Supporting Information

Unlocking XAT/HAT Selectivity: A Machine Learning-Guided Discovery of Governing Physicochemical Principles

Siqi Liu [a,*], Wenqiang Xu [b], Tian-Yu Sun [b,c], Longjie Huang [a], Yun-Dong Wu [b,d,*], Xinhao Zhang [b,c,*]

a. College of Chemistry and Materials Science, Shanghai Normal University, Shanghai, 200234, China.

b. Key Laboratory of Computational Chemistry and Drug Design, State Key Laboratory of Chemical Oncogenomics, Shenzhen Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen 518055, China.

c. Institute of Chemical Biology, Shenzhen Bay Laboratory, Shenzhen 518132, China.

d. College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

*. Corresponding authors

Contents

1. Details of the dataset used in model training

1.1 Detailed substrates

The distribution of organic halides within our dataset is illustrated in Figure S1, with corresponding optimization output files hosted on our GitHub repository (https://github.com/liusqchem/machine-learning.git). Sourced from the *i*BonD 2.0 database, the dataset includes 113 distinct C–H bonds and 61 C–X bonds (Figure S1). This selection was designed with a dual purpose. First, it ensures a broad representation of chemical and structural diversity, encompassing alkyl, cycloalkyl, (hetero)benzylic, aldehydic, and (hetero)aryl C–H and C–X motifs. The calculated bond dissociation energies (***BDE***s) range from 80–115 kcal/mol for C–H and 40–130 kcal/mol for C–X bonds (Figure S3), confirming that the dataset covers a wide and representative chemical space. Second, these ***BDE*** values serve as a benchmark to validate the reliability of our quantum chemical calculations (Figure S2). High-quality, chemically relevant training data is essential, as it directly governs the accuracy of the resulting machine learning (ML) models. To model radical abstraction, we considered a variety of abstracting agents, including C-, B-, N-, Si-, and S-centered radicals. Specifically, we selected 38 carbon-centered radicals—ranging from alkyl and α-heteroatom-substituted species to captodative and sp²-hybridized radicals—leading to a comprehensive set of 4,294 HAT and 2,318 XAT reactions. Furthermore, to assess the generalizability of the H/X selectivity patterns, we incorporated an additional array of experimental boryl, nitrogen-centered, silyl, and thiol radicals.

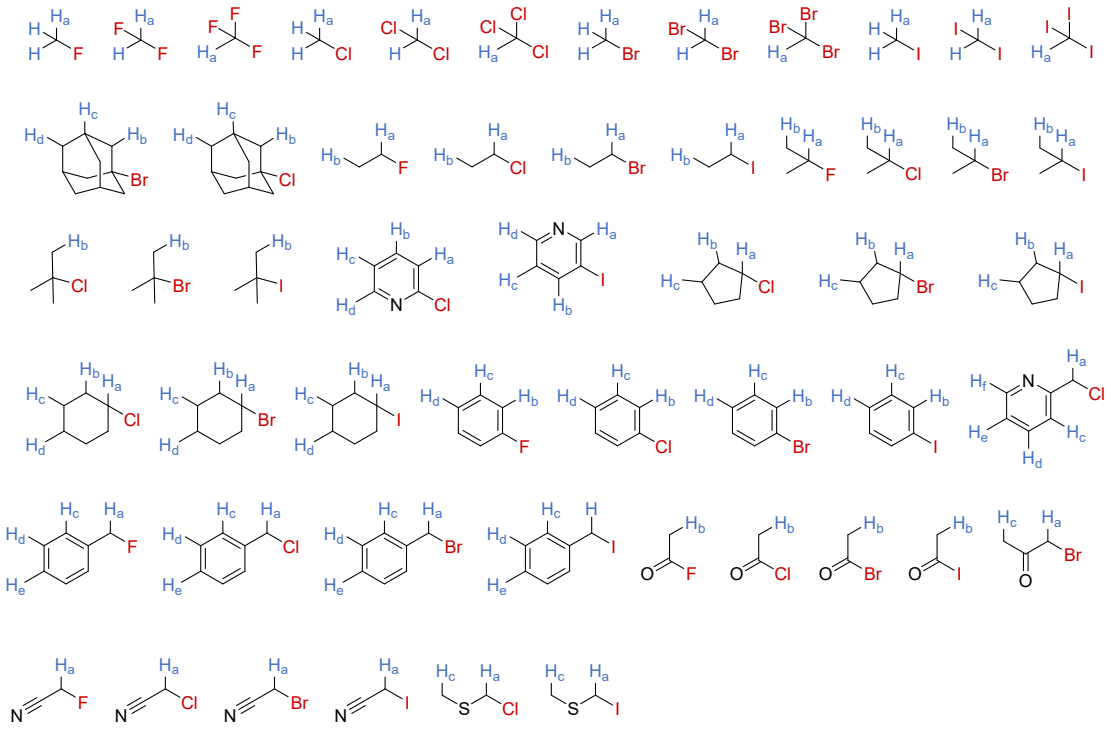


Figure S1. Compound details of organic halides

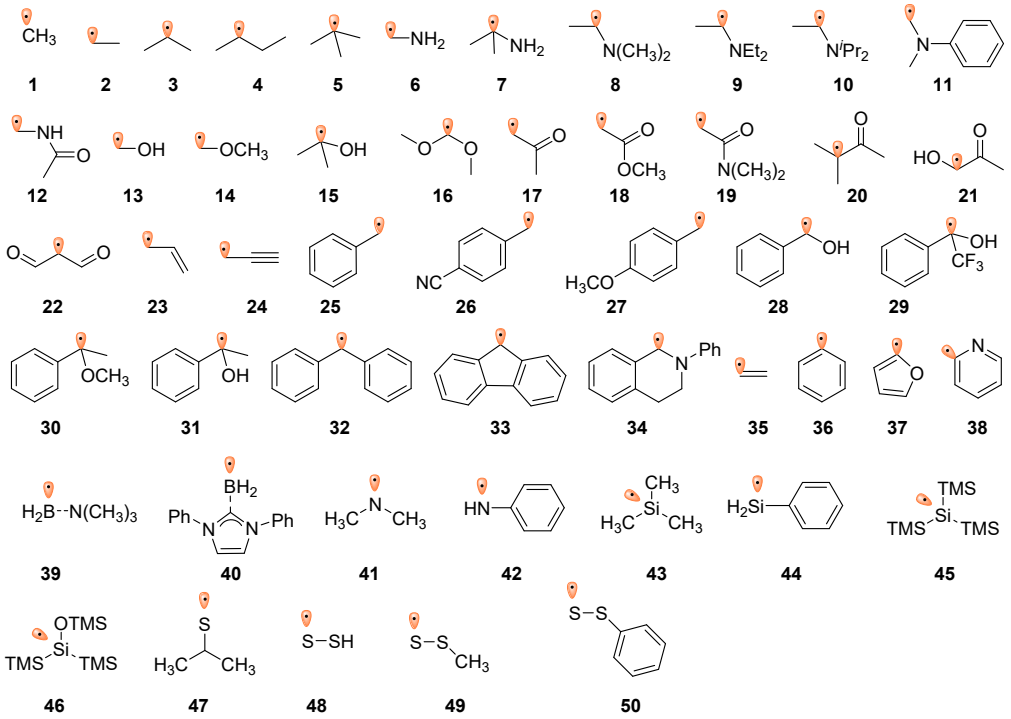


Figure S2. Compound details of radicals

1.2 Benchmark of the theoretical levels for geometry optimization

All DFT calculations were conducted with Gaussian 16[2]. Structural optimizations were performed at the (U)B3LYP-D3(BJ) [3-6] / Lanl2dz+d with ECP for I[7, 8], and (U)B3LYP-D3(BJ)/ 6-31+G(d,p) [9] for other atoms in the gas phase. Single point calculations were conducted at the (U)B3LYP-D3(BJ)/6-311++G(3df,3pd) [9, 10]-SDD[11, 12]. The calculated $\boldsymbol{BDE_{C\text{-}H}}$ and $\boldsymbol{BDE_{C\text{-}X}}$ and experimental $\boldsymbol{BDE_{C\text{-}H}}$ and $\boldsymbol{BDE_{C\text{-}X}}$ of organic halides are shown in Table S1.

Table S1. The calculated $\boldsymbol{BDE_{C\text{-}H}}$ and $\boldsymbol{BDE_{C\text{-}X}}$ and experimental $\boldsymbol{BDE_{C\text{-}H}}$ and $\boldsymbol{BDE_{C\text{-}X}}$ of organic halides (kcal/mol).

Bond type		$BDE_{exp.}$	$BDE_{calc.}$
$BDE_{C\text{-}H}$	CHF_3	106.4	103.9
	CH_2F_2	103.2	99.1
	CH_3F	101.3	99.1
	$CHCl_3$	93.8	90.9
	CH_2Cl_2	95.7	93.8
	CH_3Cl	100.1	97.5
	$CHBr_3$	95.4	91.0
	CH_2Br_2	98.6	95.2
	CH_3Br	102.1	98.9
	CHI_3	101.1	93.0
	CH_2I_2	103.0	97.6
	CH_3I	103.5	100.9
	C_2H_5Br (H_b)	99.2	96.3
	C_2H_5Cl (H_a)	97.2	95.0

	C_2H_5Cl (H_b)	101.1	98.8
	C_2H_5F (H_a)	98.2	96.6
	C_2H_5F (H_b)	103.6	101.3
	C_7H_7Cl (H_a)	84.2	84.2
	$CNCH_2F$	88.6	88.6
BDE_{C-X}	CHF_3	127.6	123.0
	CH_2F_2	118.6	116.7
	CH_3F	110.0	109.2
	$CHCl_3$	74.4	70.7
	CH_2Cl_2	80.8	76.1
	CH_3Cl	83.7	81.5
	$CHBr_3$	65.7	59.2
	CH_2Br_2	66.0	65.2
	CH_3Br	70.3	70.6
	CHI_3	52.7	48.1
	CH_2I_2	51.8	53.5
	CH_3I	56.9	58.3
	$C_{10}H_{15}Br$	70.8	74.8
	$C_{10}H_{15}Cl$	90.4	86.9
	C_2H_5F	111.7	111.0
	C_2H_5Cl	84.2	81.7
	C_2H_5Br	69.6	70.3
	C_2H_5I	56.3	57.8
	C_3H_7F	110.1	112.2
	C_3H_7Cl	84.6	81.3
	C_3H_7Br	70.5	69.4
	C_3H_7I	55.7	56.8
	C_5H_4NCl	90.5	90.7
	C_5H_4NI	66.6	68.7
	C_5H_9Cl	82.4	79.8
	C_5H_9Br	69.8	68.1
	C_5H_9I	55.4	55.8
	C6H11F	117.5	112.8
	$C_6H_{11}Cl$	86.1	82.7
	$C_6H_{11}Br$	69.9	70.8
	$C_6H_{11}I$	55.5	58.4
	C_6H_5F	127.2	125.8
	C_6H_5Cl	97.1	95.0
	C_6H_5Br	80.4	82.1
	C_6H_5I	67.4	68.9
	C_7H_7F	98.7	96.9
	C_7H_7Cl	71.7	68.9
	C_7H_7Br	57.2	58.1

	C_7H_7I	44.9	46.9
	CH_3COF	122.3	119.3
	CH_3COCl	84.6	82.3
	CH_3COBr	69.8	70.2
	CH_3COI	53.3	56.7
	$CNCH_2Cl$	66.4	66.8
	$CNCH_2Br$	56.8	56.3
	$CNCH_2I$	44.7	45.2
	SC_2H_5Cl	72	73.4
	SC_2H_5I	51.8	50.8
	CH_3COCH_2Br	62.5	60.0
	C_6H_6NCl	69.9	69.1
	C_4H_9Cl	84.1	80.5
	C_4H_9Br	70	68.3
	C_4H_9I	53.4	55.6
Mean absolute error		/	2.3

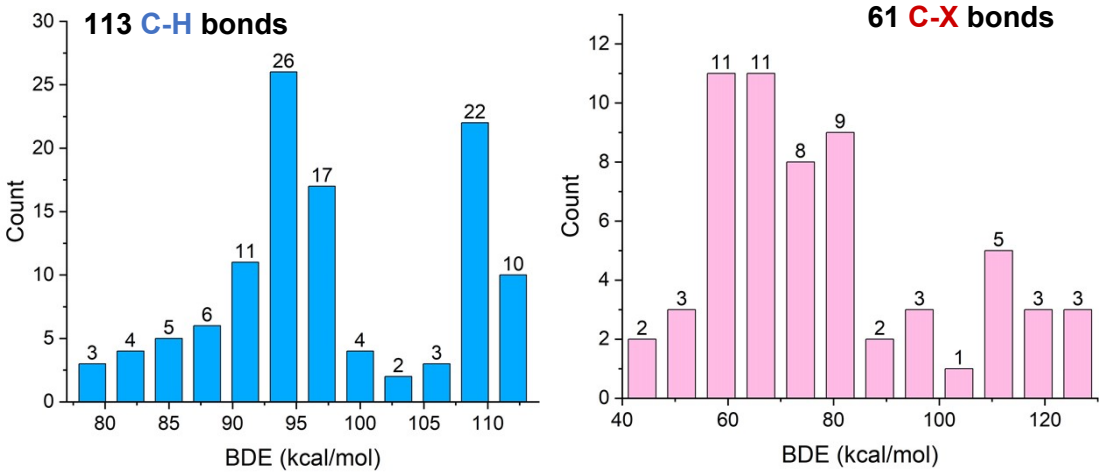


Figure S3. Distribution of calculated ***BDE***s(kcal/mol).

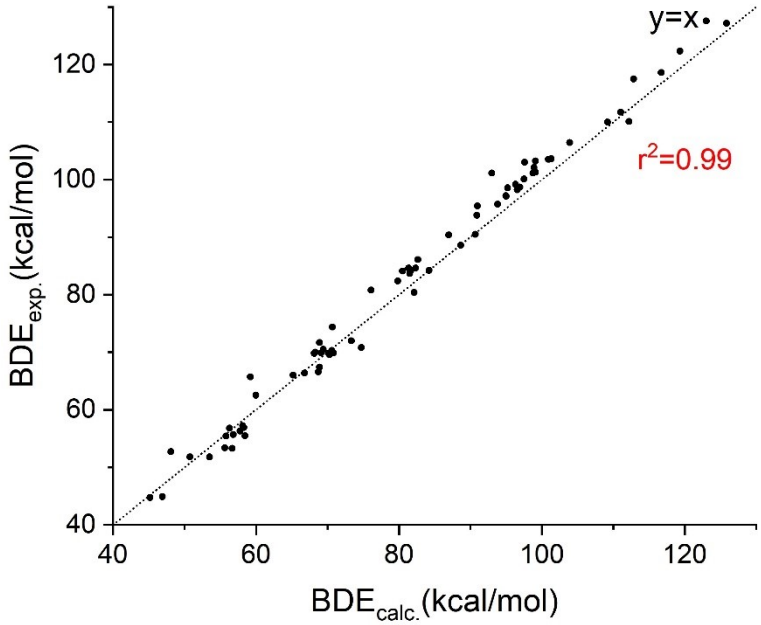


Figure S4. Correlations between the $\boldsymbol{BDE_{calc.}}$ with $\boldsymbol{BDE_{exp.}}$ (kcal/mol).

1.3 Calculated *BDE* and *BO*

Table S2. The calculated $\boldsymbol{BDE_{C\text{-}H}}$ and $\boldsymbol{BO_{C\text{-}H.}}$

Bond type	BDE	BO	Bond type	BDE	BO	Bond type	BDE	BO
$C_{10}H_{15}Br$ (H_b)	99.178	0.923	C_5H_9Cl (H_c)	94.0807	0.9342	C_7H_7Cl (H_c)	111.8924	0.9261
$C_{10}H_{15}Br$ (H_c)	97.3249	0.9142	C_5H_9I (H_a)	94.085	0.9218	C_7H_7Cl (H_d)	111.5944	0.9277
$C_{10}H_{15}Br$ (H_d)	97.9949	0.9305	C_5H_9I (H_b)	79.258	0.9098	C_7H_7Cl (H_e)	111.7443	0.9277
$C_{10}H_{15}Cl$ (H_b)	99.1365	0.9242	C_5H_9I (H_c)	94.251	0.9332	C_7H_7F (H_a)	84.3371	0.9404
$C_{10}H_{15}Cl$ (H_c)	97.4937	0.9146	$C_6H_{11}Br$ (H_a)	94.4278	0.9125	C_7H_7F (H_c)	111.7713	0.9262
$C_{10}H_{15}Cl$ (H_d)	98.0437	0.9305	$C_6H_{11}Br$ (H_b)	97.2602	0.9222	C_7H_7F (H_d)	111.5955	0.9277
C_2H_5Br (H_a)	96.1901	0.9313	$C_6H_{11}Br$ (H_c)	96.1702	0.9291	C_7H_7F (H_e)	111.6252	0.9278
C_2H_5Br (H_b)	96.2736	0.9255	$C_6H_{11}Br$ (H_d)	96.3728	0.9286	C_7H_7I (H_a)	88.549	0.9296
C_2H_5Cl (H_a)	94.975	0.9293	$C_6H_{11}Cl$ (H_a)	93.5252	0.9108	C_7H_7I (H_c)	111.950	0.9265
C_2H_5Cl (H_b)	98.7753	0.9298	$C_6H_{11}Cl$ (H_b)	97.5474	0.9229	C_7H_7I (H_d)	111.613	0.9276
C_2H_5F (H_a)	96.5698	0.9424	$C_6H_{11}Cl$ (H_c)	96.1785	0.929	C_7H_7I (H_e)	111.822	0.9278
C_2H_5F (H_b)	101.3099	0.9371	$C_6H_{11}Cl$ (H_d)	96.3937	0.9287	CH_2Br_2	95.2	0.9252
C_2H_5I (H_a)	97.78	0.9337	$C_6H_{11}I$ (H_a)	95.650	0.9158	CH_2Cl_2	93.8	0.9201
C_2H_5I (H_b)	87.4308	0.9229	$C_6H_{11}I$ (H_b)	80.73	0.9222	CH_2F_2	99.1	0.9452
C_3H_7Br (H_a)	94.1239	0.9158	$C_6H_{11}I$ (H_c)	96.118	0.9286	CH_2I_2	97.6	0.9313
C_3H_7Br (H_b)	96.1178	0.9254	$C_6H_{11}I$ (H_d)	96.366	0.9286	CH_3Br	98.9	0.9437
C_3H_7Cl (H_a)	93.1873	0.9139	C_6H_5Br (H_b)	112.5727	0.9193	CH_3Cl	97.5	0.9414
C_3H_7Cl (H_b)	98.7035	0.9298	C_6H_5Br (H_c)	111.3946	0.9264	CH_3COBr	101.9148	0.9113
C_3H_7F (H_a)	94.6661	0.9259	C_6H_5Br (H_d)	112.2021	0.9268	CH_3COCH_2Br (H_a)	87.2073	0.9092
C_3H_7F (H_b)	101.2961	0.9378	C_6H_5Cl (H_b)	113.1296	0.92	CH_3COCH_2Br (H_c)	94.4225	0.9323
C_3H_7I (H_a)	95.419	0.9189	C_6H_5Cl (H_c)	111.5567	0.9266	CH_3COCl	95.3537	0.9125
C_3H_7I (H_b)	83.83	0.9221	C_6H_5Cl (H_d)	112.2784	0.927	CH_3COF	96.0657	0.9161
C_4H_9Br (H_b)	90.3217	0.9232	C_6H_5F (H_b)	114.0035	0.92	CH_3COI	94.956	0.9104
C_4H_9Cl (H_b)	98.6455	0.9281	C_6H_5F (H_c)	111.687	0.9276	CH_3F	99.1	0.9564

C_4H_9I (H_b)	81.26	0.9198	C_6H_5F (H_d)	112.5339	0.9269	CH_3I	100.9	0.9462
C_5H_4NCl (H_a)	111.1449	0.9252	C_6H_5I (H_b)	111.79	0.9194	$CHBr_3$	91	0.9063
C_5H_4NCl (H_b)	113.1372	0.9238	C_6H_5I (H_c)	111.273	0.9258	$CHCl_3$	90.9	0.8966
C_5H_4NCl (H_c)	106.6584	0.9238	C_6H_5I (H_d)	112.110	0.9268	CHF_3	103.9	0.9306
C_5H_4NCl (H_d)	113.097	1.0471	C_6H_6NCl (H_a)	84.296	0.9226	CHI_3	93	0.9216
C_5H_4NI (H_a)	106.541	0.9182	C_6H_6NCl (H_c)	111.9215	0.9212	$CNCH_2Br$	87.8256	0.8966
C_5H_4NI (H_b)	111.082	0.923	C_6H_6NCl (H_d)	110.8864	0.9266	$CNCH_2Cl$	86.6625	0.8946
C_5H_4NI (H_c)	111.5875	0.9229	C_6H_6NCl (H_e)	112.4097	0.9247	$CNCH_2F$	88.6433	0.9113
C_5H_4NI (H_d)	106.4375	1.0058	C_6H_6NCl (H_f)	105.9481	0.9242	$CNCH_2I$	89.407	0.8989
C_5H_9Br (H_a)	92.985	0.9183	C_7H_7Br (H_a)	86.1632	0.9163	SC_2H_5Cl (H_a)	92.1362	0.9178
C_5H_9Br (H_b)	88.0989	0.9131	C_7H_7Br (H_c)	111.9483	0.9143	SC_2H_5Cl (H_c)	93.3693	0.93
C_5H_9Br (H_c)	94.1498	0.9337	C_7H_7Br (H_d)	111.6049	0.928	SC_2H_5I (H_a)	94.619	0.9224
C_5H_9Cl (H_a)	92.0736	0.9168	C_7H_7Br (H_e)	111.7808	0.9278	SC_2H_5I (H_c)	93.145	0.9264
C_5H_9Cl (H_b)	91.9867	0.9175	C_7H_7Cl (H_a)	84.2431	0.9258			

Table S3. The calculated $\boldsymbol{BDE_{C\text{-}X}}$ and $\boldsymbol{BO_{C\text{-}X}}$.

Bond type	BDE	BO	Bond type	BDE	BO	Bond type	BDE	BO
$C_{10}H_{15}F$	118.106073	0.7914	C_6H_5Cl	94.99364313	1.056	SC_2H_5Br	62.585915	0.946
C_2H_5F	111.0421986	0.826	C_6H_6NCl	69.13774169	0.9827	$CHBr_3$	59.2	1.0145
C_3H_7F	112.2049737	0.8025	C_7H_7Cl	68.88171782	0.9701	CH_2Br_2	65.2	1.0259
C_5H_4NF	124.9703995	0.8731	CH_3COCl	82.33426648	0.954	CH_3Br	70.6	1.0235
C_5H_9F	110.2854221	0.7992	$CNCH_2Cl$	66.79650375	1.0045	$C_{10}H_{15}I$	62.939	0.9305
$C_6H_{11}F$	112.8306007	0.8078	SC_2H_5Cl	73.35711557	0.9714	C_2H_5I	57.77856472	0.9828
C_6H_5F	125.8432652	0.8746	$CHCl_3$	70.7	1.035	C_3H_7I	56.8322804	0.9482
C_7H_7F	96.94519767	0.8125	CH_2Cl_2	76.1	1.0373	C_4H_9I	55.6368748	0.9117
CH_3COF	119.3416392	0.8287	CH_3Cl	81.5	1.0268	C_5H_4NI	68.69974006	1.0058
$CNCH_2F$	95.09027959	0.8549	$C_{10}H_{15}Br$	74.7658744	0.9466	C_5H_9I	55.81634252	0.9437
SC_2H_5F	102.8230792	0.8169	C_2H_5Br	70.28733909	0.9939	$C_6H_{11}I$	58.44999989	0.9549
CHF_3	123	0.8919	C_3H_7Br	69.40380572	0.96	C_6H_5I	68.88171781	1.0091
CH_2F_2	116.7	0.8711	C_4H_9Br	68.25169828	0.927	C_7H_7I	46.91574777	0.9371
CH_3F	109.2	0.8488	C_5H_4NBr	77.52628869	1.0325	CH_3COI	56.67226547	0.892
$C_{10}H_{15}Cl$	86.94834383	0.9588	C_5H_9Br	68.13937408	0.9559	$CNCH_2I$	45.16374124	0.9489
C_2H_5Cl	81.68542165	0.9998	$C_6H_{11}Br$	70.83766493	0.9683	SC_2H_5I	50.75171334	0.9195
C_3H_7Cl	81.34154645	0.9695	C_6H_5Br	82.10961808	1.0441	CHI_3	48.1	0.9694
C_4H_9Cl	80.50821383	0.9411	C_7H_7Br	58.11992989	1.0111	CH_2I_2	53.5	0.997
C_5H_4NCl	90.68139784	1.0471	CH_3COBr	70.16685727	0.9272	CH_3I	58.3	1.0115
C_5H_9Cl	79.84932886	0.9649	CH_3COCH_2Br	59.9836331	1.0223			
$C_6H_{11}Cl$	82.66182643	0.9759	$CNCH_2Br$	56.30893747	0.9844			

2. Details of machine learning

2.1 Generation of transition states geometries and corresponding barriers

The workflow for DFT data generation is detailed in Figure S5. Initially, the organic halides

and radicals were optimized to obtain the equilibrium structures and ground-state energies of the reactants. The optimized geometries served as the basis for constructing initial transition state (TS) guesses. Specifically, Python scripts were employed to automatically assemble fragments of the organic halides and radicals into starting TS geometries for each target reaction. These guesses were then subjected to TS optimization. To ensure the localization of the correct HAT and XAT transition states, all output files were automatically validated; instances of optimization failure were rectified through manual adjustment of the guess geometries. Subsequent single-point energy calculations were performed on all verified TS structures. The automation scripts for geometry generation and output verification are hosted on our GitHub repository (https://github.com/liusqchem/machine-learning.git). The H/X selectivity of the carbon radicals is quantified by the relative activation barrier $\boldsymbol{\Delta\Delta G^{\ddagger}}$, defined as the difference between the lowest-lying HAT and XAT transition states ($\boldsymbol{\Delta G^{\ddagger}_{H} - \Delta G^{\ddagger}_{X}}$) for each organic halide substrate.

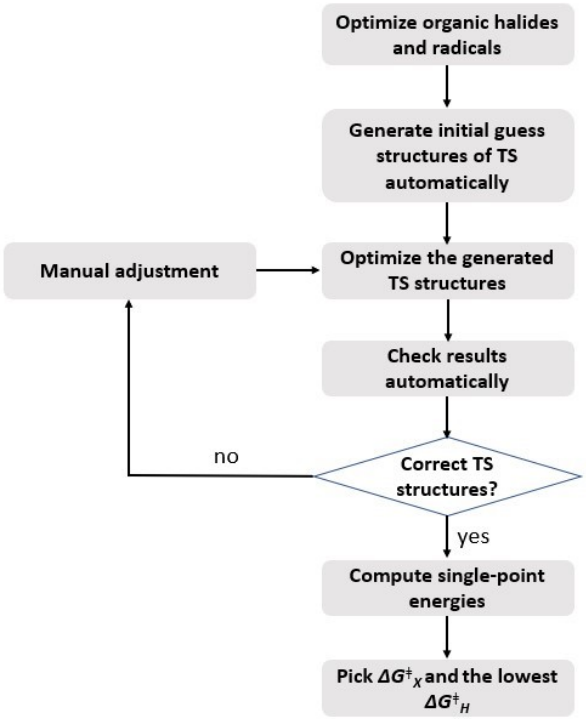


Figure S5. Flow chart of the DFT-computed data generation

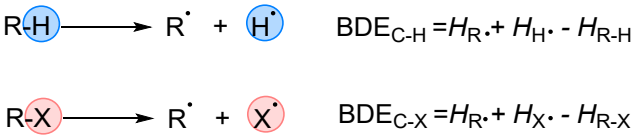
2.2 Details and calculations of descriptors

We initially select 18 physical organic descriptors categorized into eight groups: frontier molecular orbital (***FMO***) energies, atom charges, Wiberg bond orders (***BO***), bond dissociation energies (***BDE***), reaction free energies ($\boldsymbol{\Delta G}$), spin densities, ionization potentials (***IP***) and electrostatic potentials (***ESP***). These descriptors were strategically chosen to capture both the thermodynamic and polar effects governing atom transfer reactions. Thermodynamic effects are represented by three categories comprising eight descriptors: the ***BDE***s of the breaking and forming C-H/C-X bonds (4 descriptors), the Wiberg bond orders of the scissile C-H and C-X bonds (2 descriptors), and reaction energies for the HAT ($\boldsymbol{\Delta G_H}$) and XAT ($\boldsymbol{\Delta G_X}$) pathways (2 descriptors). Polar effects are encoded by the remaining ten descriptors across five categories: the spin density of carbon radical center (C_1), Natural Population Analysis (NPA) charges (encompassing the radical (C_1), the abstracted H/X atoms, and their adjacent carbon atom (C_2, C_3); 5 descriptors), ***FMO***

energies of the carbon radicals (E(LUMO) and E(SOMO) (i.e. E(HOMO)); 2 descriptors), and the ***ESP*** and ***IP*** of the carbon radicals (1 descriptor each).

2.2.1 Bond dissociation energy (***BDE***) and bond order (***BO***)

Tables S2 and S3 provide comprehensive computational details for the bond dissociation energies (***BDE***) of the scissile and nascent C–H/C–X bonds in both reactants and products. $\boldsymbol{BDE_{C\text{-}H}}$ and $\boldsymbol{BDE_{C\text{-}X}}$ were calculated as:



Additionally, bond order descriptors were derived from Wiberg bond indexes[13]. Our analysis explicitly incorporates the $\boldsymbol{BO_{C\text{-}H}}$ and $\boldsymbol{BO_{C\text{-}X}}$ values for each potential reacting carbon center.

2.2.2 Frontier orbital energy and ionization potentials

Table S4 includes the calculation details of the **FMO** energy-based descriptors. The SOMO(HOMO) and LUMO energies for radicals were obtained from single-point energy calculation. The calculated gas-phase ionization potentials (eV) were computed as indicated in Eq. (1), -0.03766 eV is the energy value of a free electron at 298K.

$$IP\ (eV) = (H_{radical\ cation} - H_{radical}) - 0.03766 \qquad \text{Eq. (1)}$$

2.2.3 Spin, atomic charge and electrostatic potentials

Natural bond orbital (***NBO***) analysis was conducted to obtain the atomic charge information. Charge (C_1) in Table S4 refers to the NPA charge[14] of radical-center carbon atom. Charge (H), charge (C_2), charge(X) and charge(C_3) in Table S5 refer to the NPA charge of abstracted H atom, the adjacent carbon atom(C_2), abstracted halogen atom and adjacent carbon atom(C_3). Spin (C_1) in Table S4 refers to Mulliken atomic spin population[15] of radical-center atom. Electrostatic potentials(***ESP***) of radicals in Table S4 are calculated using Multiwfn[16].

2.2.4 Database for descriptors

Table S4. Phys-org descriptors for radicals.

Radical	Spin (C_1)	Charge (C_1)	IP	E(SOMO)	E(LUMO)	ESP
1	1.083	-0.468	9.8	-0.23781	-0.08075	3.19233
2	1.057	-0.272	8.1	-0.20781	-0.06085	2.55279
3	1.013	-0.089	7.3	-0.18834	-0.04861	2.28046
4	1.04	-0.077	7.1	-0.18846	-0.05038	2.1719
5	0.967	0.071	6.7	-0.17686	-0.0426	2.77247
6	0.883	-0.253	6.3	-0.15677	-0.0285	4.3306
7	0.905	-0.032	5.1	-0.13958	-0.03085	2.61164
8	0.859	-0.032	4.9	-0.13397	-0.02748	3.04769
9	0.894	-0.045	4.7	-0.12858	-0.01797	1.50218
10	0.845	-0.247	5.6	-0.1504	-0.05433	3.4703
11	0.88	-0.21	6.8	-0.18295	-0.07161	3.66203
12	0.857	0.148	5.2	-0.14351	-0.02473	3.5872
13	0.931	-0.061	7.6	-0.18726	-0.04559	4.45321
14	0.924	-0.063	7.0	-0.18146	-0.04818	3.90981
15	0.916	0.317	6.1	-0.16612	-0.03462	3.85535

16	0.875	0.388	5.6	-0.17078	-0.0413	3.18867
17	0.882	-0.239	8.3	-0.27007	-0.13799	3.75155
18	0.947	-0.253	8.4	-0.26757	-0.13399	3.96379
19	0.964	-0.267	8.4	-0.24525	-0.11988	3.69976
20	0.773	0.093	7.6	-0.2268	-0.11256	2.74079
21	0.664	0.168	7.9	-0.22052	-0.10516	4.26836
22	0.747	-0.129	9.6	-0.2924	-0.19318	4.62418
23	0.688	-0.31	8.0	-0.20653	-0.09018	3.86071
24	0.798	-0.351	8.6	-0.2233	-0.10521	3.78623
25	0.797	-0.284	7.1	-0.19203	-0.09497	3.79475
26	0.759	-0.259	7.6	-0.2173	-0.12716	3.79665
27	0.797	-0.312	6.3	-0.17423	-0.08229	3.8505
28	0.647	0.187	6.3	-0.1658	-0.07376	4.50266
29	0.633	0.256	6.8	-0.19375	-0.10072	3.74628
30	0.625	0.371	5.7	-0.15627	-0.06814	3.42916
31	0.597	0.361	6.0	-0.15868	-0.06929	3.61018
32	0.652	-0.11	6.3	-0.17965	-0.10143	4.1306
33	0.629	-0.077	6.8	-0.19498	-0.11999	4.70461
34	0.543	0.022	5.3	-0.14707	-0.07331	3.08519
35	1.055	-0.054	8.5	-0.23292	-0.07954	14.11783
36	1.061	0.17	8.1	-0.25035	-0.09384	15.87553
37	1.124	0.463	9.0	-0.25046	-0.13057	17.81607
38	0.995	0.315	7.7	-0.22646	-0.08468	16.51412
39	1.057	-0.068	5.1	-0.12162	-0.02123	3.35522
40	0.409	0.036	5.2	-0.14153	-0.07161	3.06249
41	0.937	-0.292	8.7	-0.25475	-0.09153	13.91240
42	0.633	-0.521	7.9	-0.22309	-0.12405	14.48714
43	0.998	1.263	6.1	-0.18505	-0.07384	9.48910
44	0.917	0.82	6.6	-0.19973	-0.10169	10.83672
45	0.918	-0.095	6.1	-0.18941	-0.09742	6.48321
46	0.894	0.713	5.5	-0.17703	-0.08298	8.47136
47	0.981	0.108	6.5	-0.24607	-0.15358	31.09472
48	0.705	-0.135	9.3	-0.25454	-0.15118	40.49280
49	0.646	-0.184	8.6	-0.23816	-0.13939	38.27280
50	0.566	-0.203	7.8	-0.22791	-0.14265	34.90718

Table S5. Charge for organic halides.

Atom position	$C_2(C_3)$	H(X)	Atom position	$C_2(C_3)$	H(X)	Atom position	$C_2(C_3)$	H(X)
$C_{10}H_{15}Br$ (H_b)	-0.404	0.209	C_5H_9I (H_a)	-0.296	0.209	C_7H_7I (H_a)	-0.482	0.216
$C_{10}H_{15}Br$ (H_c)	-0.218	0.203	C_5H_9I (H_b)	-0.41	0.216	C_7H_7I (H_c)	-0.182	0.204
$C_{10}H_{15}Br$ (H_d)	-0.373	0.198	C_5H_9I (H_c)	-0.39	0.204	C_7H_7I (H_d)	-0.194	0.205
$C_{10}H_{15}Br$ (Br)	-0.04	-0.031	C_5H_9I (I)	-0.296	0.042	C_7H_7I (H_e)	-0.194	0.205
$C_{10}H_{15}Cl$ (H_b)	-0.399	0.209	$C_6H_{11}Br$ (H_a)	-0.21	0.195	C_7H_7I (I)	-0.482	0.058

$C_{10}H_{15}Cl$ (H_c)	-0.219	0.202	$C_6H_{11}Br$ (H_b)	-0.403	0.21	CH_2Br_2(H)	-0.535	0.212
$C_{10}H_{15}Cl$ (H_d)	-0.372	0.198	$C_6H_{11}Br$ (H_c)	-0.375	0.186	CH_2Br_2(Br)	-0.535	0.056
$C_{10}H_{15}Cl$ (Cl)	0.008	-0.087	$C_6H_{11}Br$ (H_d)	-0.379	0.2	CH_2Cl_2(H)	-0.38	0.205
C_2H_5Br (H_a)	-0.396	0.199	$C_6H_{11}Br$ (Br)	-0.21	-0.033	CH_2Cl_2 (Cl)	-0.38	-0.016
C_2H_5Br (H_b)	-0.592	0.207	$C_6H_{11}Cl$ (H_a)	-0.151	0.19	CH_2F_2(H)	0.499	0.125
C_2H_5Br (Br)	-0.396	-0.033	$C_6H_{11}Cl$ (H_b)	-0.399	0.198	CH_2F_2 (F)	0.499	-0.375
C_2H_5Cl (H_a)	-0.331	0.195	$C_6H_{11}Cl$ (H_c)	-0.376	0.202	CH_2I_2(H)	-0.771	0.214
C_2H_5Cl (H_b)	-0.59	0.206	$C_6H_{11}Cl$ (H_d)	-0.378	0.187	CH_2I_2 (I)	-0.771	0.171
C_2H_5Cl (Cl)	-0.331	-0.089	$C_6H_{11}Cl$ (Cl)	-0.151	-0.09	CH_3Br(H)	-0.601	0.206
C_2H_5F (H_a)	0.1	0.151	$C_6H_{11}I$ (H_a)	-0.292	0.197	CH_3Br (Br)	-0.601	-0.018
C_2H_5F (H_b)	-0.611	0.205	$C_6H_{11}I$ (H_b)	-0.408	0.198	CH_3Cl(H)	-0.527	0.201
C_2H_5F (F)	0.1	-0.408	$C_6H_{11}I$ (H_c)	-0.374	0.203	CH_3Cl (Cl)	-0.527	-0.076
C_2H_5I (H_a)	-0.496	0.203	$C_6H_{11}I$ (H_d)	-0.38	0.187	CH_3COBr(H)	-0.675	0.229
C_2H_5I (H_b)	-0.595	0.207	$C_6H_{11}I$ (I)	-0.292	0.059	CH_3COBr (Br)	0.527	-0.071
C_2H_5I (I)	-0.496	0.065	C_6H_5Br (H_b)	-0.226	0.219	CH_3COCH_2Br (H_a)	-0.525	0.228
C_3H_7Br (H_a)	-0.219	0.195	C_6H_5Br (H_c)	-0.185	0.207	CH_3COCH_2Br (H_c)	-0.668	0.222
C_3H_7Br (H_b)	-0.589	0.204	C_6H_5Br (H_d)	-0.205	0.207	CH_3COCH_2Br (Br)	-0.525	0.033
C_3H_7Br (Br)	-0.219	-0.046	C_6H_5Br (Br)	-0.106	0.073	CH_3COCl(H)	-0.674	0.229
C_3H_7Cl (H_a)	-0.164	0.19	C_6H_5Cl (H_b)	-0.226	0.219	CH_3COCl (Cl)	0.569	-0.104
C_3H_7Cl (H_b)	-0.586	0.213	C_6H_5Cl (H_c)	-0.185	0.207	CH_3COF(H)	-0.697	0.237
C_3H_7Cl (Cl)	-0.164	-0.098	C_6H_5Cl (H_d)	-0.206	0.207	CH_3COF (F)	0.908	-0.382
C_3H_7F (H_a)	0.239	0.149	C_6H_5Cl (Cl)	-0.035	0.003	CH_3COI(H)	-0.678	0.243
C_3H_7F (H_b)	-0.602	0.204	C_6H_5F (H_b)	-0.267	0.22	CH_3COI (I)	0.471	-0.014
C_3H_7F (F)	0.239	-0.418	C_6H_5F (H_c)	-0.183	0.206	CH_3F(H)	-0.071	0.155
C_3H_7I (H_a)	-0.304	0.196	C_6H_5F (H_d)	-0.219	0.207	CH_3F (F)	-0.071	-0.394
C_3H_7I (H_b)	-0.594	0.213	C_6H_5F (F)	0.417	-0.358	CH_3I(H)	-0.717	0.21
C_3H_7I (I)	-0.304	0.048	C_6H_5I (H_b)	-0.23	0.218	CH_3I (I)	-0.717	0.089
C_4H_9Br (H_b)	-0.591	0.213	C_6H_5I (H_c)	-0.186	0.208	$CHBr_3$(H)	-0.546	0.208
C_4H_9Br (Br)	-0.059	-0.054	C_6H_5I (H_d)	-0.203	0.207	$CHBr_3$ (Br)	-0.546	0.113
C_4H_9Cl (H_b)	-0.587	0.211	C_6H_5I (I)	-0.204	0.181	$CHCl_3$(H)	-0.302	0.032
C_4H_9Cl (Cl)	-0.013	-0.102	C_6H_6NCl (H_a)	-0.346	0.207	$CHCl_3$ (Cl)	-0.302	0.204
C_4H_9I (H_b)	-0.597	0.203	C_6H_6NCl (H_c)	-0.22	0.215	CHF_3(H)	0.958	0.11
C_4H_9I (I)	-0.13	0.04	C_6H_6NCl (H_d)	-0.159	0.208	CHF_3 (F)	0.958	-0.356
C_5H_4NCl (H_a)	-0.254	0.226	C_6H_6NCl (H_e)	-0.236	0.212	CHI_3 (H)	-0.897	0.201
C_5H_4NCl (H_b)	-0.151	0.211	C_6H_6NCl (H_f)	0.064	0.186	CHI_3 (I)	-0.897	0.232
C_5H_4NCl (H_c)	-0.247	0.214	C_6H_6NCl (Cl)	-0.346	-0.076	$CNCH_2Br$(H)	-0.523	0.249
C_5H_4NCl (H_d)	0.071	0.19	C_7H_7Br (H_a)	-0.39	0.213	$CNCH_2Br$ (Br)	-0.523	0.056
C_5H_4NCl (Cl)	-0.491	0.82	C_7H_7Br (H_c)	-0.181	0.205	$CNCH_2Cl$ (H)	-0.45	0.245
C_5H_4NI (H_a)	0.036	0.197	C_7H_7Br (H_d)	-0.195	0.205	$CNCH_2Cl$ (Cl)	-0.45	-0.013
C_5H_4NI (H_b)	-0.193	0.221	C_7H_7Br (H_e)	-0.194	0.205	$CNCH_2F$ (H)	0.006	0.202
C_5H_4NI (H_c)	-0.226	0.215	C_7H_7Br (Br)	-0.39	-0.034	$CNCH_2F$ (F)	0.006	-0.364
C_5H_4NI (H_d)	0.055	0.188	C_7H_7Cl (H_a)	-0.328	0.207	$CNCH_2I$ (H)	-0.628	0.251
C_5H_4NI (I)	-0.251	0.197	C_7H_7Cl (H_c)	-0.181	0.205	$CNCH_2I$ (I)	-0.628	0.165
C_5H_9Br (H_a)	-0.213	0.207	C_7H_7Cl (H_d)	-0.195	0.205	SC_2H_5Cl (H_a)	-0.525	0.218

C_5H_9Br (H_b)	-0.405	0.217	C_7H_7Cl (H_c)	-0.193	0.205	SC_2H_5Cl (H_c)	-0.707	0.22
C_5H_9Br (H_c)	-0.389	0.204	C_7H_7Cl (Cl)	-0.328	-0.088	SC_2H_5Cl (Cl)	-0.525	-0.079
C_5H_9Br (Br)	-0.213	-0.049	C_7H_7F (H_a)	0.089	0.163	SC_2H_5I (H_a)	-0.697	0.225
C_5H_9Cl (H_a)	-0.156	0.203	C_7H_7F (H_c)	-0.178	0.204	SC_2H_5I (H_c)	-0.71	0.221
C_5H_9Cl (H_b)	-0.402	0.216	C_7H_7F (H_d)	-0.199	0.205	SC_2H_5I (I)	-0.697	0.052
C_5H_9Cl (H_c)	-0.389	0.203	C_7H_7F (H_e)	-0.19	0.205			
C_5H_9Cl (Cl)	-0.156	-0.103	C_7H_7F (F)	0.089	-0.402			

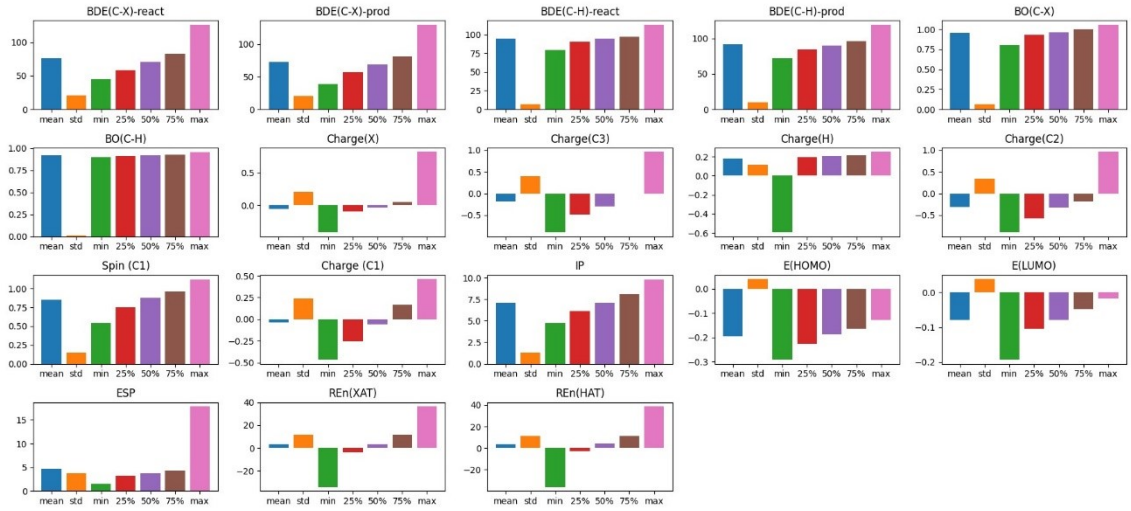


Figure S6. Distribution of phys-organic descriptors

2.3 Feature reduction based on Pearson relation coefficient

Pearson correlation analysis was performed as part of the feature selection process. The correlation coefficients for the 18 initial physical organic descriptors are presented in Figure S7, illustrating the degree of linear relationship between the features.

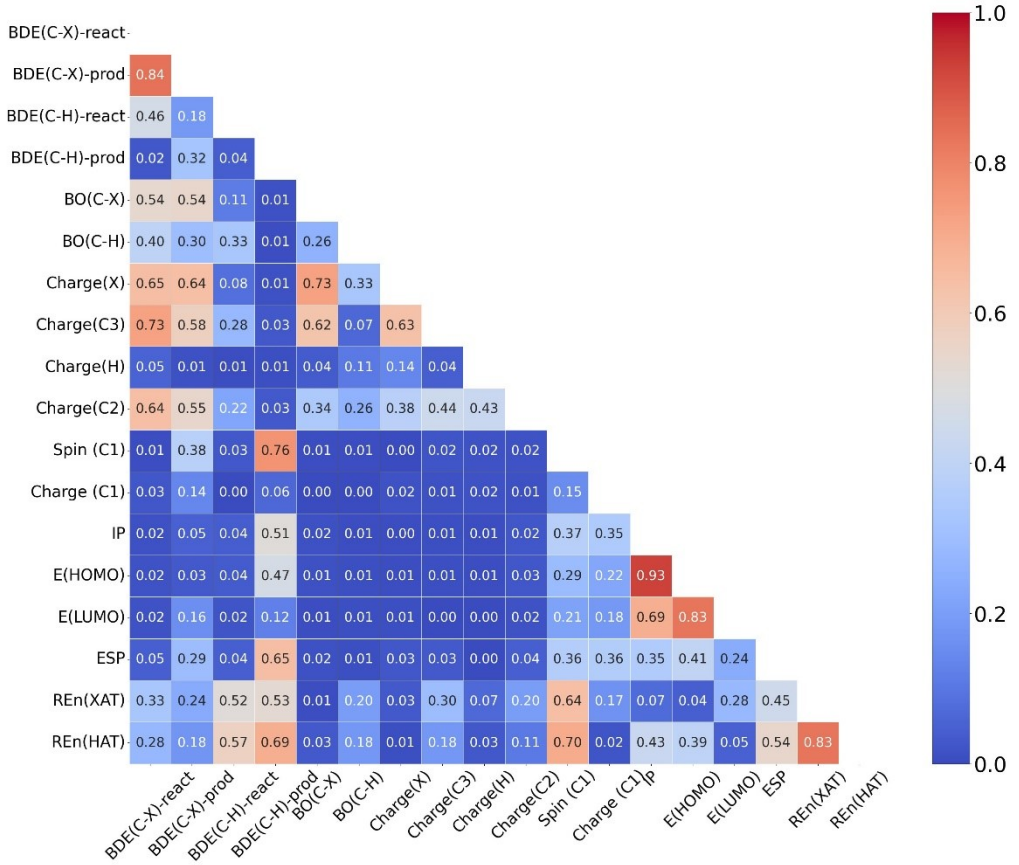


Figure S7. Pearson correlation coefficients of descriptors.

2.4 Machine learning algorithms

A series of widely used machine learning algorithms were evaluated for model development, including Bagging[17], Gradient Boosting (GB)[18, 19], Extra Tree, *k*-nearest neighbors regressor (KNN)[20], Adaboost[21], Multivariate Linear Regression (MLR)[22], Ridge Regressor[23], Lasso[24], Support Vector Machine (SVM)[25], Decision Tree (DT)[26], Random Forest (RF)[27]. All models were implemented using the scikit-learn Python library[28]. Detailed hyperparameter configurations and ensemble settings for each algorithm are provided in Table S6; for any parameters not explicitly listed, default scikit-learn settings were maintained. The complete set of model-training scripts is hosted on our GitHub repository (https://github.com/liusqchem/machine-learning.git).

Table S6. 10-fold cross validation performance and hyperparameters for the evaluated machine learning algorithms.

Model	R^2	MAE	RMSE	Main hyperparameters
Bagging	0.98852	0.86511	1.15737	estimator=None, n_estimators=10, max_samples=1.0, max_features=1.0,
Gradient Boosting	0.98856	0.8838	1.15506	loss='squared_error', n_estimators=100
Extra Tree	0.97349	1.20865	1.75862	n_estimators=100, criterion='squared_error', min_samples_split=2
***k*-nearest neighbors regressor**	0.95307	1.85303	2.33995	n_neighbors=5, weights='uniform', leaf_size=30

Adaboost	0.94562	2.08772	2.5189	n_estimators=50, learning_rate=1.0, loss='linear', random_state=None
Multivariate Linear Regressor	0.91773	2.39126	3.09815	-
Ridge Regressor	0.91716	2.39927	3.10893	alpha=1.0
Lasso	0.69257	4.43712	5.98901	alpha=1.0 '
Support Vector Machine	0.93982	1.93337	2.64977	kernel='rbf', degree=3
Decision Tree	0.98173	1.06747	1.45981	criterion='squared_error', min_samples_split=2
Random Forest	0.99051	0.78561	1.05247	n_estimators=100, criterion='squared_error', min_samples_split=2,

2.5 Feature importance

2.5.1 Feature importance scores

The relative importance was evaluated using the feature importance metrics from RF, GB, Extra Trees, DT, and AdaBoost. Table S7 summarizes these scores, providing a basis for interpreting the chemical drivers of HAT/XAT selectivity.

Table S7. Feature importance scores of RF, GB, Extra tree, DT, AdaBoost and Bagging regressors.

Feature	RF	AdaBoost	GB	DT	Extra Tree	Bagging
BDE(C-X)	0.67873	0.78153	0.58028	0.82925	0.60333	0.71955
BO(C-X)	0.16003	0.03626	0.26268	0.00613	0.19964	0.11755
IP	0.08507	0.09943	0.07969	0.08121	0.07626	0.08496
E(LUMO)	0.03489	0.04317	0.03973	0.03898	0.02762	0.03516
Charge C2	0.009	0.0132	0.01421	0.01213	0.04274	0.00937
BDE(C-H)	0.00844	0.00606	0.00845	0.00833	0.01201	0.00852
ESP	0.00567	0.00285	0.00294	0.00452	0.00954	0.00611
Spin C1	0.00524	0.01026	0.00519	0.00415	0.00581	0.00548
Charge H	0.00515	0.00197	0.00335	0.00766	0.00993	0.00515
BO(C-H)	0.00436	0.0033	0.00211	0.00313	0.00671	0.00454
Charge C1	0.00342	0.00198	0.00136	0.00451	0.00641	0.00362

2.5.2 Influence of feature set size

We investigated the impact of the number of input features on model performance by adding descriptors sequentially according to their RF-derived importance in Table S8. We found that utilizing the top seven features provides an ideal balance between simplicity and accuracy, delivering an R^2 of 0.99 and an MAE of 0.82 kcal/mol for predicting $\boldsymbol{\Delta\Delta G^{\ddagger}}$. This suggests that a subset of the most influential descriptors can effectively represent the reaction energetics.

Table S8. Influence of feature set size on the predictive accuracy (MAE and R^2) of the RF regressor.

Feature number	R^2	MAE	Detailed features
2	0.85592	3.29731	BDE(C-X), BO(C-X)
3	0.98045	1.13479	BDE(C-X), BO(C-X), IP
4	0.98603	0.95918	BDE(C-X), BO(C-X), IP, E(LUMO)

5	0.98685	0.93142	BDE(C-X), BO(C-X), IP, E(LUMO), Charge (C_2)
6	0.98722	0.91648	BDE(C-X), BO(C-X), IP, E(LUMO), Charge (C_2), BDE(C-H)
7	0.98913	0.82468	BDE(C-X), BO(C-X), IP, E(LUMO), Charge (C_2), BDE(C-H), ESP
8	0.98961	0.81308	BDE(C-X), BO(C-X), IP, E(LUMO), Charge (C_2), BDE(C-H), ESP, Spin (C_1)
9	0.98977	0.8059	BDE(C-X), BO(C-X), IP, E(LUMO), Charge (C_2), BDE(C-H), ESP, Spin (C_1), Charge (H)
10	0.98981	0.80482	BDE(C-X), BO(C-X), IP, E(LUMO), Charge (C_2), BDE(C-H), ESP, Spin (C_1), Charge (H), BO(C-H)
11	0.99007	0.79587	BDE(C-X), BO(C-X), IP, E(LUMO), Charge (C_2), BDE(C-H), ESP, Spin (C_1), Charge H, BO(C-H), Charge (C_1)

3. Results for silyl radicals

As shown in Figure S8A, the multivariate linear regression (MLR) model using $\boldsymbol{BDE_{C\text{-}X}}$ and ***IP*** as input features yields a relatively low correlation coefficient (R^2 = 0.59) for silyl radicals, suggesting that additional or alternative descriptors may better capture the selectivity trends in silyl radical-induced atom transfer reactions. To identify more suitable features, we repeated the machine-learning-guided workflow by training a model on 116 DFT-computed $\boldsymbol{\Delta\Delta G^{\ddagger}}$ values and evaluating feature importance using the Random Forest (RF) algorithm.

The RF feature importance analysis revealed that $BDE_{C\text{-}X}$ and $BDE_{C\text{-}H}$ of the organic halides are the most influential descriptors for silyl radical reactivity. Based on this insight, we constructed a new MLR model using these two features. As illustrated in Figure S8B, this model shows a significantly improved correlation with DFT-computed values (R^2 = 0.86), indicating that both C-X and C-H bond strengths play dominant roles in determining atom transfer selectivity for silyl radicals.Figure S8B also reveals that silyl radicals preferentially abstract Cl, Br, and I atoms over H atoms, and a small number of successful F atom transfer (FAT) events are also observed. These results highlight the effectiveness of the machine-learning-guided approach in uncovering reactivity patterns across diverse radical species and atom transfer scenarios.

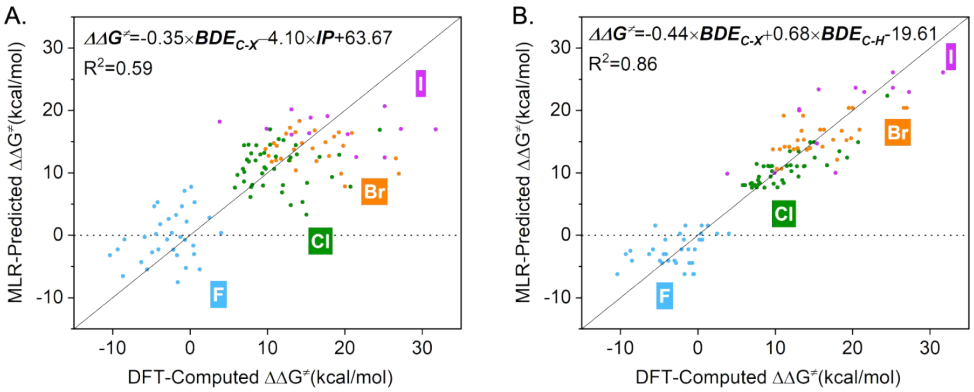


Figure S8. Correlation between DFT-computed $\boldsymbol{\Delta\Delta G^{\ddagger}}$ values and multivariate linear regression predictions for silyl radicals using (A) $\boldsymbol{BDE_{C\text{-}X}}$ and ***IP***, and (B) $\boldsymbol{BDE_{C\text{-}X}}$ and $\boldsymbol{BDE_{C\text{-}H}}$ as input features.

4. References

1. Liu, S.; Su, Y.-L.; Sun, T.-Y.; Doyle, M. P.; Wu, Y.-D.; Zhang, X., Precise introduction of

the -CH_nX_{3-n} (X = F, Cl, Br, I) moiety to target molecules by a radical strategy: A theoretical and experimental study. *J. Am. Chem. Soc.* **2021,** *143* (33), 13195-13204.

2. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

3. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32*, 1456-1465.

4. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988,** *38*, 3098-3100.

5. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988,** *37*, 785-789.

6. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

7. Wadt, W. R.; Hay, P. J., Ab initio effective core potentials for molecular calculations - Potentials for main group elements Na to Bi. *J. Chem. Phys.* **1985,** *82*, 284-298.

8. Hollwarth, A.; Bohme, M.; Dapprich, S.; Ehlers, A. W.; Gobbi, A.; Jonas, V.; Kohler, K. F.; Stegmann, R.; Veldkamp, A.; Frenking, G., A set of d-polarization functions for pseudo-potential basis-sets of the main-group elements Al-Bi and f-type polarization functions for Zn, Cd, Hg. *Chem. Phys. Lett.* **1993,** *208*, 237-240.

9. Merrick, J. P.; Moran, D.; Radom, L., An evaluation of harmonic vibrational frequency scale factors. *J. Phys. Chem. A* **2007,** *111*, 11683-11700.

10. Staroverov, V. N.; Scuseria, G. E.; Tao, J. M.; Perdew, J. P., Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003,** *119*, 12129-12137.

11. Igelmann, G.; Stoll, H.; Preuss, H., Pseudopotentials for main group elements (IIIA through VIIA). *Mol. Phys.* **1988,** *65*, 1321-1328.

12. Bergner, A.; Dolg, M.; Kuchle, W.; Stoll, H.; Preuss, H., Ab-initio energy-adjusted pseudopotentials for elements of groups 13-17. *Mol. Phys.* **1993,** *80*, 1431-1441.

13. Wiberg, K. B., Application of the pople-santry-segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron* **1968,** *24*, 1083-1096.

14. Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural population analysis. *J. Chem. Phys.* **1985,** *83*, 735-746.

15. Mulliken, R. S., Electronic population analysis on LCAO–MO molecular wave functions. I. *J. Chem. Phys.* **1955,** *23*, 1833-1840.

16. Lu, T.; Chen, F. J. J. o. c. c., Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.*

2012, *33*, 580-592.

17. Breiman, L., Bagging predictors. *Mach. Learn.* **1996,** *24*, 123-140.

18. Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I., From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020,** *2*, 56-67.

19. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. In *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*, 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 2017.

20. Song, Y.; Liang, J.; Lu, J.; Zhao, X., An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing* **2017,** *251*, 26-34.

21. Schapire, R. E.; Singer, Y., Improved boosting algorithms using confidence-rated predictions. *Mach. Learn.* **1999,** *37*, 297-336.

22. Santiago, C. B.; Guo, J.-Y.; Sigman, M. S., Predictive and mechanistic multivariate linear regression models for reaction development. *Chem. Sci.* **2018,** *9*, 2398-2412.

23. Garcia, C. B.; Garcia, J.; Lopez Martin, M. M.; Salmeron, R., Collinearity: revisiting the variance inflation factor in ridge regression. *J. Appl. Stat.* **2015,** *42*, 648-661.

24. Friedman, J.; Hastie, T.; Tibshirani, R., Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010,** *33*, 1.

25. Burges, C. J. C., A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Disc.* **1998,** *2*, 121-167.

26. Ture, M.; Tokatli, F.; Kurt, I., Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Syst. Appl.* **2009,** *36* (2, Part 1), 2017-2026.

27. Breiman, L., Random forests. *Mach. Learn.* **2001,** *45*, 5-32.

28. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E., Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011,** *12*, 2825-2830.