Electronic Supplementary Material (ESI) for New Journal of Chemistry.
This journal is © The Royal Society of Chemistry and the Centre National de la Recherche Scientifique 2022

Data-driven Approach Towards Identifying Dye-Sensitizer Molecules for Higher Power Conversion Efficiency in Solar Cells

Ganapathi Rao Kandregula[1], Dhinesh Kumar Murugaiah[1], N. Arul Murugan[2,3]*, Kothandaraman Ramanujam[1,4,5]*

[1]*Department of Chemistry, Indian Institute of Technology Madras, Chennai, India, 600 036.*

[2]*Department of Computer Science, School of Electrical Engineering and Computer Science 106 91, Stockholm, Sweden.*

[3]*Department of Computational Biology, Indraprastha Institute of Information Technology,*

New Delhi, India - 110020.

[4]*DST-IITM Solar Energy Harnessing Centre, Indian Institute of Technology Madras, Chennai, India, 600 036.*

[5]*Potential Centre of Excellence: Advanced Centre for Energy Storage and Conversion, Indian Institute of Technology Madras, Chennai, India 600036.*

**Corresponding author's e-mail: murugan@kth.se & rkraman@iitm.ac.in*

Contents:

Table S1. The details of the descriptors selected for the machine learning algorithms

Entry	Name	Details
1	nCl	Number of chlorine atoms
2	ATSC3c	Centered Broto-Moreau autocorrelation - lag 3 / weighted by charges
3	ATSC4c	Centered Broto-Moreau autocorrelation - lag 4 / weighted by charges
4	ATSC7c	Centered Broto-Moreau autocorrelation - lag 7 / weighted by charges
5	ATSC4e	Centered Broto-Moreau autocorrelation - lag 4 / weighted by Sanderson electronegativities
6	AATSC5i	Average centered Broto-Moreau autocorrelation - lag 5 / weighted by first ionization potential
7	AATSC8i	Average centered Broto-Moreau autocorrelation - lag 8 / weighted by first ionization potential
8	AATSC0s	Average centered Broto-Moreau autocorrelation - lag 0 / weighted by I-state
9	MATS3c	Moran autocorrelation - lag 3 / weighted by charges
10	MATS7c	Moran autocorrelation - lag 7 / weighted by charges
11	MATS4v	Moran autocorrelation - lag 4 / weighted by van der Waals volumes
12	MATS4p	Moran autocorrelation - lag 4 / weighted by polarizabilities
13	MATS5i	Moran autocorrelation - lag 5 / weighted by first ionization potential
14	MATS8s	Moran autocorrelation - lag 8 / weighted by I-state
15	GATS6c	Geary autocorrelation - lag 6 / weighted by charges
16	GATS4v	Geary autocorrelation - lag 4 / weighted by van der Waals volumes
17	GATS4p	Geary autocorrelation - lag 4 / weighted by polarizabilities
18	GATS5i	Geary autocorrelation - lag 5 / weighted by first ionization potential
19	GATS6i	Geary autocorrelation - lag 6 / weighted by first ionization potential
20	GATS3s	Geary autocorrelation - lag 3 / weighted by I-state
21	GATS4s	Geary autocorrelation - lag 4 / weighted by I-state
22	CrippenLogP	Crippen's LogP (Crippen Descriptor)
23	VE1_Dt	Coefficient sum of the last eigenvector from detour matrix
24	nsOm	Count of atom-type E-State: -O-
25	nsCl	Count of atom-type E-State: -Cl
26	SHCsatu	Sum of atom-type H E-State: H on C sp3 bonded to unsaturated C
27	SsNH2	Sum of atom-type E-State: -NH2
28	minHBint2	Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 2
29	minHBint4	Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 4
30	minHsOH	Minimum atom-type H E-State: -OH

31	minHCsatu	Minimum atom-type H E-State: H on C sp3 bonded to unsaturated C
32	minsssCH	Minimum atom-type E-State: >CH-
33	maxHCsatu	Maximum atom-type H E-State: H on C sp3 bonded to unsaturated C
34	maxtsC	Maximum atom-type E-State: #C-
35	maxtN	Maximum atom-type E-State: #N
36	ETA_Epsilon_2	A measure of electronegative atom count
37	ETA_Epsilon_4	A measure of electronegative atom count
38	ETA_dEpsilon_C	A measure of contribution of electronegativity
39	n4Ring	Number of 4-membered rings
40	nF11Ring	Number of 11-membered fused rings
41	nF12Ring	Number of 12-membered fused rings
42	nT4Ring	Number of 4-membered rings (includes counts from fused rings)
43	nT11Ring	Number of 11-membered rings (includes counts from fused rings)
44	nF11HeteroRing	Number of 11-membered fused rings containing heteroatoms (N, O, P, S, or halogens)
45	VE1_D	Coefficient sum of the last eigenvector from topological distance matrix
46	Active area	Total area covered by dye on TiO_2 surface
47	Coabsorbent (mMol)	Quantity of Coabsorbent used in the measurement in milli moles
48	Active layer(um)	Thickness of TiO_2 surface active layer
49	Scattering_ layer (um)	Thickness of TiO_2 surface scattering layer
50	Dye_Loding (nMol)	Quantity of dye loaded on the TiO_2 surface
51	Exposure_Time (h)	Time used for the dye loading on the TiO_2 surface
52	Absorption (nm)	Maximum peak obtained at absorption spectrum in experimental
53	Emission (nm)	Maximum peak obtained at emission spectrum in experimental
54	V_{OC} (V)	Open circuit voltage value when the light shining on the cell
55	J_{SC} (mAc/m2)	Short circuit current value when the light shining on the cell
56	FF	Fill factor value when the light shining on the cell
57	HOMO-1 (eV)	One state below the highest occupied molecular orbital state
58	HOMO (eV)	Highest occupied molecular orbital state
59	LUMO (eV)	Lowest unoccupied molecular orbital state
60	LUMO+1 (eV)	One state above the lowest unoccupied molecular orbital state
61	Energy gap (eV)	Band gap calculated (ELUMO-EHOMO)
62	Electronegativity (eV)	Electronegativity calculated (-[EHOMO+ELUMO]/2)
63	dipole	Dipole moment of the dye molecules

64	Vertical Energy (cm-1)	Vertical Energy of the molecules in in TD-DFT calculation
65	Computational wavelength(nm)	Maximum peak obtained at absorption spectrum in TD-DFT
66	Oscillator Strength	Oscillator strength of the molecules in in TD-DFT calculation
67	DRDF (eV)	Dye regeneration drying force of the molecules
68	Lifetime (ns)	Life time of the of the molecules in in TD-DFT calculation
69	LHE (%)	Light Harvesting efficiency
70	ΔGinject(eV)	Driving force for efficient electron injection

Table S2. Comparison of data set and type of descriptor used in the literature

Application	**Data set quantity**	**Descriptor quantity**	**R^2 Value (Best)**	**Type of Descriptors used**			**Reference**
				Experimental	**Structural**	**Quantum**	
DSSC	400	31	0.75	No	Yes	Yes	10.1002/jcc.23886
DSSC	600	6	0.8	Some	2D only	No	10.1016/j.saa.2021.120387
DSSC	223/10000	36	0.79	Yes	Yes	No	10.1002/solr.202000110
DSSC	1200	14	0.88	Some	2D only	No	10.1016/j.nanoen.2020.104537
DSSC	354	31	0.91	No	Yes	Yes	10.1109/ACCESS.2018.2850048
Organic Solar Cells	566	13	0.72	Yes	Yes	Yes	10.1021/acs.chemmater.0c02325
Organic Solar Cells	54	970	0.86	Yes	Yes	Yes	10.1016/j.mtcomm.2020.101604
Solar Cells	207	NA	0.99	Yes	No	No	10.1109/ITC-Egypt52936.2021.9513898

Solar Cell Materials	58/1997	1557	0.87	Some	No	Yes	10.1021/acs.chemmater.9b02166
Perovskite	333	NA	NA	Yes	No	No	10.1002/aenm.201901891
Perovskite	540	20+	NA	No	No	Yes	10.1038/s41524-019-0177-0
Perovskite	5158	18	NA	Yes	Yes	Simulated	10.1016/j.nanoen.2020.105380
CIGS solar cells	300	11	0.98	Yes	No	No	10.1016/j.solener.2021.09.031

Table S3. Absorption Maximum modified based on the structure

Code of the dye	No of additional groups	Type of bond (C=C/C≡C)	Addition (nm)	Parent dye absorption maximum (nm)	New dye absorption maximum (nm)
d1029a	1	C=C	30	536	566
d1029b	2	C=C	60	536	596
d1029c	2	C=C	60	536	596
d1067a	1	C=C	30	543	573
d1067b	2	C≡C	60	543	603
d824a	1	C=C	30	512	542
d824b	2	C≡C	60	512	572
d1130a	1	C=C	30	552	582
d1130b	2	C≡C	60	552	612
d279a	0	NO	0	470	470
d279b	0	NO	0	470	470
d1313a	0	NO	0	624	624

Table S4. The dependency of molar extension coefficient and oscillator strength on PCE

Dye Name	ε (M^{-1} cm^{-1})	Oscillator strength	PCE (%)	Reference
DT1	63,779	2.08	5.01	10.1016/j.dyepig.2015.10.022
DT2	59,430	2.04	5.27	10.1016/j.dyepig.2015.10.022
DT3	**70,802**	**2.17**	5.86	10.1016/j.dyepig.2015.10.022

DT4	59,522	2.04	**8.05**	10.1016/j.dyepig.2015.10.022
ZL001	**37,687**	NA	12.8	10.1021/acsenergylett.9b00141
ZL003	29,042	NA	**13.6**	10.1021/acsenergylett.9b00141
ND01	**46,700**	NA	7.46	10.1039/C4CC08539F
ND02	39,600	NA	**8.3**	10.1039/C4CC08539F
ND03	31,700	NA	6.69	10.1039/C4CC08539F
ND04	38,200	NA	7.01	10.1039/C4CC08539F
K1	46,300	0.39	4.25	10.1039/C5TA03807C
K2	57,400	0.47	4.69	10.1039/C5TA03807C
K3	86,100	0.7	4.78	10.1039/C5TA03807C
K4	75,400	0.6	4.82	10.1039/C5TA03807C
K5	82,100	0.56	3.95	10.1039/C5TA03807C
K6	**1,38,600**	0.49	3.82	10.1039/C5TA03807C
K7	1,21,100	0.55	**5.1**	10.1039/C5TA03807C
K8	49,200	0.31	5.15	10.1039/C5TA03807C
YR6	2,79,000	NA	**6.5**	10.1021/cm5045946
T-PA	**3,11,000**	NA	4.6	10.1021/cm5045946
TS3	1,60,000	NA	5.6	10.1021/cm5045946
DTP-PA	2,04,000	NA	2.8	10.1021/cm5045946
DTT-CA	2,31,000	NA	6	10.1021/cm5045946
DTT-PA	**2,59,000**	NA	1.8	10.1021/cm5045946
DTS-CA	2,57,000	NA	**8.9**	10.1021/cm5045946
DTS-PA	2,14,000	NA	5	10.1021/cm5045946

BOX S1. The QSAR model equation of the LR algorithm in ML

PCE(%) = -1.5121 * ATSC3c + 0.3734 * ATSC4c + -0.3759 * ATSC7c +-0.5806* AATSC5i + 0.4958 * AATSC8i + 0.1775 * AATSC0s + 2.6998 * MATS3c + 1.4004 * MATS7c +1.3417 * GATS6c + 1.9504 * GATS4v + -1.2119 * GATS4p + 0.8015 * GATS6i + -1.3404 * GATS3s + -0.6374 * GATS4s + 0.0169 * CrippenLogP + 0.2303 * SHCsatu + 0.619 * SsNH2 + -0.2324 * minHBint4 + 0.4864 * minHsOH + -1.0607 * minHCs+ 0.2218 * maxtsC + -0.0406 * maxtN + 8.2954 * ETA_Epsilon+-29.561*ETA_Epsilo+ -15.3294 * ETA_dEps+ -0.5072 * n4Ring + -0.3997 * nF11Ring + 0.1697 * nF12Ring+ -0.5072 * nT4Ring + -0.3997 * nT11Ring + 0.9832 * nF11HeteroRi+ -0.6378 * Active Area + 0.0156 * Coabsorbent + 0.0362 * Active layer + -0.1234 * Sctering_lay+ 0.0128 * Exp_Time+ 0.0026 * Absorpt Max+ 0.0049 * Emission Max+ -0.1897 * OsciStreg+ 0.6017 * LHE(%) + -0.0822 * Ginject (eV) + 0.0102 * VOC (V) + 0.0147 * JSC (mA)+ -1.3612