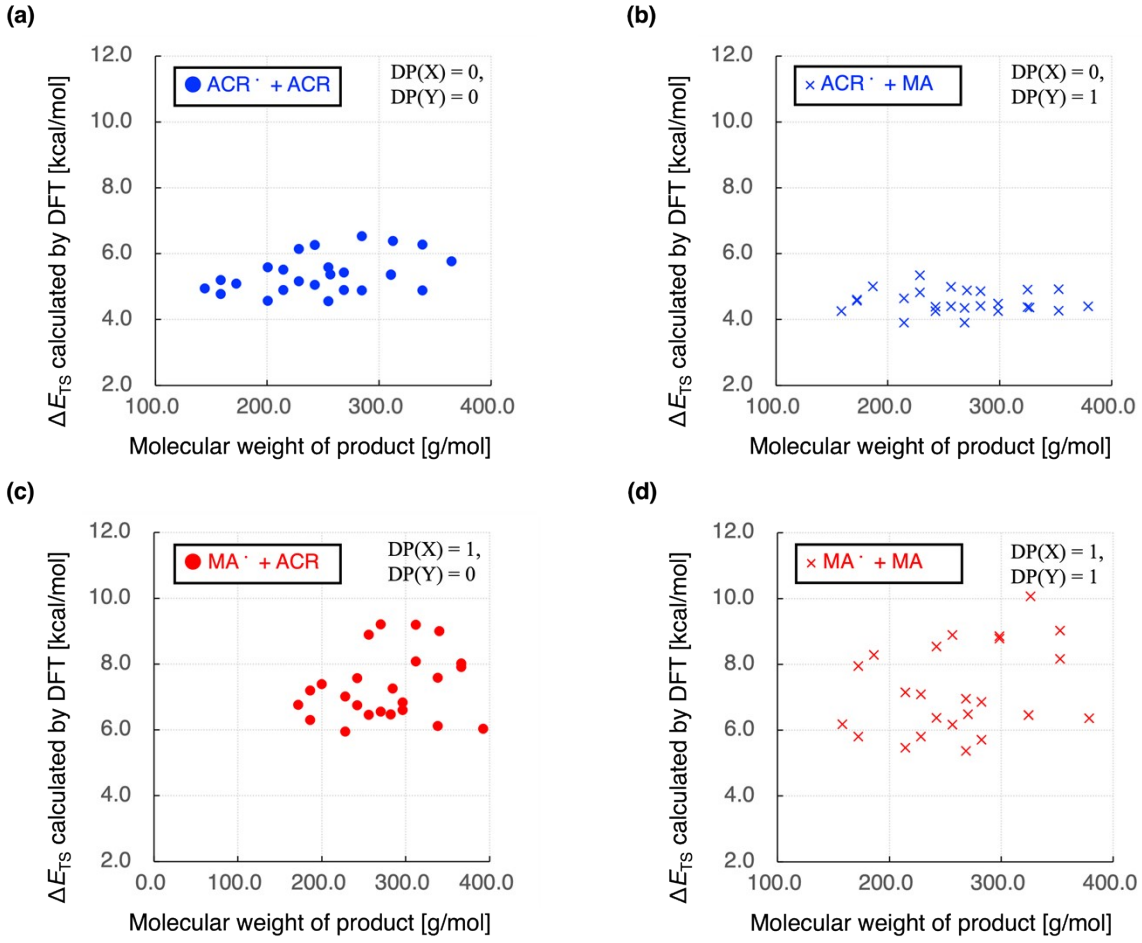

Fig. S6 Comparisons of ΔE_{TS} values predicted by DFT-based calculations and the molecular weight of product for each category. (a) ACR · + ACR (b) ACR · + MA (c) MA · + ACR (d) MA · + MA.

S6. Predictions of ΔE_{TS} based on threoretical regression models after estimations of ΔE_{prod} based on random forest model with only molecular weight and DP(m)) descriptors

In this section, we describe an collobration approach between theoretical regression and ML models, where ΔE_{prod} is first estimated based on RF-based model and then ΔE_{TS} is predicted using BEP and Marcus-like theoretical models (see also **Eq. (3)** and **(4)**). For the ΔE_{prod} estimation part, we constructed a RF-based model with $R^2 = 0.92$, using only four descriptors (molecular weight and the DP(m) for each reactant species (**Fig. S7 (a)**). Then, we predicted ΔE_{TS} based on theoretical regression models with ΔE_{prod}, where R^2 values of 0.75 and 0.76 are obtained from the BEP and Marcus-like models, respectively (**Fig. S7 (c)** and **(d)**.

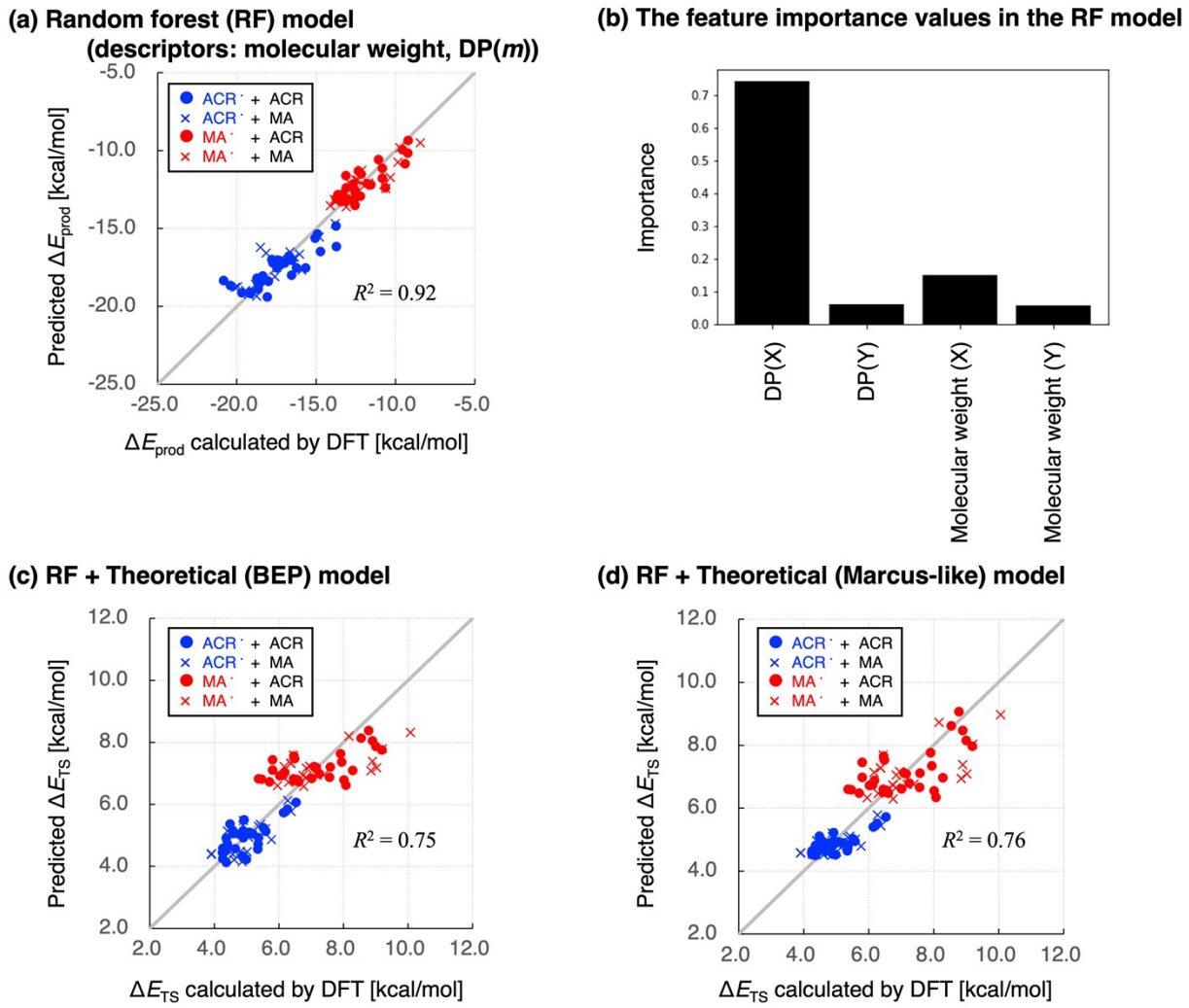

Fig. S7 (a) Comparisons of ΔE_{prod} values predicted by DFT-based calculations and the RF model using four descriptors; the dummy parameter (DP(m)) and molecular weight of each reactant. The blue and red colors represent ACR · and MA · for the reactant radical (X ·), respectively. Conversely, ● and × represent ACR and MA for the reactant monomer (Y), respectively. (b) Feature-importance values in the RF-based model. Comparisons of ΔE_{TS}s predicted by DFT-based calculations and the RF + theoretical models: (d) BEP model and (d) Marcus-like model.