Supplementary Materials

Accelerated discovery of high-performance small-molecule hole transport materials via molecular splicing, high-throughput screening, and machine learning

Jiansen Wen[1,2,#], Shuwen Yang[1,3,4,#], Linqin Jiang[2,*], Yudong Shi[1], Zhihan Huang[4], Ping Li[2], Hao Xiong[2], Ze Yu[4], Xushan Zhao[4], Bo Xu[4], Bo Wu[1,3,*], Baisheng Sa[1,*], Yu Qiu[2]

[1]Multiscale Computational Materials Facility & Materials Genome Institute, School of Materials Science and Engineering, Fuzhou University, Fuzhou 350100, Fujian, China.

[2]Key Laboratory of Green Perovskites Application of Fujian Province Universities, Fujian Jiangxia University, Fuzhou 350100, Fujian, China.

[3]Materials Design and Manufacture Simulation Facility, School of Advanced Manufacturing, Fuzhou University, Jinjiang 362200, Fujian, China.

[4]Fujian Science and Technology Innovation Laboratory for Energy Devices (21C-Lab), Contemporary Amperex Technology Co., Limited (CATL), Ningde 352100, Fujian, China.

[#]Authors contributed equally.

[*]**Correspondence to:** Prof. Linqin Jiang, Key Laboratory of Green Perovskites Application of Fujian Province Universities, Fujian Jiangxia University, No. 2 Xiyuangong Road, Fuzhou 350100, Fujian, China, E-mail: linqinjiang@fjjxu.edu.cn; Prof. Bo Wu, Prof. Baisheng Sa, Multiscale Computational Materials Facility & Materials Genome Institute, School of Materials Science and Engineering, Fuzhou University, No. 2 Wulongjiang Avenue, Fuzhou 350100, Fujian, China, E-mail: wubo@fzu.edu.cn; bssa@fzu.edu.cn

Supplementary Table 1. Molecular properties of common hole transport materials

Molecule	HOMO (eV)	Hole reorganization energy (eV)	Solvation free energy (eV)	Maximum absorption peak (nm)	LogP	SAScore	Ref
spiro-OMeTAD	-4.217	0.152	-56.33	395.89	13.392	3.952	This work
	-4.18						[1]
	-4.34						[2]
	-4.44						[3]
DTPC8-ThDTPA	-4.247	0.251	-51.08	531.73	15.413	3.590	This work
	-4.22						[1]
DTPC13-ThTPA	-4.266	0.252	-52.15	533.86	16.704	4.261	This work
	-4.25						[4]
DTP-C6Th	-4.292	0.261	-52.22	533.16	16.054	3.887	This work
	-4.30						[5]
YZ18	-4.531	0.122	-41.18	405	10.389	2.472	This work
YZ22	-4.625	0.124	-42.66	429.12	7.358	2.686	This work
TPA-TVT-TPA	-4.336	0.256	-40.01	515.28	10.413	2.714	This work
	-4.49						[6]

Supplementary Table 2. The scores of the top 20 ranked molecules

Molecule	Score
M1	0.788
M2	0.768
M3	0.754
M4	0.753
M5	0.752
M6	0.750
M7	0.749
M8	0.748
M9	0.748
M10	0.747
M11	0.745
M12	0.743
M13	0.742
M14	0.742
M15	0.739
M16	0.739
M17	0.738
M18	0.738
M19	0.735
M20	0.734

Supplementary Table 3. The molecular IDs and their occurrence frequencies among the top 10 candidates across all weighting scenarios

Molecular ID	Frequency
2901	124
4024	108
3959	88
3499	86
3621	80
3948	70
4186	70
4551	66
4853	60
3129	60

Supplementary Table 4. Molecular properties of six screened SM-HTM molecules M1-M6

Molecule	HOMO (eV)	Hole reorganization energy (eV)	Solvation free energy (eV)	Maximum absorption peak (nm)	LogP	SAScore
M1	-4.287	0.218	-41.94	395.59	10.130	3.531
M2	-4.306	0.168	-44.91	387.55	10.094	4.131
M3	-4.236	0.224	-42.76	393.74	9.156	3.566
M4	-4.284	0.173	-41.60	384.75	9.359	3.670
M5	-4.281	0.190	-39.85	391.27	9.332	3.474
M6	-4.278	0.158	-43.58	370.60	9.610	4.038

Supplementary Table 5. The performance indices for the hole reorganization energy, solvation free energy, maximum absorption, and LogP across the RF, GBDT, and XGBoost models

Performance indices	Model	R^2	MAE	RMSE
Hole reorganization energy	RF	0.739	0.035 eV	0.079 eV
	GBDT	0.865	0.028 eV	0.055 eV
	XGBoost	0.901	0.024 eV	0.047 eV
Solvation free energy	RF	0.996	0.394 eV	0.637 eV
	GBDT	0.997	0.405 eV	0.552 eV
	XGBoost	0.998	0.263 eV	0.406 eV
Maximum absorption	RF	0.923	17.908 nm	30.606 nm
	GBDT	0.946	18.363 nm	25.695 nm
	XGBoost	0.969	12.954 nm	19.423 nm
LogP	RF	0.991	0.125 eV	0.192 eV
	GBDT	0.992	0.140 eV	0.186 eV
	XGBoost	0.996	0.078 eV	0.131 eV

Supplementary Table 6. Detailed information of the top 15 important features in the RF model trained by the hole reorganization energy dataset

	Importance	Feature	Explanation
1	0.37701	fr_hdrzine	Number of hydrazine groups
2	0.08119	NumAromaticRings	The number of aromatic rings of the molecule
3	0.05751	VSA_EState8	Descriptors represented by a mixture of van der Waals surface area and electrical topological state indices
4	0.04455	SlogP_VSA4	Descriptors for Van der Waals surface area and LogP mixed representations
5	0.04366	BCUT2D_MWLOW	The minimum eigenvalue of the atomic weight
6	0.03217	VSA_EState4	Descriptors represented by a mixture of van der Waals surface area and electrical topological state exponents
7	0.03101	FpDensityMorgan1	Morgan fingerprint density
8	0.02200	VSA_EState2	Descriptors represented by a mixture of van der Waals surface area and electrical topological state indices
9	0.01916	BCUT2D_LOGPLOW	The minimum eigenvalue of MolLogP
10	0.01747	qed	The weighted sum of the ADS mapping attributes
11	0.01235	BCUT2D_MRLOW	Minimum eigenvalue of MolMR
12	0.01108	VSA_EState9	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
13	0.01005	SlogP_VSA5	Descriptor for a mixture of refractive index representations calculated from van der Waals surface area and molecular properties
14	0.00875	SlogP_VSA8	Descriptors for Van der Waals surface area and LogP mixed representations
15	0.00797	BCUT2D_LOGPHI	The maximum unique value of MolLogP

Supplementary Table 7. Detailed information of the top 15 important features in the RF model trained by the solvation free energy dataset

	Importance	Feature	Explanation
1	0.49475	Chi1n	Molecular connectivity index 1n
2	0.40574	Chi0n	Molecular connectivity index 0n
3	0.04238	MolMR	Relative molecular weight
4	0.00970	NumValenceElectrons	The number of valence electrons of the molecule
5	0.00612	Chi0	Molecular connectivity index 0
6	0.00365	qed	The weighted sum of the ADS mapping attributes
7	0.00277	VSA_EState7	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
8	0.00277	Chi1	Molecular connectivity index 1
9	0.00256	SlogP_VSA2	Descriptors for Van der Waals surface area and LogP mixed representations
10	0.00237	NumRotatableBonds	Number of rotatable keys
11	0.00229	VSA_EState2	Descriptors represented by a mixture of van der Waals surface area and electrical topological state indices
12	0.00222	fr_NH1	The amount of secondary amine
13	0.00176	LabuteASA	Approximate surface area of Labute
14	0.00175	Chi0v	Molecular connectivity index 0v
15	0.00175	BalabanJ	Balaban's J value

Supplementary Table 8. Detailed information of the top 15 important features in the RF model trained by the maximum absorption dataset

	Importance	Feature	Explanation
1	0.258558	NumRadicalElectrons	The number of free radical electrons in a molecule
2	0.24698	fr_allylic_oxid	Number of allyl oxidation sites
3	0.066981	Kappa3	Kappa Shape Index 3
4	0.055103	qed	The weighted sum of the ADS mapping attributes
5	0.048191	VSA_EState10	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
6	0.029353	SMR_VSA7	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
7	0.022103	PEOE_VSA7	Descriptor for the mixed representation of Van der Waals surface area and atomic local charge
8	0.021737	Kappa2	Kappa Shape Index 2
9	0.017433	SlogP_VSA4	Descriptors for Van der Waals surface area and LogP mixed representations
10	0.016589	VSA_EState8	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
11	0.015948	BCUT2D_MWLOW	The minimum eigenvalue of the atomic weight
12	0.013212	SMR_VSA10	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
13	0.012968	VSA_EState9	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
14	0.010629	BCUT2D_CHGHI	The maximum intrinsic value of the Gasteiger charge
15	0.00879	BCUT2D_CHGLO	The minimum intrinsic value of the Gasteiger charge

Supplementary Table 9. Detailed information of the top 15 important features in the RF model trained by the LogP dataset

	Importance	Feature	Explanation
1	0.695163	Kappa2	Kappa Shape Index 2
2	0.191859	MolLogP	Lipid-water partition coefficient
3	0.032059	FractionCSP3	SP3 hybrid carbon atom ratio
4	0.013841	TPSA	Topological molecular polar surface area
5	0.004573	MolMR	The amount of relative molecular mass
6	0.003969	SMR_VSA7	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
7	0.003205	Chi3v	Molecular connectivity index 3v
8	0.002969	VSA_EState6	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
9	0.002777	Kappa3	Kappa Shape Index 3
10	0.002501	Kappa1	Kappa Shape Index 1
11	0.002319	fr_Ar_NH	The amount of aromatic amines
12	0.002252	fr_Nhpyrrole	Number of H-pyrrole nitrogen
13	0.001869	SlogP_VSA2	Descriptors for Van der Waals surface area and LogP mixed representations
14	0.001816	VSA_EState9	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
15	0.001683	VSA_EState10	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices

Supplementary Table 10. Detailed information of the top 15 important features in the GBDT model trained by the hole reorganization energy dataset

	Importance	Feature	Explanation
1	0.319715	fr_hdrzine	Number of hydrazine groups
2	0.059361	VSA_EState8	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
3	0.054819	BCUT2D_LOGPLOW	The minimum eigenvalue of MolLogP
4	0.054197	SlogP_VSA4	Descriptors for Van der Waals surface area and LogP mixed representations
5	0.053922	BCUT2D_MWLOW	The minimum eigenvalue of the atomic weight
6	0.051922	NumAromaticRings	The number of aromatic rings of the molecule
7	0.037224	VSA_EState4	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
8	0.032078	FpDensityMorgan1	Morgan fingerprint density
9	0.028174	BCUT2D_MRLOW	Minimum eigenvalue of MolMR
10	0.024034	qed	The weighted sum of the ADS mapping attributes
11	0.019701	SlogP_VSA8	Descriptors for Van der Waals surface area and LogP mixed representations
12	0.012555	VSA_EState9	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
13	0.012102	VSA_EState2	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
14	0.011656	PEOE_VSA8	A descriptor for the van der Waals surface area and atomic local charge mixture representation
15	0.010848	VSA_EState1	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices

Supplementary Table 11. Detailed information of the top 15 important features in the GBDT model trained by the solvation free energy dataset

	Importance	Feature	Explanation
1	0.361111	Chi1n	Molecular connectivity index 1n
2	0.302687	Chi0n	Molecular connectivity index 0n
3	0.118406	NumRotatableBonds	Number of rotatable keys
4	0.039749	qed	The weighted sum of the ADS mapping attributes
5	0.034523	MolMR	Relative molecular weight
6	0.033001	NumValenceElectrons	The number of valence electrons of the molecule
7	0.028487	Chi0	Molecular connectivity index 0
8	0.028147	VSA_EState7	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
9	0.025879	BalabanJ	Balaban's J value
10	0.005531	NHOHCount	Number of NHs or OHs
11	0.003546	VSA_EState2	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
12	0.003345	SMR_VSA3	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
13	0.00282	Chi2n	Molecular connectivity index 2n
14	0.002469	SlogP_VSA2	Descriptors for Van der Waals surface area and LogP mixed representations
15	0.0022	Kappa2	Kappa Shape Index 2

Supplementary Table 12. Detailed information of the top 15 important features in the GBDT model trained by the maximum absorption dataset

	Importance	Feature	Explanation
1	0.259278	fr_allylic_oxid	Number of allyl oxidation sites
2	0.245641	NumRadicalElectrons	The number of free radical electrons in a molecule
3	0.072799	qed	The weighted sum of the ADS mapping attributes
4	0.055005	SMR_VSA7	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
5	0.039573	PEOE_VSA7	A descriptor for the van der Waals surface area and atomic local charge mixture representation
6	0.039153	VSA_EState10	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
7	0.038396	VSA_EState8	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
8	0.032512	Kappa3	Kappa Shape Index 3
9	0.023293	SlogP_VSA4	Descriptors for Van der Waals surface area and LogP mixed representations
10	0.021563	VSA_EState9	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
11	0.018034	SlogP_VSA5	Descriptors for Van der Waals surface area and LogP mixed representations
12	0.017581	BCUT2D_MWLOW	The minimum eigenvalue of the atomic weight
13	0.014932	SMR_VSA10	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
14	0.011221	BCUT2D_CHGLO	The minimum intrinsic value of the Gasteiger charge
15	0.008814	MaxAbsEStateIndex	The maximum absolute value of the electrical topological state index for all atoms

Supplementary Table 13. Detailed information of the top 15 important features in the GBDT model trained by the LogP dataset

	Importance	Feature	Explanation
1	0.430777	Kappa2	Kappa Shape Index 2
2	0.320665	MolLogP	Lipid-water partition coefficient
3	0.075336	SMR_VSA7	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
4	0.036489	FractionCSP3	SP3 hybrid carbon atom ratio
5	0.031081	Kappa3	Kappa Shape Index 3
6	0.025256	SlogP_VSA6	Descriptors for Van der Waals surface area and LogP mixed representations
7	0.00884	PEOE_VSA7	A descriptor for the van der Waals surface area and atomic local charge mixture representation
8	0.007813	TPSA	Topological molecular polar surface area
9	0.005967	VSA_EState10	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
10	0.005728	SMR_VSA10	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
11	0.005535	PEOE_VSA6	A descriptor for the van der Waals surface area and atomic local charge mixture representation
12	0.005284	fr_Nhpyrrole	Number of H-pyrrole nitrogen
13	0.00409	VSA_EState6	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
14	0.003355	VSA_EState9	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
15	0.003222	NHOHCount	Number of NHs or OHs

Supplementary Table 14. Detailed information of the top 15 important features in the XGBoost model trained by the hole reorganization energy dataset

	Importance	Feature	Explanation
1	0.563537	fr_hdrzine	Number of hydrazine groups
2	0.178515	NumAromaticRings	The number of aromatic rings of the molecule
3	0.069851	SlogP_VSA4	Descriptors for Van der Waals surface area and LogP mixed representations
4	0.013171	VSA_EState8	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
5	0.009369	SlogP_VSA8	Descriptors for Van der Waals surface area and LogP mixed representations
6	0.008896	NHOHCount	Number of NHs or OHs
7	0.007586	SMR_VSA5	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
8	0.007575	EState_VSA9	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
9	0.007543	SlogP_VSA1	Descriptors for Van der Waals surface area and LogP mixed representations
10	0.006211	RingCount	The number of rings
11	0.005231	Chi1n	Molecular connectivity index 1n
12	0.005164	NumAliphaticCarbocycles	The number of aliphatic (containing at least one non-aromatic bond) carbon ring of the molecule
13	0.004898	Chi3n	Molecular connectivity index 3n
14	0.004549	fr_aniline	The amount of aniline
15	0.004515	qed	The weighted sum of the ADS mapping attributes

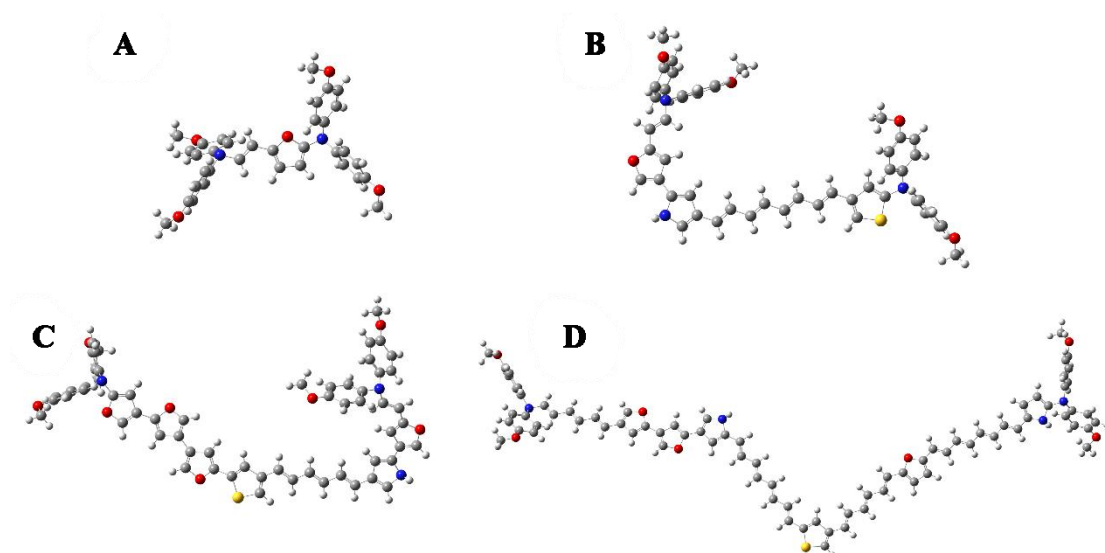**Supplementary Table 15. Detailed information of the top 15 important features in the XGBoost model trained by the solvation free energy dataset**

	Importance	Feature	Explanation
1	0.387199134	Chi0n	Molecular connectivity index 0n
2	0.165416628	Chi1n	Molecular connectivity index 1n
3	0.148235649	NumRotatableBonds	Number of rotatable keys
4	0.123958759	NumValenceElectrons	The number of valence electrons of the molecule
5	0.093575604	MolMR	Relative molecular weight
6	0.035778038	Chi0	Molecular connectivity index 0
7	0.012434168	Kappa1	Kappa Shape Index 1
8	0.011918288	VSA_EState7	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
9	0.006515387	BalabanJ	Balaban's J value
10	0.002878686	Chi2n	Molecular connectivity index 2n
11	0.002295791	NHOHCount	Number of NHs or OHs
12	0.001895435	Kappa2	Kappa Shape Index 2
13	0.000985838	SlogP_VSA2	Descriptors for Van der Waals surface area and LogP mixed representations
14	0.000567186	fr_allylic_oxid	Number of allyl oxidation sites
15	0.000534783	SMR_VSA3	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties

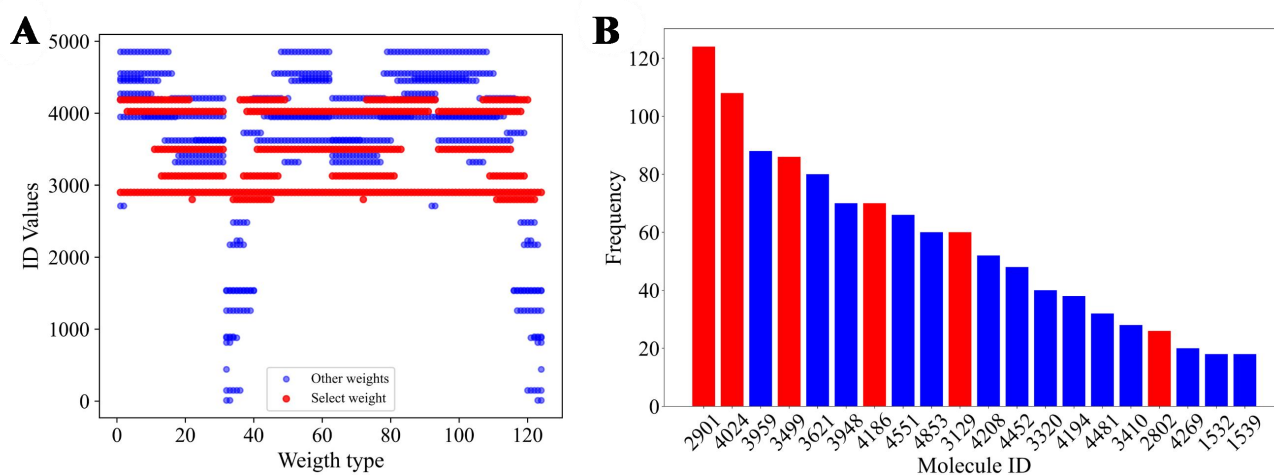**Supplementary Table 16. Detailed information of the top 15 important features in the XGBoost model trained by the maximum absorption dataset**

	Importance	Feature	Explanation
1	0.364299	NumRadicalElectrons	The number of free radical electrons in a molecule
2	0.181128	fr_allylic_oxid	Number of allyl oxidation sites
3	0.08823	NumAliphaticCarbocycles	The number of aliphatic (containing at least one non-aromatic bond) carbon ring of the molecule
4	0.070412	Kappa3	Kappa Shape Index 3
5	0.037964	Kappa2	Kappa Shape Index 2
6	0.034381	SlogP_VSA4	Descriptors for Van der Waals surface area and LogP mixed representations
7	0.034225	SMR_VSA7	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
8	0.021496	VSA_EState10	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
9	0.017372	qed	The weighted sum of the ADS mapping attributes
10	0.011444	PEOE_VSA7	A descriptor for the van der Waals surface area and atomic local charge mixture representation
11	0.010118	NumAliphaticRings	The number of fat rings
12	0.007983	VSA_EState8	A descriptor represented by a mixture of van der Waals surface area and electrotopological state indices
13	0.006885	fr_aniline	The amount of aniline
14	0.00618	fr_NH0	The amount of tertiary amine
15	0.006159	SMR_VSA10	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties

Supplementary Table 17. Detailed information of the top 15 important features in the XGBoost model trained by the LogP dataset

	Importance	Feature	Explanation
1	0.342064	Kappa2	Kappa Shape Index 2
2	0.327091	MolLogP	Lipid-water partition coefficient
3	0.090669	SMR_VSA7	A descriptor for a mixed representation of the refractive index calculated from van der Waals surface area and molecular properties
4	0.055303	fr_Ar_NH	The amount of aromatic amines
5	0.051836	FractionCSP3	SP3 hybrid carbon atom ratio
6	0.014837	SlogP_VSA6	Descriptors for Van der Waals surface area and LogP mixed representations
7	0.013762	NHOHCount	Number of NHs or OHs
8	0.010818	LabuteASA	Calculate the approximate surface area of the Labute
9	0.008291	NumAliphaticHeterocycles	The number of aliphatic (containing at least one non-aromatic bond) heterocycle of the molecule
10	0.006548	Kappa3	Kappa Shape Index 3
11	0.006399	Kappa1	Kappa Shape Index 1
12	0.005264	MolMR	The amount of relative molecular mass
13	0.00449	PEOE_VSA7	A descriptor for the van der Waals surface area and atomic local charge mixture representation
14	0.004179	TPSA	Topological molecular polar surface area
15	0.004071	SlogP_VSA2	Descriptors for Van der Waals surface area and LogP mixed representations

Supplementary Table 18. Calculation of absolute and mean absolute errors of DFT values and machine learning predictions
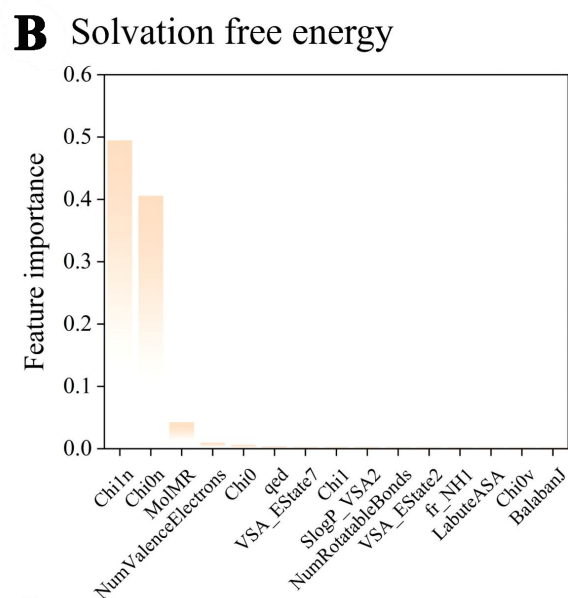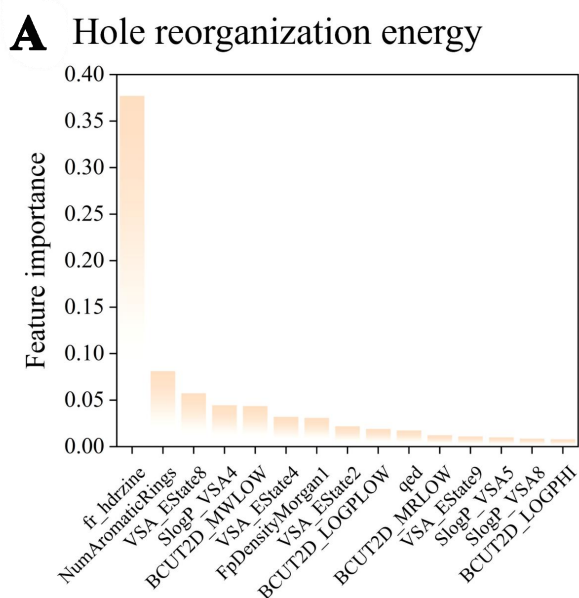
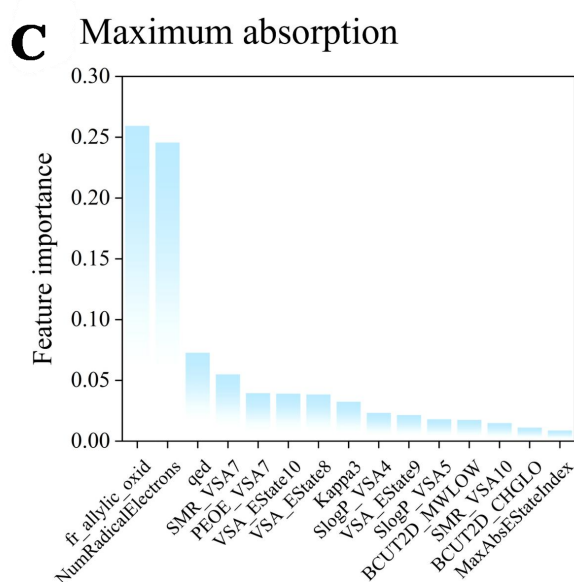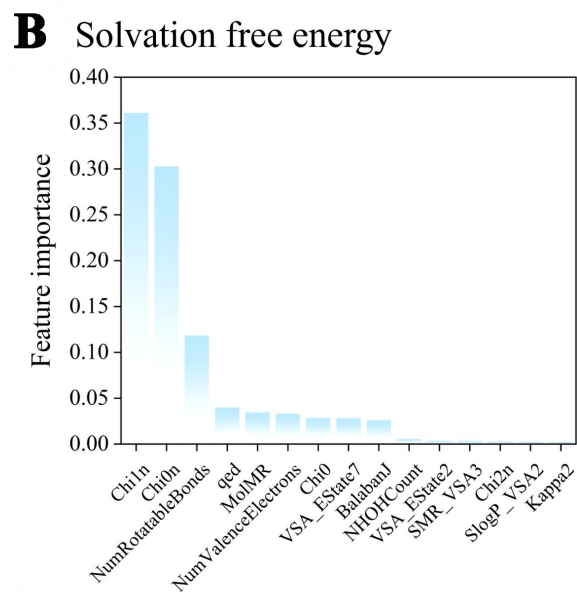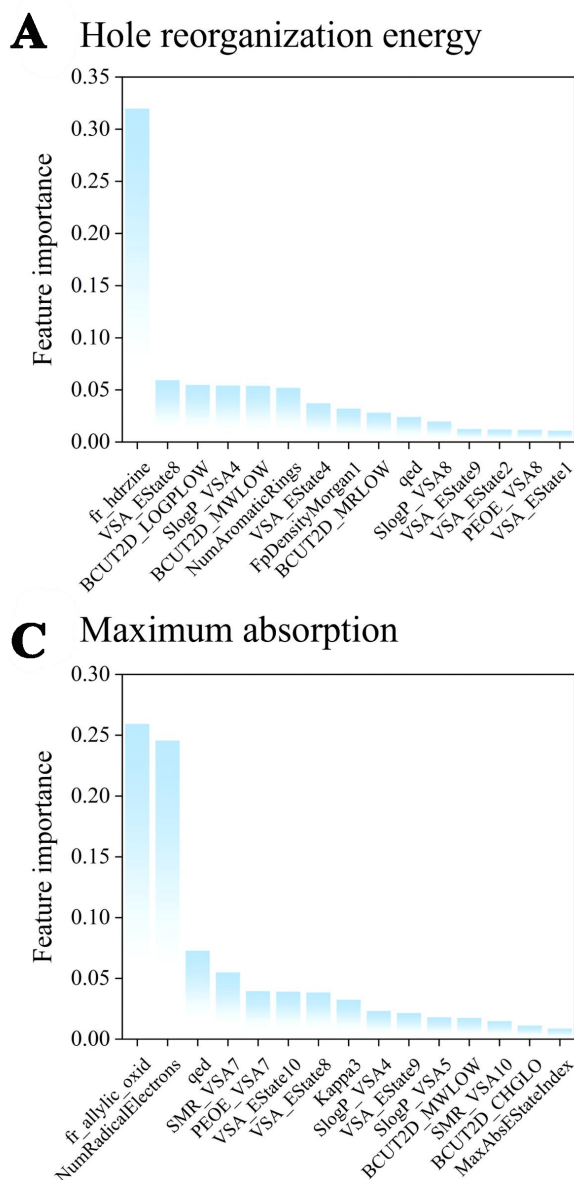| Property	Method	Spiro-O MeTAD	DTP-C6Th	DTPC8-T hDTPA	DTPC1 3-ThTP A	TPA-T VT-TP A	YZ18	YZ22	Mean absolute error
Hole reorganizati on energy	RF	0.058	0.062	0.065	0.002	0.014	0.018	0.030	0.036
	GBDT	0.072	0.044	0.088	0.018	0.012	0.054	0.058	0.049
	XGBoost	0.015	0.046	0.003	0.055	0.024	0.035	0.019	0.028
Solvation free energy	RF	16.547	10.415	8.761	12.888	0.534	0.807	1.445	7.343
	GBDT	4.033	5.470	4.469	8.027	0.232	1.398	0.245	3.410
	XGBoost	1.460	1.941	2.181	3.520	4.547	7.068	6.958	3.953
Maximum absorption	RF	22.500	73.481	74.737	58.645	43.859	22.269	7.669	43.308
	GBDT	97.715	43.665	55.679	29.353	25.680	35.700	13.011	42.972
	XGBoost	58.411	65.969	78.624	84.993	27.374	22.306	51.41	55.584
LogP	RF	0.593	1.376	0.788	1.980	0.378	0.661	1.049	0.975
	GBDT	0.918	1.548	0.995	2.063	0.325	0.592	1.342	1.112
	XGBoost	2.329	3.163	2.854	3.262	1.881	1.785	0.083	2.194

Supplementary Figure 1. Specific examples of the D-π-D splicing results: (A)-(D) represent four typical D-π-D spliced molecules with different numbers of atoms.

Supplementary Figure 2. (A) Distribution of the top 10 molecules with the highest weighted scores for all candidate molecules; (B) Ranking of the 20 most frequent molecules in all weighted cases. The red dots and columns represent the HTM molecular candidates that we have screened.
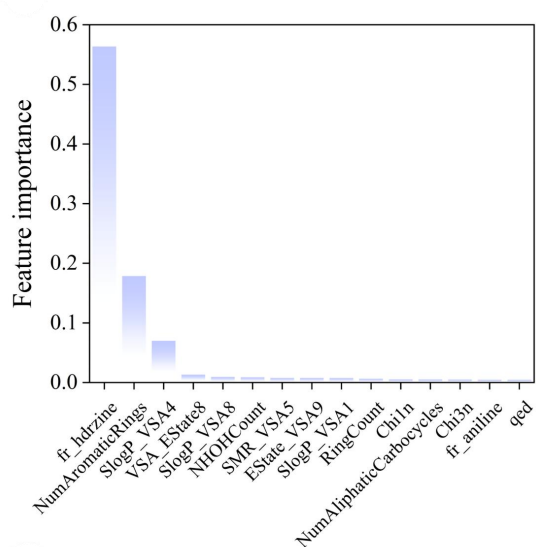
A Hole reorganization energy

B Solvation free energy

C Maximum absorption

D LogP

Supplementary Figure 3. The importance ranking of 15 descriptors for (A) Hole reorganization energy, (B) Solvation free energy, (C) Maximum absorption, and (D) LogP obtained based on the RF model, respectively.
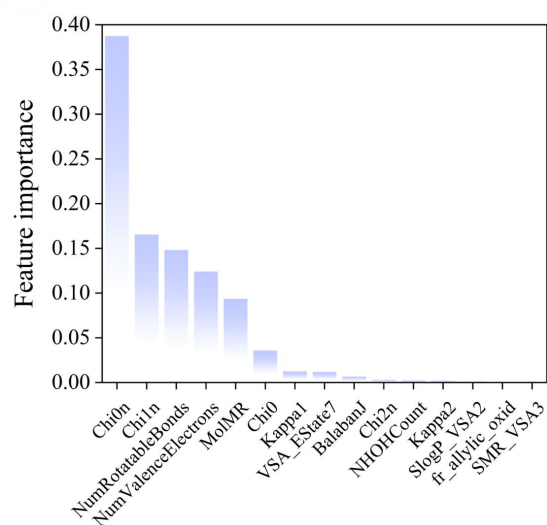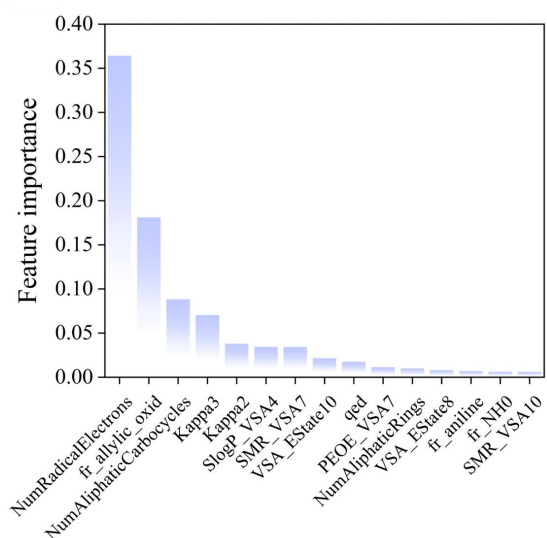
Supplementary Figure 4. The importance ranking of 15 descriptors for (A) Hole reorganization energy, (B) Solvation free energy, (C) Maximum absorption, and (D) LogP obtained based on the GBDT model, respectively.
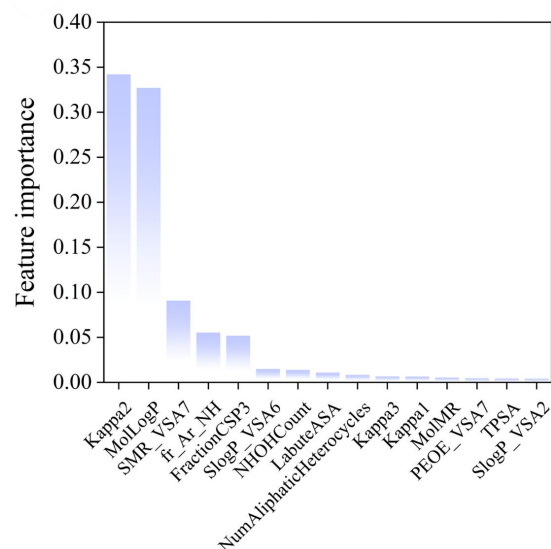
A Hole reorganization energy

B Solvation free energy

C Maximum absorption

D LogP

Supplementary Figure 5. The importance ranking of 15 descriptors for (A) Hole reorganization energy, (B) Solvation free energy, (C) Maximum absorption, and (D) LogP obtained based on the XGBoost model, respectively.

References

[1] Dong, Z.; Yin, X.; Ali, A., et al. A dithieno[3,2-b:2′,3′-d]pyrrole-cored four-arm hole transporting material for over 19% efficiency dopant-free perovskite solar cells. *J. Mater. Chem. C* **2019**, *7*, 9455-9459. DOI: 10.1039/C9TC03111A

[2] Xu, J.; Liang, L.; Mai, C., et al. Lewis-base containing spiro type hole transporting materials for high-performance perovskite solar cells with efficiency approaching 20%. *Nanoscale* **2020**, *12*, 13157-13164. DOI: 10.1039/D0NR01961E

[3] Sandoval-Torrientes, R.; Zimmermann, I.; Calbo, J., et al. Hole transporting materials based on benzodithiophene and dithienopyrrole cores for efficient perovskite solar cells. *J. Mater. Chem. A* **2018**, *6*, 5944-5951. DOI: 10.1039/C7TA11314E

[4] Zhou, J.; Yin, X.; Dong, Z., et al. Dithieno[3,2-b:2′,3′-d]pyrrole cored p-type semiconductors enabling 20 % efficiency dopant-free perovskite solar cells. *Angew. Chem., Int. Ed.* **2019**, *58*, 13717-13721. DOI: 10.1002/anie.201905624

[5] Yin, X.; Zhou, J.; Song, Z., et al. Dithieno[3,2-b:2′,3′-d]pyrrol-cored hole transport material enabling over 21% efficiency dopant-free perovskite solar cells. *Adv. Funct. Mater.* **2019**, *29*, 1904300. DOI: 10.1002/adfm.201904300

[6] Pham, H. D.; Hu, H.; Feron, K., et al. Thienylvinylenethienyl and naphthalene core substituted with triphenylamines—highly efficient hole transporting materials and their comparative study for inverted perovskite solar cells. *Sol. RRL* **2017**, *1*, 1700105. DOI: 10.1002/solr.201700105