Electronic Supplementary Material (ESI) for Sustainable Energy & Fuels. This journal is © The Royal Society of Chemistry 2021

Supplementary Information for

Using High-Throughput Virtual Screening to Explore the Optoelectronic Property Space of Organic Dyes; Finding Diketopyrrolopyrrole Dyes for Dye-Sensitized Water Splitting and Solar Cells

Isabelle Heath-Apostolopoulos,^a Diego Vargas-Ortiz,^b Liam Wilbraham,^a Kim E. Jelfs,^b Martijn A. Zwijnenburg^{a*}

^a Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ, UK.

^b Department of Chemistry, Imperial College London, Molecular Sciences Research Hub, White City Campus, Wood lane, London, W12 0BZ, UK.

* email: <u>m.zwijnenburg@ucl.ac.uk</u>



 $\mathcal{A}_{\mathcal{P}}$ -top-root, after after $\mathcal{A}_{\mathcal{P}}$, $\mathcal{A}_{\mathcal{P}}$, after root, after root, after $\mathcal{A}_{\mathcal{P}}$, root, ap the ap ap we ap we ap an \$ 15 m. ap m. ap B. M. M. M. M. M. D. D. D. M. W. W. D. -an- (i) rai. & ing ing rai, rai, ing ing au an rai. Rai the the the set of the set of the set of the set the the the the the the the the set of the the set of

Figure S1: All building blocks used in this work (explicitly shown hydrogen atoms, other than those connected to nitrogen atoms, indicate carbon atoms where building blocks can be connected to other building blocks in the dyes).



Figure S2: Calibration curve for xTB *vs*. DFT -IPs (left) and -EAs (right) for the combination of dyes and building blocks, the red dashed line is the x=y line and the black line is the line of best fit.



Figure S3: Calibration curves for sTDA/DFTbaby *vs*. TD-DFT optical gaps for the combination of non-sulfur containing dyes and building blocks, the red dashed line is the x=y line and the black line is the line of best fit.



Figure S4: Calibration curve for sTDA vs. TD-DFT optical gaps with (left) and without (right) sulfurcontaining structures.



Figure S5: Calibration curve for xTB v. DFT -IPs (left) and -EAs (right) for the dyes only, the red dashed line is the x=y line and the black line is the line of best fit.



Figure S6: Calibration curve for sTDA/DFTbaby *vs.* TD-DFT optical gaps for the non-sulfur containing dyes only, the red dashed line is the x=y line and the black line is the line of best fit.



Figure S7: 2-dimensional histograms for the uncalibrated -IP, -EA and optical gap values for the ABCBA, ACA and ACA' dyes in water calculated with (IPEA/sTDA)-xTB.



Figure S8: Convex hulls for uncalibrated -IP, -EA and optical gap values for the ABCBA, ACA and ACA' dyes in water solvent calculated with (IPEA/sTDA)-xTB.



Figure S9: 2-dimensional histograms for the uncalibrated -IP, -EA and optical gap values for the ABCBA, ACA and ACA' dyes in benzene calculated with (IPEA/sTDA)-xTB.



Figure S10: Convex hulls for the uncalibrated -IP, -EA and optical gap values for the ABCBA, ACA and ACA' dyes in benzene calculated with (IPEA/sTDA)-xTB.



Figure S11: 2-dimensional histograms for the calibrated -IP, -EA and optical gap values for the ABCBA, ACA and ACA' dyes in benzene calculated with (IPEA/sTDA)-xTB.



Figure S12: Convex hulls for the calibrated -IP, -EA and optical gap values for the ABCBA, ACA and ACA' dyes in benzene calculated with (IPEA/sTDA)-xTB.



Figure S13: Kernel density plots and histograms for the DFT-calculated properties of the building blocks with the value for diketopyrrolopyrrole (DPP) for each property indicated by a black dashed line.



Figure S14: ACA' dyes with the smallest optical gap (left) and fundamental gap (right) relative to the minimum gap out of the corresponding ACA and A'CA' dyes.



Figure S15: Convex hulls for building block shown in Table 2 (a) – top corner.



Figure S16: Convex hulls for building block shown in Table 2 (b) – bottom corner.



Figure S17: Convex hulls for building block shown in Table 2 (c) – bottom corner.



Figure S18: Convex hulls for building block shown in Table 2 (d) – middle region.



Figure S19: Convex hulls for building block shown in Table 2 (e) – middle region.



Figure S20: Convex hulls for building block shown in Table 2 (f) - middle region.

Fitting Data	Property	Slope	Intercept
Dyes + building blocks	IP	0.9044	-0.1399
	EA	0.873	-0.0660
	Δo(sTDA)	0.9735	0.2703
	∆o(DFTbaby)	1.3238	0.8929
Dyes	IP	0.8194	-0.5718
	EA	0.9153	0.1478
	Δo(sTDA)	0.9435	0.2068
	∆o(DFTbaby)	1.1587	0.6944

Table S1: Fitting data for the calibration performed on the combined dyes and building blocks dataset (D+M), and the dyes (D) only dataset in water (non-sulfur in case of Δo).

Table S2: MAE values for the different fits performed in water.

Fitting parameters from Data fitted		IP MAE / eV	EA MAE / eV	Δ _o MAE / eV
Dyes + Building Blocks	Dyes + Building Blocks	0.16	0.13	0.23
Dyes	Dyes + Building Blocks	0.19	0.16	0.54
Dyes + Building Blocks (no sulfur)	Dyes + Building Blocks	0.19	0.18	0.23
Dyes (no sulfur)	Dyes + Building Blocks	0.20	0.16	0.26
Dyes + Building Blocks	Dyes	0.14	0.12	0.23
Dyes	Dyes	0.13	0.12	0.20
Dyes + Building Blocks (no sulfur)	Dyes	0.21	0.21	0.25
Dyes (no sulfur)	Dyes	0.14	0.12	0.21
Dyes + Building Blocks (no sulfur)	Dyes + Building Blocks (no sulfur)	0.19	0.15	0.20
Dyes (no sulfur)	Dyes + Building Blocks (no sulfur)	0.22	0.17	0.26
Dyes + Building Blocks (no sulfur)	Dyes (no sulfur)	0.23	0.18	0.17
Dyes (no sulfur)	Dyes (no sulfur)	0.12	0.09	0.12

Fitting Data	Property	Slope	Intercept
	IP	0.7971	-0.8852
Dyes	EA	0.8810	0.3957
	Δo(sTDA)	1.3065	-0.7695

Table S3: Fitting data for the calibration performed on dyes (D) only dataset in benzene.

Table S4: MAE values for the different fits performed in benzene.

Fitting parameters from	Data fitted	IP MAE / eV	EA MAE / eV	Δ _o MAE / eV
Dyes + Building Blocks Dyes + Building Blocks		0.22	0.13	0.24
Dyes	Dyes + Building Blocks	0.33	0.13	0.58
Dyes + Building Blocks (no sulfur)	3uilding Blocks Dyes + Building Blocks 10 sulfur)		0.13	0.25
Dyes (no sulfur)	Dyes + Building Blocks	0.35	0.12	0.34
Dyes + Building Blocks	Dyes	0.23	0.12	0.25
Dyes	Dyes	0.14	0.11	0.23
Dyes + Building Blocks (no sulfur)	Building Blocks Dyes o sulfur)	0.28	0.12	0.29
Dyes (no sulfur)	Dyes	0.15	0.11	0.27
Dyes + Building Blocks (no sulfur)	Dyes + Building Blocks (no sulfur)	0.23	0.13	0.22
Dyes (no sulfur)	Dyes + Building Blocks (no sulfur)	0.43	0.13	0.34
Dyes + Building Blocks	Dyes			
(no sulfur) (no sulfur)		0.34	0.11	0.21
Dyes	Dyes	0.10	0.00	0.14

Property	ID	No anchor	Phosphonic acid	Carboxylic acid	Cyanocarboxylic acid
IP	37382	-5.48	-5.25	-5.23	-5.26
	4386	-5.66	-5.51	-5.53	-5.56
	7121	-5.19	-5.66	-5.69	-5.67
	aca_174	-5.24	-5.67	-5.76	-5.69
	aca_190	-5.65	-5.34	-5.43	-5.39
	acb_2361	-5.75	-4.59	-4.61	-4.63
	acb_296	-5.45	-4.26	-4.29	-4.24
	acb_2361	-5.75	-4.59	-4.66	-4.62
	acb_296	-5.45	-4.30	-4.29	-4.29
	37382	-4.00	-3.38	-3.35	-3.42
	4386	-3.69	-3.95	-3.98	-4.00
	7121	-3.36	-3.70	-3.79	-3.74
	aca_174	-3.06	-3.45	-3.60	-3.64
EA	aca_190	-3.40	-3.09	-3.19	-3.32
	acb_2361	-3.66	-3.99	-3.95	-4.22
	acb_296	-3.00	-3.83	-3.76	-4.00
	acb_2361	-3.66	-3.91	-4.11	-4.00
	acb_296	-3.00	-3.75	-3.94	-3.89
Optical gap	37382	2.34	2.56	2.58	2.60
	4386	2.63	2.35	2.34	2.33
	7121	2.57	2.59	2.60	2.65
	aca_174	2.84	2.73	2.70	2.69
	aca_190	2.75	2.83	2.82	2.78
	acb_2361	2.71	3.99	3.95	4.22
	acb_296	2.98	3.83	3.76	4.00
	acb_2361	2.71	3.91	4.11	4.00
	acb_296	2.98	3.75	3.94	3.89

Table S5: Effect of three different anchoring group on the properties of a selection of dyes from each dye class in water.