

# PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

## Data Mining with Molecular Design Rules Identifies New Class of Dyes for Dye-Sensitised Solar Cells

Jacqueline M. Cole<sup>a,b\*</sup>, Kian Sing Low<sup>a</sup>, Hiroaki Ozoe<sup>c</sup>, Panagiota Stathi<sup>d</sup>, Chitoshi Kitamura<sup>c</sup>, Hiroyuki Kurata<sup>e</sup>, Petra Rudolf<sup>d</sup>, Takeshi Kawase<sup>c</sup>

<sup>a</sup> Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge, CB3 0HE, UK.

<sup>b</sup> Argonne National Laboratory, 9700 S Cass Avenue, Argonne, IL 60439, USA.

<sup>c</sup> Graduate School of Engineering, University of Hyogo, 2167 Shosha, Himeji, Hyogo 671-2280, Japan

<sup>d</sup> Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747AG Groningen, The Netherlands

<sup>e</sup> Department of Chemistry, Graduate School of Science, Osaka University, Machikaneyama-Cho 1-1, Toyonaka, Osaka, 560-0043, Japan

Author for Correspondence (J. M. Cole): [jmc61@cam.ac.uk](mailto:jmc61@cam.ac.uk)

### Abstract

A major deficit in suitable dyes is stifling progress in the dye-sensitised solar cell (DSC) industry. Materials discovery strategies have afforded numerous new dyes; yet, corresponding solution-based DSC device performance has little improved upon 11% efficiency, achieved using the N719 dye over two decades ago. Research on these dyes has nevertheless revealed relationships between the molecular structure of dyes and their associated DSC efficiency. Here, such structure-property relationships have been codified in the form of molecular dye design rules, which have been judiciously sequenced in an algorithm to enable large-scale data mining of dye structures with optimal DSC performance. This affords, for the first time, a DSC-specific dye-discovery strategy that predicts new classes of dyes from surveying a representative set of chemical space. A lead material from these predictions is experimentally validated, showing DSC efficiency that is comparable to many well-known organic dyes. This demonstrates the power of this approach.

## 1. Introduction

There is currently a world-wide deficit in suitably efficient materials for photovoltaic applications, despite their crucial role in realizing new generation power sources. This is particularly relevant to the field of dye-sensitized solar cells (DSCs), which is one of the strongest contenders for next generation photovoltaic technology. Although less efficient than Silicon-based solar cells, DSCs are far more cost-effective to the extent that their price-performance ratio achieves ‘grid-parity’ status, *i.e.* they are competitive with fossil-fuel energy production.

DSCs arose from the pioneering work of O'Regan and Grätzel<sup>1</sup> and are based upon a chemical photosynthetic redox process. The molecular dye is a particularly critical component of a DSC since it is responsible for both the light-harvesting of energy from the sun, and electron injection that initiates the chemical redox reaction of the solar cell. Consequently, there have been extensive efforts to discover new materials that outperform the ruthenium-based dye, N719 (10.4% solar-cell efficiency<sup>2</sup>), which remained the world's most efficient dye for DSCs for over 20 years. A major breakthrough finally came in 2011 with the report<sup>3</sup> of a zinc porphyrin-based dye which broke this world record, affording 12.3 % under 1 sun illumination, when coupled with a cobalt-based electrolyte in a Grätzel cell. Earlier this year, a solid-state DSC employing a lead-iodide based pigment surpassed solution-based DSC records with a 15% efficiency<sup>4</sup>.

This said, metal-based dyes are extremely expensive relative to organic dyes, and are less environmentally sound. Given that price-to-performance ratio governs ‘grid parity’, economic incentives are heavily motivating the development of efficient organic dye alternatives. In this regard, carbazole-based organic dyes with ferrocene-based electrolytes have received particular acclaim, demonstrating 7.5% solar-cell efficiency<sup>5</sup>; 10.3% DSC efficiency has even been

reported using a thiophene-based dye, albeit in a highly volatile electrolyte<sup>6</sup>. But clearly there is still a long way to go in this area.

Dye materials discovery is still based on serendipity or iterative chemical substitution methodologies that are based on an *a priori* known DSC active chemical exemplar. Such methods are very valuable within a specific area of synthetic chemistry; see for example, the work by Lee *et al.*<sup>7</sup> on designing a porphyrin class of dyes by systematic chemical substitution. However, this approach can only ever afford improvements *within* a currently known class of DSC dyes. Computational efforts have similarly restricted themselves to predicting or rationalising new dyes within the framework of an existing class of chemical dyes. Such methods therefore have no capacity to reveal entirely new classes of suitable DSC dyes. If one is going to make a step-change in dye performance in DSCs, a distinct shift in the approach to dye design is therefore urgently needed that specifically targets the materials discovery of new classes of dyes. To this end, we present a systematic large-scale data-mining procedure that successfully identifies new classes of dyes, whose intrinsic molecular characteristics predict good DSC functionality. Its underlying search methodology relies on a judicious encoding of molecular design rules for dyes that have recently begun to develop in the DSC field.

In that regard, molecular design rules are developing via the application of chemical intuition to empirical surveys of DSC dye discovery in the literature. Certain chemical groups and molecular architectures that commonly feature in a well-performing DSC dye have been spotted and translated into generic design principles<sup>8,9</sup>. For example, the importance of having an electron donor (D) and acceptor (A), separated via a  $\pi$ -conjugating framework, in an organic DSC dye was quickly realized. This D- $\pi$ -A molecular architecture enables a charge-separated resonance structure, from which an electron-hole pair can be generated. Within this molecular design

structure, A then has the additional role of injecting the electronic charge into the conduction band of the TiO<sub>2</sub> semiconductor electrode, to which the dye is adsorbed via A. Thus, the electrical circuit is initiated.

Recent work has shown that A need not also be the substituent that adsorbs to TiO<sub>2</sub>, with the advantage that a natural adsorbing substituent (Ads), most commonly a carboxyl group, is not the best acceptor<sup>10,11</sup>. To this end, a D- $\pi$ -A- $\pi_{\text{phenyl}}$ -Ads chemical motif has demonstrated in excess of 6 x DSC efficiency improvements over its analogous D- $\pi$ -A molecular architecture<sup>12</sup>. The judicious choice of a phenyl ring to separate A and Ads provides sufficient aromatic ring stabilization to inhibit the back-transfer of electronic charge from TiO<sub>2</sub> to A, thereby circumventing undesirable electron recombination<sup>9</sup>.

This decoupling of A and Ads has exciting implications for systematic molecular design. For the first time, we have the opportunity to discover new classes of dyes by searching chemical space for molecules that contain strong D- $\pi$ -A moieties, devoid of the constraints that Marcus theory<sup>10</sup> imposes on Ads.

## 2. Algorithmic Dye Prediction Strategy

This work realizes this opportunity by automatically sampling a diverse range of 118,465 organic molecules with an algorithm (see Fig. 1), designed to first identify molecules containing D- $\pi$ -A moieties, and then rank them according to D- $\pi$ -A strength using bond-length alternation<sup>13</sup> and molecular dipole moment as criteria.

To this end, the Cambridge Structural Database<sup>14</sup> provided a naturally diverse set of chemical space (118,465 organic molecules) for this study. Data quality control was pre-assured using the crystallographic R1 < 0.07 statistic as a compound inclusion threshold for data mining, while all polymeric and ionic salts were excluded. Semi-empirical (AM1) calculations on all 118,465

molecules provided their molecular dipole moments,  $\mu$ , automatically by coupling the search engine of the database to MOPAC<sup>15</sup>. To ensure that prospective dyes have sufficient charge-separation, a  $\mu > 5$  Debye data filter was imposed. This minimum threshold was selected for two reasons. Firstly,  $\mu \sim 6-7$  Debye is commonly observed in mid-sized DSC dyes such as coumarins<sup>16</sup>, and  $\mu$  generally increases when creating larger dye motifs such as D149<sup>17</sup>. Secondly, this minimum corresponds closely to the charge separation of an electron and proton separated by 1 Å ( $1 \text{ eÅ} = 4.8$  Debye), suggesting that a molecule with  $\mu > 5$  Debye can sustain an electron-hole pair separated by at least one organic chemical bond length (typically  $\sim 1-1.5$  Å). The resulting 24,903 molecules were then passed through a second data filter that exclusively selects dyes with good D- $\pi$ -A intramolecular charge transfer (ICT) characteristics. To this end, Hammett constants<sup>18</sup> were employed to identify molecules containing an electron donor, D, and the bond-length alternation (BLA) parameter<sup>13</sup> then determined the level of  $\pi$ -delocalisation between D and A; a BLA  $< 0.04$  Å was used as the threshold for good ICT. This series of calculations employed an algorithm similar to that of Cole & Weng<sup>15</sup>, but with one important exception: the electron acceptor, A, was not defined by Hammett constants; rather, BLA was calculated sequentially along all bond paths emanating from D until a BLA  $> 0.04$  Å was exceeded; the atom at which this inequality falters is deemed to be an acceptor, A. If BLA  $< 0.04$  Å was not realised after the second BLA calculation along the bond path from D, it is discarded as a donor in that molecule. This ICT calculation is illustrated in Figure 2 using, as an exemplar, the molecular dye precursor that ultimately identified the target material.

Trivial molecules and known classes of D- $\pi$ -A dye motifs were then manually removed from the resulting 526 initial short-list of molecules. SciFinder<sup>®</sup> acted as a further filter to target molecules where a known synthetic route for realizing a target with a *para*-phenyl group

substituted to A was available; this provided a suitable electronic pathway to generate a D- $\pi$ -A- $\pi$ -X motif. The 10 most novel classes of chemicals where synthetic feasibility was assured comprised the final short-list. These ‘top 10’ molecules then underwent simulated ‘molecular mutation’ where X was replaced by a cyanoacrylate (Ads) group using DFT calculations. This group was selected as the Ads moiety given its particularly good adsorption characteristics<sup>19</sup>. DFT then assessed the energetic suitability of the resulting D- $\pi$ -A- $\pi$ -Ads target molecules to DSC function. All DFT calculations used a B3LYP functional and a 6-31++G(d,p) basis set. This final step identified **1** (Fig. 1 bottom) as the lead material.

<Insert Figure 1>

<Insert Figure 2>

### 3. Dye Candidate Selection

Compound **1** stood out from the final short-list of molecules via manual inspection of its respective HOMO-LUMO charge-transfer distributions and associated band gaps, generated via Density Functional Theory (DFT) calculations (see SI). A particular attraction of compound **1** was its near-complete shift of charge-density distribution from the D- $\pi$ -A moiety in the HOMO to the  $\pi$ -Ads moiety in the LUMO (Figure 3). On the one hand, this indicates that the charge-separated state is stable so that **1** can avoid undesirable electron recombination from TiO<sub>2</sub>-to-dye electron back transfer. On the other hand, this reveals that Ads in **1** is electron-rich for efficient electron injection into the TiO<sub>2</sub> electrode. The associated HOMO and LUMO energy levels of **1** are also favourable as illustrated in Figure 2; with the LUMO residing  $\sim 0.8\text{eV}$  above the TiO<sub>2</sub> conduction band edge, and the HOMO lying  $\sim 0.8\text{eV}$  below that of the driving voltage of the iodide/triiodide redox couple, the most commonly used DSC electrolyte. A minimum driving potential difference of  $\Delta V \sim 0.2\text{eV}$  is required in each case, to ensure good electron injection<sup>20</sup> and

dye regeneration<sup>21</sup>, respectively. Premature dye regeneration has also been associated with structural aspects of dye molecules; in particular, conjugation length and the type of side chain on the dye have demonstrated relationships<sup>9</sup>. These factors were not considered by the algorithm presented herein. However, it has been shown that such structural influences have only minor effects if  $\Delta V_{\text{electrolyte-HOMO}} > 0.2-0.25 \text{ eV}$ <sup>22</sup> as in **1**.

<Insert Figure 3>

Compound **1** was therefore synthesized (see Methods and SI) in order that its dye performance in a DSC could be experimentally validated. A related compound, **2**, was also synthesised by substituting the  $\pi$ -NMe<sub>2</sub> unit of **1** for a triarylamine group. This enabled a check for possible dye aggregation issues in **1** via a DSC performance comparison of **1** and **2**: triarylamines are proving very popular as bulky donor groups in dyes to offset intrinsic DSC photocurrent losses owing to dye aggregation<sup>25</sup>.

## 4. Experimental Validation of Dye Candidate

### 4.1 Materials Characterisation of **1** and **2**

The molecular structure of **1** (Figure 4) was confirmed by single-crystal X-ray diffraction (see SI for full details). This data also evidenced experimentally the high D- $\pi$ -A intramolecular charge-transfer (ICT), predicted by DFT. To this end, the NMe<sub>2</sub> group of **1** (D) carries a definite  $\delta^+$  charge in its ground-state given that all three N-C bonds display partial double-bonded character, with that of N1-C11 being particularly marked.  $\pi$ -density from the three phenyl rings of **1** conjugates with the intersecting carbon, C1 (A), resulting in its heavily  $\delta^-$  character. In fact, the C1-C2 bond length (1.381 Å) is particularly short since it forms part of a classically bonded quinodimethene unit, whose high bond-length alternation<sup>13</sup> somewhat isolates it from the primary ICT pathway; this takes place in the delocalised regions of a molecule, *i.e.* via the other

two (more delocalised) phenyl rings as indicated via HOMO-to-LUMO charge redistributions (Fig 3b). In the ground-state, the charge is essentially held within the D- $\pi$ -A moiety between the highly  $\delta^+$  and  $\delta^-$  atom centres;  $\pi$ -conjugation from the adjoining  $\pi_{\text{phenyl-Ads}}$  moiety is also hindered, being it twisted out of the plane. Fig 3 shows that photo-absorption enables the ICT of **1** to channel almost entirely through to the  $\pi_{\text{phenyl-Ads}}$  moiety, thereby rendering a stable charge-separated excited state. This suggests good resistance of **1** to premature dye regeneration.

<Insert Figure 4>

The crystal structure of **2** was not available. However, one assumes that the substitution of NMe<sub>2</sub> for NPh<sub>2</sub> would place the nitrogen in a less  $\delta^+$  environment owing to the lower electron donating ability of Ph relative to Me; as such, **2** is expected to possess slightly less pronounced charge-separation. Indeed, this effect is manifest in the DFT charge-distributions of Fig 3.

Prior to incorporating **1** or **2** into a full DSC device, their optical absorption characteristics, when adsorbed onto an open TiO<sub>2</sub> electrode, were evaluated. UV/vis absorption spectroscopy revealed a good panchromatic nature of both **1** and **2**, in methanol solution and when dye-sensitized onto TiO<sub>2</sub> nanoparticles (Fig. 4a). The minor red/blue shifting observed between the solution and dye-sensitized TiO<sub>2</sub> spectra indicate successful dye adsorption to TiO<sub>2</sub>. This was further evidenced by X-ray Photoelectron Spectroscopy (XPS) of dye-sensitized TiO<sub>2</sub> as compared with bare TiO<sub>2</sub>; the 2p Ti XPS peaks displayed a -0.2-0.3 eV shift upon dye-sensitization for both **1** and **2**, as is characteristic of adsorption of organic chromophores to TiO<sub>2</sub><sup>26,27</sup>. Corresponding shifts in 1s O XPS spectra, reminiscent of surface Ti...HO-OC interactions in **1** and **2**, were also identified (see Supplementary Information for further XPS details).

Time-dependent (TD)-DFT calculations permitted the assignment of optical transitions to the three main optical absorption peaks,  $\lambda_{\text{max}}$ , of **1** and **2**, respectively: HOMO to LUMO charge-

transfer excitation (499, 481 nm); a HOMO-1 to LUMO  $\pi$ - $\pi^*$  transition (372, 371 nm); a HOMO to LUMO+1  $\pi$ - $\pi^*$  transition (321, 300 nm). The greater amount of charge shifting from the HOMO in **2**, relative to **1**, results in its slightly lower  $\lambda_{\max}$  values for the first and third peaks.

#### 4.2 DSC Device Fabrication and Testing of **1** and **2**

Full DSC device fabrication and testing was then undertaken, employing the doctor-blade method to lay the nanoporous TiO<sub>2</sub> and Pt electrodes onto transparent conducting glass substrates, sensitising the TiO<sub>2</sub> by its overnight soaking in a methanol solution of the dye, and heat-sealing the two electrodes together using a thermoplastic resin (Surlyn), while placing a I<sup>-</sup>/I<sub>3</sub><sup>-</sup> redox couple betwixt to act as an electrolyte. A full description of these methods is given in the Supplementary Information.

The photocurrent-voltage (J-V) characteristics under standard solar simulation reporting conditions (AM 1.5G illumination at 1000 W/m<sup>2</sup> at 298K) were determined for each DSC, yielding its power conversion efficiency (PCE,  $\eta$ ) according to:

$$\eta = P_{\text{out(max)}}/P_{\text{in}} = J_{\text{sc}} \times V_{\text{oc}} \times \text{FF} / P_{\text{in}}$$

where  $J_{\text{sc}}$  is the short-circuit current density,  $V_{\text{oc}}$  is the open-circuit voltage, FF is the fill factor (maximum power observed in the J-V curve (Fig. 4b) divided by  $J_{\text{sc}}V_{\text{oc}}$ ), and  $P_{\text{in}}$  is the intensity of simulated sunlight impinging on the active area of the solar cell.

<Insert Table 1 and Figure 5>

Results in Table 1 show that DSC devices employing **1** and **2** deliver ~30% of the solar-cell performance efficiency realized using N719. The results of **1** and **2** were normalized to in-house results on DSCs containing N719, since interpretations based on this  $\eta_{\text{Dye}} : \eta_{\text{N719}}$  ratio, rather than  $\eta$ , obviate the common variability in DSC efficiency reporting. Such variability is caused by the

renowned acute sensitivity of  $J_{sc}$  to the fabrication conditions<sup>28</sup>; while  $J_{sc(\text{dye})}:J_{sc(\text{N719})}$  is generally consistent between duplicate runs, as in this case. Absolute  $V_{oc}$  and FF values are consistent and typical for organic DSCs bearing an  $I/I_3$  electrolyte<sup>25</sup>; thus good electron injection and DSC device quality are implied, respectively.

Corresponding Incident-Photon-to-Current Efficiency (IPCE) measurements were undertaken from 375 to 700 nm in 5 nm increments. Results (Fig. 4c) on each dye show a red shift relative to the dye in solution or as adsorbed on  $\text{TiO}_2$  (open electrode), as is commonly observed<sup>29</sup>. **1** retains a greater panchromatic absorption over **2** in the DSC by an extent of  $\sim 45$  nm at long-wavelengths, judging from the IPCE spectra.

The same  $\eta_{\text{dye}}:\eta_{\text{N719}}$  ratio observed for both **1** and **2** suggests that **1** exhibits no significant dye aggregation problems; recall that the presence of the bulky donor group, triarylamine, in **2** was used to test for this issue. The molecular structure of **1** holds a possible explanation for its lack of dye aggregation, despite hosting a non-bulky donor group: its three rings are somewhat twisted relative to each other (see SI), affording a distinctly 3-D molecule.  $\pi\dots\pi$  intermolecular interactions that typically cause dye aggregation<sup>9</sup> in planar dye molecules therefore cannot easily form. The two *tert*-butyl groups in **1** will further hinder such intermolecular interactions with neighbouring molecules.

There are two common structural causes of DSC performance losses that were not considered by the molecular design approach presented herein. One of these is electron recombination owing to undesirable  $\text{TiO}_2$ -to-electrolyte electron back transfer. To some extent, this has been implicitly tested via experimental validation of **2** since the triarylamine offers a hydrophobic substitute for the more commonly used long alkyl chains that physically block the approach of the electrolyte to the  $\text{TiO}_2$  surface. Finally, long-term chemical and photo-chemical stability is naturally a key

asset of a dye for DSCs. While it is too soon to assess the long-term photo-chemical stability of **1** or **2**, they present unusually high melting points (182 and 208°C) for a small organic material which suggests good chemical stability.

## 5. Conclusions

In summary, we have discovered a new class of dyes for DSC applications, which were predicted from large-scale data-mining that encodes molecular design rules. Two dye specifications, **1** and **2**, exhibit solution-based DSC efficiencies of  $\sim 1/3^{\text{rd}}$  that of those containing world's best performing dyes. This is very encouraging, given that these findings emanate from the first experimental validation phase of this algorithmic approach to DSC dye materials-by-design. Moreover, **1** does not appear to suffer from any of the common drawbacks of DSC dyes, as discussed above. Therefore, this class of dyes now has the potential to be fully developed by the manual, but chemical intuitive substitution strategies, as has transpired for many classes of *a priori* known organic dyes; such developments would lead to a brand new range of dyes for DSCs, with the prospect that some new specifications are likely to be more superior than the D- $\pi$ -A- $\pi$ -Ads building blocks, **1** or **2**. Indeed, one would expect better performing dyes to unfold from these ideal 'ICT molecular cores'; in the same way that famous organic dyes such as MK2<sup>30</sup> were generated by extension of its core building block, the MK44 dye. The likes of MK44 had, of course, been identified by means of chemical intuition. This new algorithm-based molecular design rule strategy has the innate strength to successfully predict DSC dye molecular building blocks before even turning to the chemical bench; and its systematic nature already offers many short-cuts for the synthetic chemist; with further refinement, the predictive power and reach of this computational approach stands to deliver even greater dye design foresight and versatility.

New structure-property relationships are continuously evolving from such chemistry, which can also be fed into the next steps in the algorithmic prediction for DSC applications. To this end, next steps in algorithmic development will not only embed further dye design rules based on structure, but will combine both structural and energetic considerations in a fully automated and concerted manner. This will naturally extend into the prediction of not only the dye, but all DSC material components simultaneously, since their roles are so entangled; this has the tantalising scope of realising new electrolytes, electrodes, or substrates as well as the dyes themselves. Ultimately, we seek an algorithmic abstraction of the material influences on DSC performance efficiency, whose verisimilitude to its device functionality is sufficient to yield a superior DSC technology.

## 6. Experimental Methods

**Synthesis.** **1** and **2** were synthesized by reacting 4-bromo-2,6-di(*t*-butyl)-1-(trimethylsiloxy)benzene with 4-*N,N*-dialkylamino-4'-formylbenzophenone ethylene acetal (where alkyl = methyl<sup>31</sup> or phenyl<sup>32</sup> in the presence of *n*-butyl lithium in tetrahydrofuran at -78°C, followed by a Knoevenagel condensation reaction with cyanoacetic acid in acetonitrile<sup>14</sup>. Full synthetic details are given in the Supplementary Information.

**DSC fabrication and testing.** DSCs were fabricated using Dyesol TiO<sub>2</sub> DSL 18NR-T paste. The TiO<sub>2</sub> was deposited onto cleaned fluorine-doped tin oxide (FTO) conductive glass (Solaronix S. A.) via the doctor-blade method. The composite was then successively sintered for 10 minutes at 100°C, 10 minutes at 150°C, 30 minutes at 325°C, 5 minutes at 400°C, and 30 minutes at 300°C, before allowing to cool to 70°C where it was dipped into a 0.5 mM dye solution using methanol

for **1** and **2**, and a 1:1 ratio of acetonitrile:tert-butanol solvent for N719. The platinum counter-electrode was prepared by spin coating a  $\text{H}_2\text{PtCl}_6$  solution (52 mM in isopropanol) onto the FTO coated glass and then annealing at 450°C. The electrolyte comprised 50 mM iodide/triiodide in acetonitrile (Iodolyte AN-50 from Solaronix S. A.). The photoelectrode and counter-electrode were sealed together using 25 mm Surlyn. J-V characteristics of the resulting DSCs were determined under the equivalent of 100  $\text{mWcm}^{-2}$  AM 1.5G illumination with a calibrated ABET Sun 2000 solar simulator, corrected for spectral mismatch.

**Supporting Information:** Preparation and Chemical Characterization of **1** and **2**; Density Functional Theory and Time-Dependent Density Functional Theory of **1** and **2**; X-ray Photoelectron Spectroscopy of **1** and **2** Supporting Information is available from XXXXXXXX.

### Acknowledgements

J.M.C. is indebted to the Fulbright Commission for a UK-US Fulbright Scholar Award, and to Argonne National Laboratory where work done was supported by DOE Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. J.M.C. would also like to thank Peter Littlewood at Argonne for helpful discussions. K.S.L. acknowledges the EPSRC for a DTA PhD studentship (EP/P504120/1). Partial financial support also came from the “Top Research School” program of the Zernike Institute for Advanced Materials under the Bonus Incentive Scheme (BIS) of the Netherlands’ Ministry of Education, Science, and Culture and from the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'. This study was also supported by the Hyogo prefecture and the Ministry of Education, Culture, Sports, Science and Technology, Japan.

### Author contributions

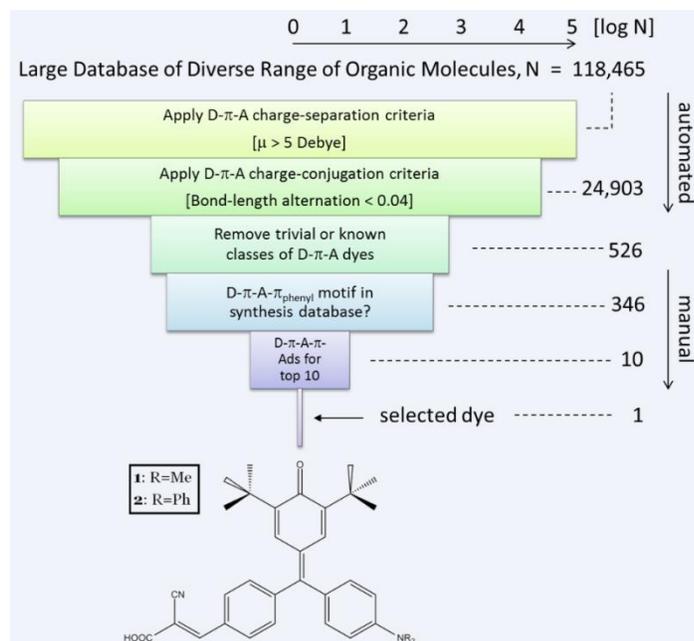
J.M.C. conceived the project, designed and carried out the data-mining work, interpreted the primary results according to DSC molecular design criteria, and drafted the paper. K.S.L. conducted the UV/vis absorption spectroscopy, DSC fabrication and testing, and detailed DFT and TD-DFT calculations on **1** and **2**, under the supervision of J.M.C. P.S. undertook the XPS characterization and associated data interpretation of **1** and **2**, under the direction of P.R. H.O. synthesised **1** and **2** and did the Cyclic Voltammetry; C.K. determined the crystal structure of **1** from X-ray diffraction; H.K. performed the mass spectrometry for **1** and **2**. H.O., C.K., and H.K. were supervised by T.K.

### References

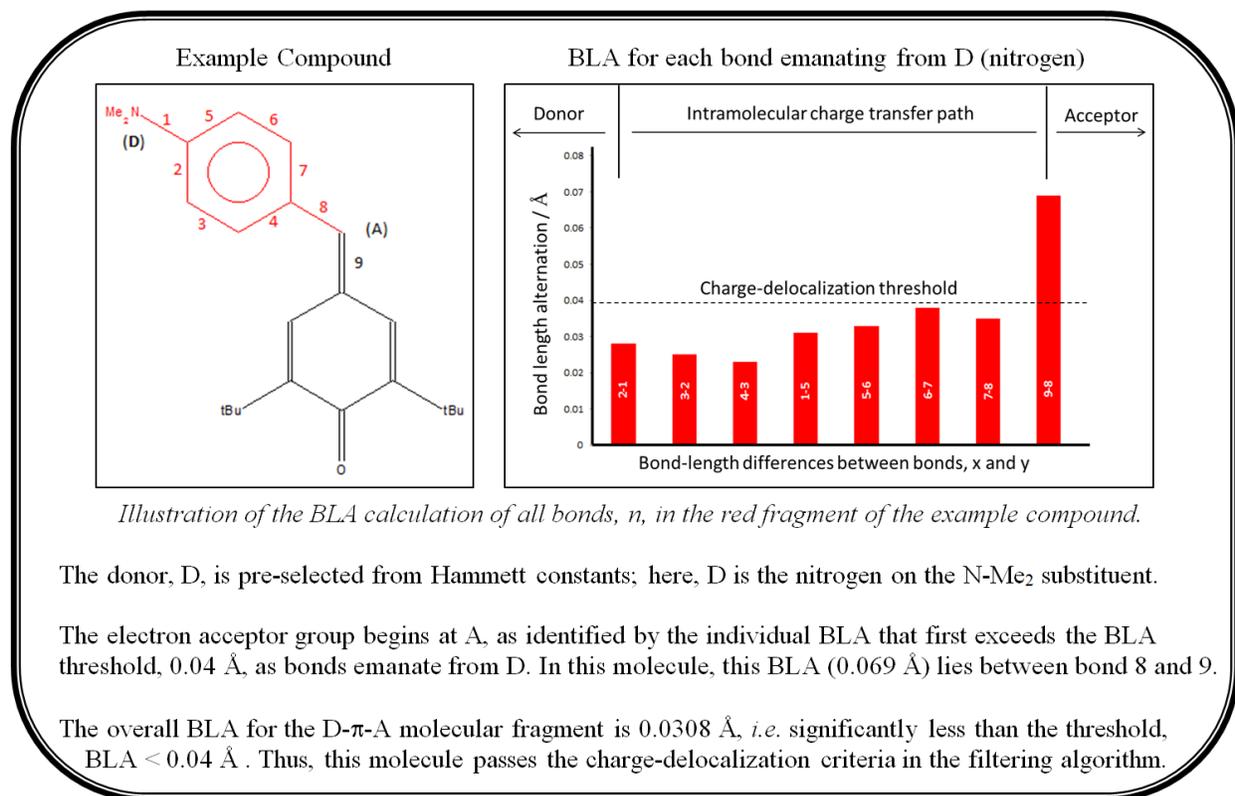
1. B. O'Regan, M. Grätzel, *Nature* 1991, **353**, 737.
2. M. K. Nazeeruddin, A. Kay, I. Rodicio, R. Humphry-Baker, E. Mueller, P. Liska, N. Vlachopoulos, M. Graetzel, *J. Am. Chem. Soc.* 1993, **115**, 6382.
3. A. Yella, H.-W. Lee, H. N. Tsao, C. Yi, A. K. Chandiran, Md. K. Nazeeruddin, E.W.-G. Diao, C.-Y. Yeh, S. M. Zakeeruddin, M. Grätzel, *Science* 2011, **334**, 629.
4. J. Burschka, N. Pellet, S. J. Moon, R. Humphry-Baker, P. Gao M. K. Nazeeruddin, M. Grätzel, *Nature*, 2013, **499**, 316.
5. T. Daeneke, T. H. Kwon, A. B. Holmes, N. W. Duffy, U. Bach, L. Spiccia, *Nat. Chem.* 2011, **3**, 211.
6. W. Zeng, Y. Cao, Y. Bai, Y. Wang, Y. Shi, M. Zhang, F. Wang, Y. Pan, P. Wang, *Chem. Mater.* 2010, **22**, 1915.

7. C.-W. Lee, H.-P. Lu, C.-M. Lan, Y.-L. Huang, Y.-R. Liang, W.-N. Yen, Y.-C. Liu, Y.-S. Lin, E. W.-G. Diau, C.-Y. Yeh, *Chem. Eur. J.* 2009, **15**, 1403.
8. A. Mishra, M. K. R. Fischer, P. Baeuerle, *Angew. Chem., Int. Ed.* 2009, **48**, 2474.
9. B.-G. Kim, K. Chung, J. Kim, *Chem. Eur. J.* 2013, **19**, 5220.
10. R. A. Marcus, *J. Chem. Phys.* 1956, **24**, 979.
11. H. Tian, X. Yang, R. Chen, Y. Pan, L. Li, A. Hagfeldt, L. Sun, *Chem. Commun.* 2007, 3741.
12. S. Haid, M. Marszalek, A. Mishra, M. Wielopolski, J. Teuscher, J. E. Moser, R. Humphry-Baker, S. M. Zakeeruddin, M. Grätzel, P. Baeuerle, *Adv. Funct. Mater.* 2012, **22**, 1291.
13. C. B. Gorman, S. R. Marder, *Proc. Natl. Acad. Sci* 1993, **90**, 11297.
14. F. H. Allen, *Acta Cryst. B* 2002, **58**, 380.
15. J. M. Cole, Z. F. Weng, *Adv. Mat. Res.* 2010, **123-125**, 959.
16. M. S. Zakerhamidi, A. Ghanadzadeh, M. Moghadam, *Spectrochimica Acta Part A* 2011, **78**, 961.
17. A. El-Zohry, A. Orthaber, B. Zietz, *J. Phys. Chem. C* 2012, **116**, 26144.
18. L. P. Hammett, *J. Am. Chem. Soc.* 1937, **59**, 96.
19. D. P. Hagberg, T. Marinado, K. M. Karlson, K. Nonomura, P. Qin, G. Boschloo, T. Brinck, A. Hagfeldt, L. Sun, *J. Org. Chem.* 2007, **72**, 9550.
20. K. Hara, T. Sato, R. Katoh, A. Furube, Y. Ohga, A. Shinpo, S. Suga, K. Sayama, H. Sugihara, H. Arakawa, *J. Phys. Chem. B* 2003, **107**, 597.
21. S. Wenger, P. A. Bouit, Q. Chen, J. Teuscher, D. Di Censo, R. Humphry-Baker, J. E. Moser, J. L. Delgado, N. Martin, S. M. Zakeeruddin, M. Grätzel, *J. Am. Chem. Soc.* 2010, **132**, 5164.

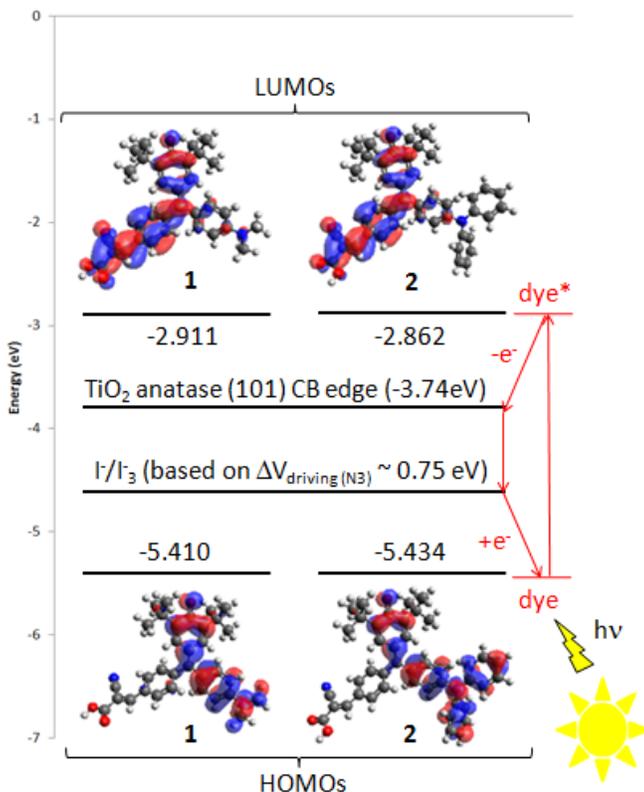
22. T. Daeneke, A. J. Mozer, Y. Uemura, S. Makuta, M. Fekete, Y. Tachibana, N. Koumura, U. Bach, L. Spiccia, *J. Am. Chem. Soc.* 2012, **134**, 16925.
23. F. Labat, C. Adamo, *J. Phys. Chem. C* 2007, **111**, 15034.
24. G. Boschloo, A. Hagfeldt, *Acc. Chem. Res.* 2009, **42**, 1819.
25. M. Liang, J. Chen, *Chem. Soc. Rev.* 2013, **42**, 3453.
26. E. M. J. Johansson, M. Hedlund, H. Siegbahn, H. Rensmo, *J. Phys. Chem. B* 2005, **109**, 22256.
27. E. M. J. Johansson, M. Hedlund, M. Odellius, H. Siegbahn, H. Rensmo, *J. Chem. Phys.* 2007, **126**, 244303.
28. C. Y. Jen, L. V. Munukutla, S. Radhakrishnan, A. M. Kannan, A. Htun, *J. Nanosci. Nanotechnol.* 2012, **12**, 1829.
29. M. Grätzel, *Inorg. Chem.* 2005, **44**, 6841.
30. N. Koumura, Z-S. Wang, S. Mori, M. Miyashita, E. Suzuki, K. Hara, *J. Am. Chem. Soc.* 2006, **128**, 14256.
31. T. Kumagai, T. Anki, T. Ebi, A. Konishi, K. Matsumoto, H. Kurata, T. Kubo, K. Katsumoto, C. Kitamura, T. Kawase, *Tetrahedron* 2010, **66**, 8968.
32. G. P. Schiemenz, G. Stein, *Tetrahedron* 1970, **26**, 2007.



**Figure 1. Step-wise filtering algorithm to identify trial DSC dye candidates from large sets of chemical data.** A series of DSC molecular design criteria were sequentially applied to an initial set of 118,465 molecules: (a) compounds containing a D- $\pi$ -A structural motif were automatically sifted from the initial chemical set via charge (i) separation (threshold  $\mu > 5$  Debye; 24903 compounds) and (ii) delocalisation (bond-length alternation) considerations, yielding an initial short-list of 526 molecules. (b) Manual intervention then removed (i) trivial or known D- $\pi$ -A classes of molecules and (ii) those where there are no examples of D- $\pi$ -A- $\pi_{\text{para}}$  in a synthesis database; (c) the 10 most novel chemical motifs in the resulting final short-list underwent ‘molecular mutation’ to create the target D- $\pi$ -A- $\pi$ -Ads (Ads = cyanoacrylate) compounds; DFT was employed to test their energetic feasibility for application as DSC dyes. The lead material, **1**, was identified; it was then synthesized, so that its DSC dye prospects could be experimentally validated.

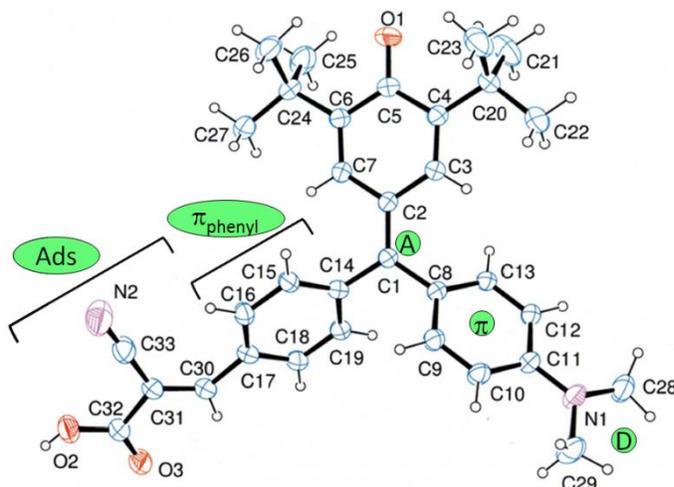


**Figure 2. Example of BLA calculation encoded to automatically evaluate D- $\pi$ -A intramolecular charge transfer.** The compound chosen as the exemplar is the ‘molecular mutation’ precursor to the lead material, **1**.



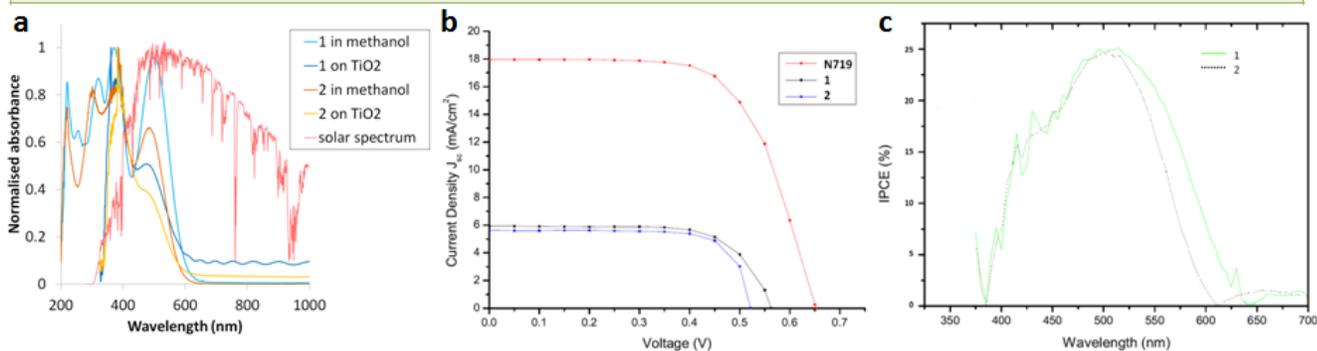
**Figure 3.** The energy relationship between frontier molecular orbitals of **1** and **2** and key quantum levels in other DSC components (left), which drives device operation (right, in red). DFT calculations afforded highest-occupied and lowest-unoccupied molecular orbitals (HOMOs and LUMOs) of **1** and **2**. These computational results are backed up by Cyclic Voltammetry experiments (see Supplementary Information). Results show that a near-complete shift of orbital population from the D- $\pi$ -A group to the  $\pi_{\text{phenyl}}$ -Ads moiety occurs in the HOMO-to-LUMO transition for both dyes. This prospects good electron injection efficiency and minimal dye-to-TiO<sub>2</sub> electron back transfer. The 0.8 eV energy gap between the LUMO of the dye and TiO<sub>2</sub> conduction-band edge, assuming preferential dye adsorption to the (101) face of TiO<sub>2</sub> anatase<sup>15</sup>, is also favorable for electron injection. A similar driving voltage between the HOMO

of each dye and the redox level of the  $\Gamma/\Gamma_3$  electrolyte is also indicative of good dye regeneration; the redox level is calculated with reference to the maximum 0.75 eV  $\text{TiO}_2$ -to- $\Gamma/\Gamma_3$  driving voltage, c.f. for the cationic N719, i.e.  $\text{N}3^{16}$ . All of these factors suggest that **1** and **2** would be attractive DSC dyes.



**Figure 4. Molecular structure of 1.** Bond lengths evidence (i) high levels of D- $\pi$ -A intramolecular charge transfer (ICT), as predicted, which enable stable charge-separation; (ii) low levels of  $\pi$ -conjugation in the quinodimethene group which shows that this moiety is largely isolated from ICT, i.e. it does not play a significant electronic role in DSC function; (iii)  $\pi$ -conjugation is somewhat hindered between the  $\pi_{\text{phenyl}}$ -Ads and D- $\pi$ -A moieties; indeed, the respective phenyl rings lie out-of-plane with respect to each other, by  $72.1(1)^\circ$ . As such, the likelihood of undesirable  $\text{TiO}_2$ -to-dye electron back-transfer is diminished.

Dye	$V_{oc}$ (V)	$J_{sc}$ (mA/cm <sup>2</sup> )	FF	$\eta$ (%)	$\eta_{dye}:\eta_{N719}$ (%)
<b>1</b>	0.56	5.92	0.69	2.32	30.8
<b>2</b>	0.52	5.62	0.75	2.20	29.2
N719	0.65	17.96	0.64	7.54	100



**Table 1 and Figure 5. Optoelectronic and photovoltaic characteristics of 1 and 2.** Table 1

shows that the DSC performance efficiency for devices containing **1** and **2** can be ascribed to 30% of that of N719. Interpretations are based on this  $\eta_{Dye} : \eta_{N719}$  ratio rather than  $\eta$  in order to avoid the common variability in absolute DSC efficiency reporting<sup>21</sup>. The DSC device fabrication and testing for **1**, **2** and N719 were all obtained from the same in-house fabrication and testing methods; conducted in triplicate to ensure good consistency.  $V_{oc}$  and FF values evidence efficient electron injection and DSC device quality, respectively. Figure (a) UV/vis absorption spectra of **1** and **2** in solution and adsorbed onto TiO<sub>2</sub>; this show that **1** and **2** capture well the solar spectrum (overlaid) up to ~600 nm, as is common in DSC dyes. (b) J-V curves from the DSC photovoltaic tests. (c) Corresponding IPCE plots of **1** and **2** in the functioning DSC.

## Table of Contents Graphic for PCCP

