

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.

Artificial evolution of coumarin dyes for dye sensitized solar cells[†]

Vishwesh Venkatraman,^a Sailesh Abburu^a, and Bjørn Kåre Alsberg^{*a}

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXX 200X

DOI: 10.1039/b000000x

The design and discovery of novel molecular structures with optimal properties has been an ongoing effort for materials scientists. This field has in general been dominated by experiment driven trial-and-error approaches that are often expensive and time-consuming. Here, we investigate if a *de novo* computational design methodology can be applied to the design of coumarin-based dye sensitizer with improved properties for use in Grätzel solar cells. To address the issue of synthetic accessibility of the designed compounds, a fragment-based assembly is employed, wherein the combination of chemical motifs (derived from existing databases of structures) are carried out with respect to an user-adaptable set of rules. Rather than using computationally intensive Density Functional Theory (DFT)/*ab initio* methods to screen candidate dyes, we employ quantitative structure-property relationship (QSPR) models (calibrated from empirical data) for rapid estimates of the property of interest, which in this case is the product of the short circuit current (J_{sc}) and the open circuit voltage (V_{oc}). Since QSPR models have limited validity, pre-determined applicability domain criteria are used to prevent unacceptable extrapolation. DFT analysis of the top-ranked structures provide supporting evidence of their potential for dye sensitized solar cell applications.

1 Introduction

The direct exploitation of the abundant solar energy resources has been a long, on-going research effort¹. In the search for low-cost and commercially viable alternatives to silicon-based counterparts, dye-sensitized solar cells (DSSCs) have received widespread attention^{2–4}. The typical device architecture contains a transparent conducting oxide (fluorine-doped tin oxide), a semiconductor film (e.g. TiO_2 , ZnO) that is dyed with a sensitizer (adsorbed onto the surface), an electrolyte solution with a redox mediator (e.g. iodide/tri-iodide (I^-/I_3^- , cobalt(II/III))) and a counter electrode. When the sensitizing dye (an organic or inorganic transition metal complex) absorbs a photon, an electron is injected into the conduction band (CB) of the nanocrystalline semiconductor and is then transferred to the counter electrode (generally platinum-based) while passing through the transparent electrode and the external load. The oxidized dye is restored to the ground state by the redox mediator, which is in turn regenerated at the counter electrode. The DSSC is thus, essentially modular with functions such as light absorption (dye), electron (oxide) and hole (electrolyte) transport performed by different components that interact with each other⁵.

With the highest reported solar energy to current conversion efficiency of only 12%⁶ (for a zinc-porphyrin dye com-

bined with a cobalt based redox mediator), more needs to be done to improve the efficiency and stability of DSSCs. With this aim in view, research has focused on optimizing the different components ranging from the electrodes^{7–10} to electrolytes^{6,11–13} and dyes^{14–16}. In contrast to the other materials used in the DSSC, the properties of the dyes (light harvesting efficiency, stability etc) are more easily adjusted by making subtle changes to their chemical composition. Thus, a variety of organic dyes such as triphenylamines¹⁵, perylenes¹⁷, and inorganic ruthenium-based¹⁸ and zinc-porphyrin dyes¹⁹ have been explored. Although ruthenium-based dyes²⁰ have traditionally shown good efficiencies (up to 11%²¹), organic metal-free dyes too are slowly catching up with recent estimates closely matching inorganic counterparts. above^{22–24}. Furthermore, factors such as high molar extinction coefficients, structural diversity, environmentally friendliness, and availability of materials have resulted in increased interest in organic sensitizer²⁵.

With the dye sensitizer controlling key aspects of photovoltaic performance such as the light harvesting capacity, device stability and charge transfer efficiency, focus is on engineering them with specific properties suited for their application. Though most dyes typically possess a donor- π bridge-acceptor (D- π -A) architecture, other forms such as D-A- π -A have also been reported²⁶. The principal means of improving the DSSC efficiency has been to make molecular modifications (systematic variation of the π -spacer and anchoring groups²⁷) to existing dyes which have been shown to be promising. Given that the costs associated with these experiment-driven approaches tend to be quite high, the ap-

[†] Electronic Supplementary Information (ESI) available: [Structures and properties of the dye molecules, additional plots and tables]. See DOI: 10.1039/b000000x/

^a Department of Chemistry, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. E-mail: bjorn.k.alsberg@ntnu.no

proach is limited to investigating a rather small number of candidate dyes.

A better understanding of the relationship between the molecular structure and the photovoltaic performance would certainly facilitate a faster and more rational design of dyes²⁵. While a majority of this effort has been through experiments, there is a rising trend towards using computational methods, in particular quantitative structure-property relationship (QSPR) models^{28–33}. These methods, however, are not limited to dyes. For instance, machine learning and data mining tools have been used to model solar cell properties of metal oxides³⁴ and in other photovoltaic applications³⁵. Taking this a step further, in a recent article, the QSPR models were used to guide the evolutionary *de novo* design of phenothiazine dyes³⁶. Results showed that an increase of around 2% (in the estimated power conversion efficiency (*PCE*)) could be achieved for the proposed structures. The advantage of the approach lies in its ability to identify novel molecules and rapidly assess potential high-performance dyes while filtering out the less promising ones.

Until now, the highest reported efficiencies (*PCE*) for the coumarin dyes have been around 7–8%³⁷. With the aim of improving cell performance, a number of theoretical studies have been reported where existing coumarin dyes have been modified by introducing different substituent groups^{38–40}. In this contribution, we investigate the use of an efficient, computational screening scheme based on artificial evolution to the automated *in silico* design of coumarin-based dyes. Starting with a set of fragments (single atom or groups of atoms) that are combined in a synthetically tractable manner, the algorithm explores the molecular space by iteratively refining existing scaffold structures. The Darwinian operators of selection, mutation and crossover are used to create small changes in the geometric and electronic properties of the dye structure^{41–43}. Promising coumarin dyes emerging from these simulations are validated using quantum chemical calculations.

2 Methods and Materials

2.1 Artificial Molecular Evolution

Figure 1 provides a general overview of the *de novo* methodology being used and is only discussed here briefly. Readers are referred to previous articles for additional details of the procedure^{36,41,44}. At the core of the computational procedure is a genetic algorithm (GA) that explores different combinations of small fragments (building blocks), while taking into account aspects of synthetic accessibility of the proposed structures and the need for high performance. In the GA, the molecules are encoded as trees³⁶ whose nodes represent fragments that are connected by edges i.e. bonds. The mechanisms driving the evolution towards better solutions are

mainly, crossover and mutation which are applied so that genetic material in the form of fragments are either exchanged or substituted.

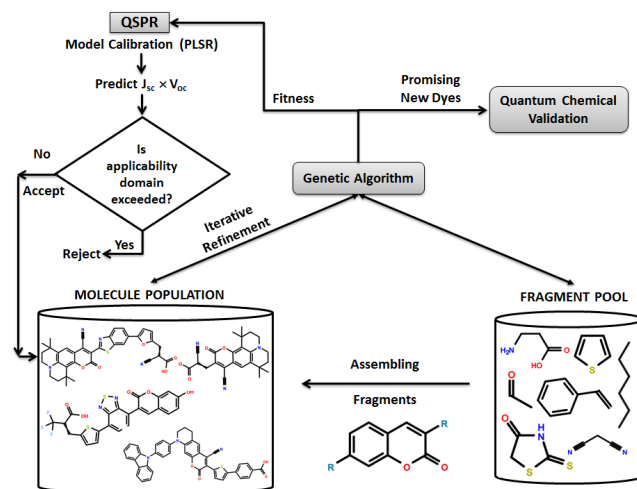


Fig. 1 Overview of the QSPR-guided *de novo* approach.

To maximize the probability of being able to synthesize the new structures, we have employed the concept of a fragment compatibility matrix (coded as a lookup table) that defines the connection rules between fragments^{36,44,45}. Each building block (fragment) stores information about the sites (atoms to which other fragments may be connected) called attachment points and the type of substituents that can be bound (see Figure 2). To create the relevant building blocks for the molecular assembly, dyes reported in the literature were subjected to a pseudo-retrosynthetic⁴⁴ decomposition. Attachment point annotation is therefore based on the type of substituents existing on each side of the cleaved bond. Over 100 such fragments containing moieties such as benzothiadiazole, thienopyrazine etc. that are found in several dyes were extracted. A second list of fragments consisting of coumarin scaffolds (different attachment points, chemical structure) were additionally identified (see bottom part of Figure 2) following a similar procedure.

Molecular assembly is carried out by selecting a coumarin scaffold (randomly from the list) and then looping through the available attachment points, adding fragments according to the linkage rules. Over a number of cycles, the GA modifies each structure (using crossover and mutation) in an attempt to improve its fitness, which here corresponds to a measure of the desired level of a given property. Both crossover and mutation are also subject to the fragment compatibility rules, thus ensuring realistic structures. The *de novo* simulations were carried out using in-house software written in Java.

Optimizing the efficiency of the DSSC requires a delicate balance between several processes. Although some of these

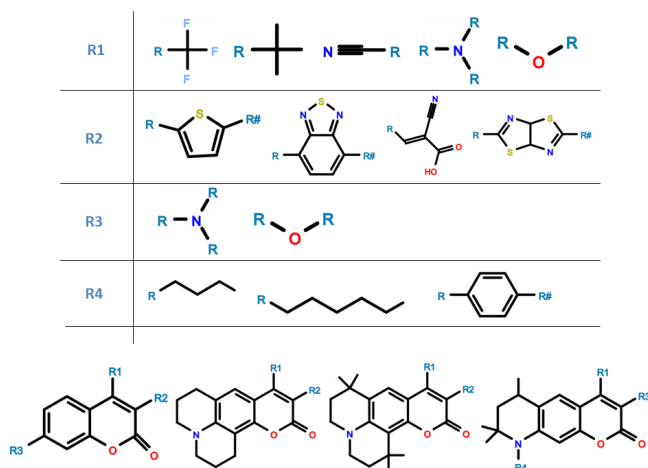


Fig. 2 Figure shows the compatibility matrix which governs the connectivity between fragments and thus constrains the chemical space to more realistic structures. The genetic algorithm attaches different fragments at the specified $R\#$ according to the linkage rules.

procedures can be modelled theoretically, only a few properties can be predicted (e.g. absorption wavelength, injection efficiency) reliably using quantum chemical methods⁴⁶. In addition, such properties also tend to be only weakly correlated with the *PCE*. For a typical GA run, where thousands of structures need to be assessed, use of DFT/TDDFT level electronic structure methods becomes computationally prohibitive given the computational costs. A more rapid estimation is afforded by QSPR methods which are now being increasingly used in computational chemistry to approximate properties of interest^{47–49}.

2.2 QSPR Modelling

The overall solar-to-electrical energy conversion efficiency *PCE*, is given by the following equation:

$$PCE(\%) = \frac{J_{sc} \cdot FF \cdot V_{oc}}{P_{inc}} \times 100 \quad (1)$$

where *FF* is the fill factor (*FF*), J_{sc} the short circuit current (measured in mA cm^{-2}) and V_{oc} the open circuit voltage (measured in mV). While the *FF* has been shown to be influenced by external factors such as electrode thickness and composition of the electrolyte etc⁵⁰, the J_{sc} and to some extent the V_{oc} are more directly influenced by the dye. Since the short circuit current and open circuit voltage largely depend on the molecular properties of the donor and acceptor moieties in the dye, we have chosen to model the product $J_{sc} \times V_{oc}$.

The experimental data on the photovoltaic properties for 49 coumarin dyes was taken from the literature^{37,51,52,52–63} where the *PCE* values for the molecules range from 0.3% to

7.7% (see Table S1 in the supplementary material). Previous studies have demonstrated the utility of vibrational frequency based eigenvalue (EVA) descriptors⁶⁴ for defining structure-property models for coumarin²⁹ and phenothiazine^{30,36} dyes. Here, we extend the approach to additionally include molecular orbital energies (eigenvalues of the Schrödinger equation). Unlike comparative molecular field analysis (CoMFA⁶⁵)-type methods, both EVA and electronic eigenvalue (EEVA)⁶⁶ do not require any 3D molecular alignment and are physically sound representations of the intrinsic electronic and physico-chemical properties of molecules.

The molecules in the data set were constructed using MarvinSketch⁶⁷ and energy-minimized using the Dreiding force field⁶⁸. The structures were then geometry optimized using the AM1 Hamiltonian in MOPAC⁶⁹. For each molecule, the vibrational frequencies (f_i) and molecular orbital energies (e_i) from the semi-empirical calculation are projected onto a bounded scale: $1 - 4000\text{cm}^{-1}$ for the EVA and $-45 - 10\text{ eV}$ for the EEVA. For a chosen sampling interval L , a Gaussian smoothing function of fixed standard deviation σ is applied to each frequency(energy) value and the EVA(EEVA) descriptor (D_x) at each point x of the scale is calculated as:

$$D_x^{EVA} = \sum_{i=1}^{3A-6} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-f_i)^2}{2\sigma^2}\right) \quad (2)$$

$$D_x^{EEVA} = \sum_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-e_i)^2}{2\sigma^2}\right) \quad (3)$$

where, A is the number of atoms in the molecule, N is the number of energy values. For the success of the method there is a need to identify the optimal set of σ and L values. σ and L were, therefore, varied as follows: $\sigma = \{0.025, 0.050, 0.075, 0.100, 0.200\}$ and $L = \{0.025, 0.050, 0.100\}$ for the EEVA and $\sigma = \{2, 4, 6, 8, 10\}$ and $L = \{1, 2, 5, 8, 10\}$ for the EVA.

Due to the efficiency in treating large numbers of descriptors, partial least squares regression (PLSR)⁷⁰ was used to predict the photovoltaic properties. The predictive power of the models was assessed with respect to independent training (80%) and test sets(20%) that were created using the Kennard-Stone algorithm⁷¹. Prior to modelling, all variables were autoscaled to zero mean and unit variance. The optimal number of latent variables (LVs) was determined using 10-fold cross-validation (CV), repeated 50 times, thus producing different sets of CV folds (500 total resamples). The number of LVs corresponding to the lowest root mean square error (*RMSE*) was then taken as the optimal value. The predictive ability of the model was evaluated in terms of the cross-validated correlation coefficient (R_{cv}^2) and the *RMSE* that are given by:

$$R_{cv}^2 = 1 - \frac{\sum (y_{obs,i} - \widehat{y}_{cv,i})^2}{\sum (y_{obs,i} - \overline{y}_{obs})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum (y^{obs} - y^{pred})^2}{n}} \quad (5)$$

where, $\widehat{y}_{cv,i}$ is the predicted value for the excluded (cross-validated) i^{th} compound, $y_{obs,i}$ is the corresponding experimental value and $\overline{y}_{obs}$ is the mean of the experimental values. Variable selection was performed using the variable importance on projection (VIP) scores that take into account the explained variance of each PLSR dimension⁷². Here, we have chosen to include only those variables with VIP scores greater than 1. Furthermore, to guard against overfitting, y -randomization tests⁷³ (repeated 1000 times) were also carried out. Using the above-described approach, multiple PLSR models were computed by varying the σ and L and the combination yielding the best R_{cv}^2 was then used in further analysis. The PLSR modelling was carried out using the *pls*⁷⁴ package available in the R ⁷⁵.

2.3 Applicability Domain

Although QSPR models can rapidly predict quantities of interest, given the finite size of their applicability domain (AD), the reliability of the predictions needs to be evaluated. Here, we have employed two such diagnostics. For both measures, high values suggest that the predictions should be treated with caution.

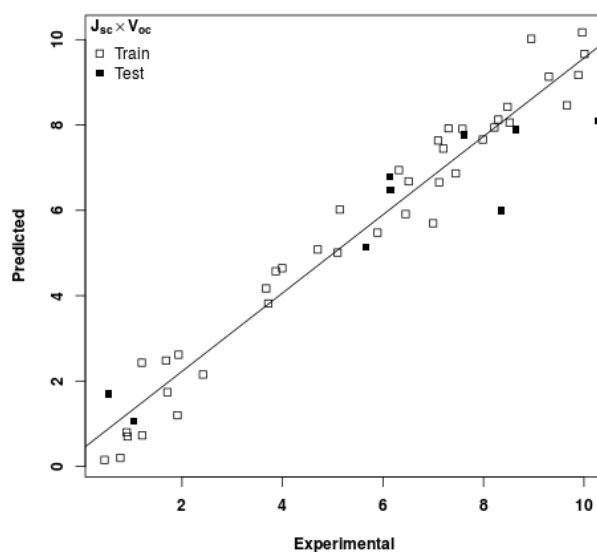
1. The leverage⁷⁶ $h_i = \mathbf{t}_i'(\mathbf{T}'\mathbf{T})^{-1}\mathbf{t}_i$ where $\mathbf{t}_i$ is the scores vector for the i^{th} compound and $\mathbf{T}$ is the model scores matrix. The warning leverage threshold is typically set to $3 \frac{N_{comp}}{N_{train}}$ where N_{comp} is the number of PLSR components (LVs) in the model and N_{train} , the number of compounds in the training set.
2. The uncertainty associated with each prediction $\tilde{y}_i = \sqrt{\frac{\sum_{j=1}^M (\hat{y}_j - \bar{y})^2}{M-1}}$ is calculated using a bootstrapping technique^{36,77}. In this work, 500 bootstrap models were created and the standard deviation of the predictions was taken as a measure of the uncertainty.

3 Results and Discussion

3.1 QSPR Results

In order to identify the best combination of the EVA and EEVA descriptors, over 500 models were computed. The settings of $\sigma = 8$ ($\sigma = 0.050$), $L = 10$ ($L = 0.025$) for the descriptors were found to yield the best models. For the calibration set containing 49 structures, a 3-component PLSR

Fig. 3 Plot of the observed vs predicted $J_{sc} \times V_{oc}$ values.



model with $R_{cv}^2 = 0.83$ and $RMSE = 1.25 \text{ (mA} \cdot \text{V cm}^{-2}\text{)}$ was obtained. Corresponding figures of merit for the test set (9 compounds) were $R_{test}^2 = 0.89$ and $RMSE_{test} = 1.21$ which is suggestive of a fairly predictive model despite the structural diversity of the molecules in the dataset. Table 1 summarizes the results for the QSPR model calculated for $J_{sc} \times V_{oc}$. The scatter plot in Figure 3 further shows the linear relationship between experimental and predicted values for both training and test set data. Further, examination of the score plot (see Figure F1 in the Electronic Supplementary Information (ESI)) clearly shows a separation between low ($< 2\%$ PCE) and high efficiency structures. For the predicted $J_{sc} \times V_{oc}$, applicability domain estimates in the form of leverage values (h_i) are well within the warning threshold of 0.225 with maximum uncertainties of up to $\pm 2 \text{ mA} \cdot \text{V cm}^{-2}$.

3.2 de novo Evolution Runs

The GA experiments were carried out using population sizes of 100-200 molecules with 5 offspring produced during each generation (100-500 cycles). For each proposed molecule, the QSPR model was used to rapidly assess the $J_{sc} \times V_{oc}$. Both crossover and mutation probabilities were set to 0.50. During the runs, structures that violated the warning leverage threshold were discarded. Over 6000 unique dye molecules were generated with predicted $J_{sc} \times V_{oc}$ ranging between 2-12 $\text{mA} \cdot \text{V cm}^{-2}$, an increase of up to 2 units with respect to the calibration set data. Although very few structures violated the leverage threshold, prediction uncertainties can rise up to $\pm 20 \text{ mA} \cdot \text{V cm}^{-2}$. In addition to the commonly occurring D- π -A configuration, other molecular architectures containing

Table 1 QSPR model performance for the calibration and test set molecules (text in italics). The "QSPR" column contains the predicted $J_{sc} \times V_{oc}$ with calculated uncertainties. h_i corresponds to the leverage for the QSPR prediction.

Molecule	$J_{sc} \times V_{oc}$	QSPR	h_i	Molecule	$J_{sc} \times V_{oc}$	QSPR	h_i
C01 ⁵⁶	1.68	2.59 ± 0.77	0.06	C26 ⁵⁶	0.92	1.21 ± 0.71	0.07
C02 ⁵²	1.20	2.50 ± 1.01	0.06	C27 ⁵⁶	1.71	1.94 ± 0.59	0.06
C03 ⁷⁸	8.52	8.46 ± 0.46	0.03	C28 ⁵⁶	7.00	4.86 ± 1.57	0.05
C04 ⁷⁸	6.51	7.65 ± 0.97	0.02	C29 ⁵⁶	7.10	7.98 ± 0.61	0.04
C05 ⁷⁸	7.58	8.60 ± 0.88	0.03	C30 ³⁷	7.99	7.94 ± 0.50	0.03
C06 ⁷⁹	8.29	8.29 ± 0.34	0.02	C31 ⁸⁰	8.64	8.15 ± 0.42	0.03
C07 ⁷⁹	7.45	7.02 ± 0.49	0.01	C32 ³⁷	10.30	8.22 ± 0.43	0.03
C08 ⁵³	1.21	1.08 ± 0.43	0.07	C33 ³⁷	10.01	9.66 ± 0.63	0.05
C09 ⁵³	1.04	1.23 ± 0.37	0.06	C34 ⁵⁷	9.30	9.05 ± 0.60	0.05
C10 ⁵³	0.54	1.90 ± 0.49	0.04	C35 ⁵⁸	9.66	8.77 ± 1.06	0.04
C11 ⁵³	0.77	1.14 ± 0.52	0.07	C36 ⁸¹	7.30	8.14 ± 0.62	0.04
C12 ⁵³	0.46	0.39 ± 0.47	0.08	C37 ⁵⁷	9.96	10.00 ± 0.42	0.08
C13 ⁵³	0.90	0.41 ± 0.49	0.09	C38 ⁵⁹	7.20	6.13 ± 1.08	0.07
C14 ⁵⁴	7.12	6.51 ± 0.63	0.04	C39 ⁵⁹	8.48	7.05 ± 1.20	0.10
C15 ⁵⁴	6.14	6.86 ± 0.66	0.06	C40 ⁵⁹	8.95	9.25 ± 0.91	0.17
C16 ⁵⁴	6.32	7.07 ± 0.88	0.08	C41 ⁶⁰	3.72	3.29 ± 0.47	0.04
C17 ⁵⁵	8.22	8.68 ± 0.46	0.04	C42 ⁶⁰	1.93	2.60 ± 0.88	0.03
C18 ⁵⁵	5.89	6.91 ± 0.98	0.07	C43 ⁶⁰	4.00	3.90 ± 0.54	0.01
C19 ⁵⁵	3.87	5.24 ± 1.04	0.00	C44 ⁶⁰	2.42	2.00 ± 0.56	0.07
C20 ⁵⁶	3.68	5.03 ± 1.11	0.01	C45 ⁶⁰	5.10	4.24 ± 0.65	0.01
C21 ⁵⁶	8.36	6.48 ± 0.52	0.02	C46 ⁶⁰	1.91	0.98 ± 0.78	0.10
C22 ⁵⁶	4.70	5.22 ± 0.42	0.02	C47 ⁶¹	5.14	5.11 ± 0.69	0.02
C23 ⁵⁶	6.45	6.27 ± 0.52	0.04	C48 ⁶¹	7.62	6.44 ± 1.22	0.02
C24 ⁵⁶	6.14	6.98 ± 0.55	0.02	C49 ⁶¹	9.89	7.29 ± 2.00	0.04
C25 ⁵⁶	5.66	5.88 ± 0.78	0.02				

double donors (D-D- π -A), multiple π -A and D- π -A- π based systems were also observed. The top ranking structures (see dyes M01-M06 in Table 2) contained electron donating substituents such as furan, auxiliary acceptors such as quinoxaline and benzothiazole. Dyes containing secondary electron donors such as fluorene⁸² and triphenylethylene⁸³ have also been explored. While on one hand they can enhance photoinduced intramolecular charge transfer, owing to their twisted non-planar structure, these moieties have been shown to slow down electron recombination kinetics and lengthen electron lifetimes^{84,85}. Additionally, the presence of such bulky groups can also limit aggregation through steric hindrance⁸⁶. While the carboxylic acid based anchoring group occurred in a majority of the dyes, some molecules instead contained a nitro group as an alternative anchoring group which has been shown to increase the electron injection ability²⁷.

3.3 Theoretical Validation of Designed Dyes

Selected promising structures (see Table 2) emerging from the GA runs were subjected to additional calculations at the DFT level of theory. Solvation effects were included by means of the conductor-like polarizable continuum model (CPCM)⁸⁷.

The analysis focuses on important features such as the UV-vis spectra and reorganization energy³¹. All calculations were carried out using the Gaussian software⁸⁸.

3.3.0.1 HOMO-LUMO distribution: Preliminary examination of the ground state isodensity plots (computed at the B3LYP/6-31 G(d,p) level) in Table 2, shows that the orbitals are well distributed over the linker conjugation. While the HOMO is generally localized over the coumarin scaffold (donor), the LUMO is distributed around the acceptor and anchoring groups. This localization is expected to be advantageous for charge separation and electron injection^{4,89}.

3.3.0.2 Spectral properties: In addition to a high light harvesting efficiency (*LHE*), with reference to Equation 1, the short circuit current J_{sc} ^{90,91} is defined as:

$$J_{sc} = \int LHE(\lambda) \phi_{inj} \eta_{col} d\lambda \quad (6)$$

where $LHE(\lambda)$ is the light harvesting efficiency at wavelength λ , ϕ_{inj} is the electron injection efficiency and η_{col} is the charge collection efficiency. Since the cells only differ in the dyes being used, it is reasonable to assume that η_{col} is constant. Thus, the two main factors influencing J_{sc} are the *LHE* and ϕ_{inj} . The

LHE of the dye is calculated as $LHE(\lambda) = 1 - 10^{-f(\lambda)}$ where $f(\lambda)$ is the oscillator strength of the dye at absorption wavelength (λ). In order to model the absorption spectra, we have employed time-dependent density functional theory⁹². Since the choice of the functional affects the accuracy of the DFT calculated spectral properties, we analysed a number of hybrid functionals. Among those tested, the CAM-B3LYP⁹³ functional with a DGDZVP⁹⁴ basis set was found to provide the best agreement between theoretical and experimental data (see Table S2 in the ESI). For the proposed coumarin dyes, the computed maximum absorption wavelengths λ_{max} (in ethanol), oscillator strengths (f_{max}) and the light harvesting efficiency *LHE* are listed in Table 3.

3.3.0.3 Dye-TiO₂ complex: For dyes containing the carboxylic acid anchoring group (M01-M04), the optical response after adsorption onto the TiO₂ surface was also studied. Assuming a bidentate chelating configuration, TDDFT calculations were performed for dyes bound to a (TiO₂)₉ cluster, which has been shown to reproduce electronic absorption spectra with reasonable accuracy^{96,97}. Here, the geometry optimizations were carried out using the B3LYP functional and 6-31G(d,p) basis set for the C, H, O, N, S atoms and the effective core potential (ECP) LANL2DZ basis set for the Ti atom. With the exception of dye M01 (for which convergence could not be achieved), all three dyes tested (M02, M03, M04) show only a marginal decrease in the (~ 10 nm) in the λ_{max} (see Table 4).

3.3.0.4 Injection efficiency: In order to maximize the photocurrent (J_{sc}), the electron injection efficiency ϕ_{inj} also needs to be enhanced. Since ϕ_{inj} is related to the thermodynamic driving force of electron injection^{38,91,98} (ΔG_{inj}) i.e. the free energy change (in eV) that can be expressed as:

$$\Delta G_{inj} = E_{ox}^{dye*} - E_{CB}^{TiO_2} \quad (7)$$

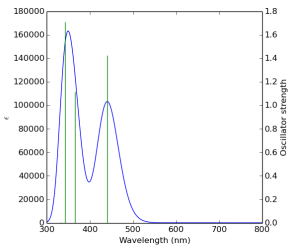
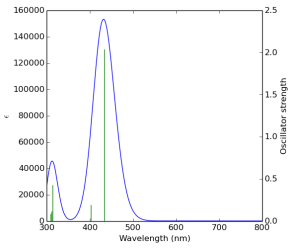
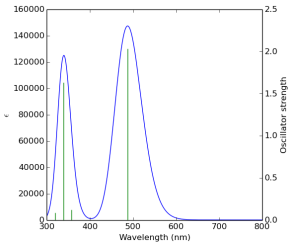
where $E_{ox}^{dye*} = E_{ox}^{dye} - \lambda_{max}$ is the oxidation potential of the dye in the excited state and E_{CB} is the reduction potential of conduction band of the semiconductor (experimental value of -4.0 eV vs. vacuum is used⁹⁹). The excited-state oxidation potential is calculated from the redox potential of the ground-state dye and the absorption energy corresponding to the λ_{max} :

$$E_{ox}^{dye*} = E_{ox}^{dye} - \lambda_{max} \quad (8)$$

Another important requirement is that of low reorganization energies to ensure rapid electron transfer during regeneration¹⁰⁰. Here, we focus on the inner-sphere reorganization energy λ_i ¹⁰¹ which is calculated as:

$$\lambda_i = [E_0(G_+) + E_+(G_0)] - [E_+(G_+) + E_0(G_0)] \quad (9)$$

Table 4 Simulated TDDFT spectra (in ethanol) of dyes adsorbed on TiO₂.

Mol	Spectra	λ_{max} (nm/eV)	f
M02		440 (2.81)	1.42
M03		447 (2.86)	2.04
M04		488 (2.54)	2.03

where E_j is the energy of the system in charge state j , G_k is the geometry of the system in charge state k , while 0 and + correspond to the neutral and cationic charge states respectively. Table 5 lists the injection efficiencies and reorganization energies for the 6 dyes analyzed. The driving force of the electron injection (ΔG_{inj}) decreases in the order M03 > M04 > M05 > M01 > M06 > M02. The corresponding trends for the calculated reorganization energies proceeds as M01 > M04 > M06 > M02 > M05 > M03. Among these, dye M03 containing both fluorene and tetraphenylethylene units possesses the highest ΔG_{inj} that favours electron injection from dye excited state into the TiO₂ conduction band edge. In addition a low reorganization energy should also support rapid electron transfer. In dye M04, the presence of methine moieties can increase the conjugation length as reflected in the spectral response (see Table 3). However, larger values of λ_i may result in slower injection kinetics. Furthermore, the stronger electron localization in rhodanine-3-acetic acid acceptor moiety that is present in M04, can cause the lowest unoccupied molecular orbital (LUMO) to be shifted from the anchoring group, thus impeding electron injection into the conduction band of TiO₂^{102,103}. For the dye M01, which contains al-

ternating electron withdrawing- π conjugation groups, the impact of removing these substituents was studied. The existing molecular architecture of D-A- π -A was modified manually by either rearranging or replacing the moieties. A significant decrease by up to 5 units in the predicted $J_{sc} \times V_{oc}$ values for the modified dyes, although associated uncertainties are much smaller (see Table S4 in the ESI). Two dyes (M05, M06) contain a nitro group as the acceptor/anchoring that have been shown to improve dye/TiO₂ coupling and facilitate electron transfer^{27,104}.

Table 5 Estimated HOMO, LUMO, ΔG_{inj} , reorganization energies λ_i (in eV) for the proposed dyes.

Dye	HOMO	LUMO	ΔG_{inj} (eV)	λ_i (eV)
M01	-4.94	-2.51	-1.71	0.206
M02	-4.97	-2.35	-1.89	0.149
M03	-5.18	-2.59	-1.40	0.097
M04	-5.22	-2.91	-1.55	0.202
M05	-5.07	-2.91	-1.69	0.148
M06	-5.04	-2.92	-1.81	0.150

The present work makes use of a QSPR-driven evolutionary molecular design approach that allows for the estimation of the $J_{sc} \times V_{oc}$ in a matter of minutes (2-3 minutes on an Intel Core i5-2400 CPU @ 3.10GHz). This can be easily integrated into a virtual high-throughput screening framework where large libraries of dyes can be analyzed by using theoretical techniques¹⁰⁵. Although the EVA/EEVA descriptors are not easily amenable to interpretation, their use is purely motivated by the prospect of performing accelerated dye discovery. Furthermore, the existing model does not account for the interactions between the cell components (dye-electrolyte¹⁰⁶, dye-dye¹⁰⁷, dye-TiO₂¹⁰⁸). Predictive models applied in this study that are based solely on the dye can therefore be limited in practice. Attention must also be drawn to the experimental variability and their impact on the DSSC performance indicators. While it is desirable to obtain data measured according to recommended practices^{109,110}, we do observe differences in the procedures for the preparation (different electrolyte compositions, TiO₂ layer thickness) and evaluation (different active areas, use of additives such as tert-butyl pyridine, deoxycholic acid in different concentrations, lack of stability tests) of cells. From a QSPR perspective, the variability in the measurement protocols introduces greater uncertainty in the models. In this respect, data generated by combinatorial chemistry based approaches emerging from a single lab (thus avoiding inter-lab variations) would be an promising alternative¹¹¹.

4 Conclusions

In this article, we have investigated the feasibility of using a fully automated *de novo* approach based on artificial evolution to design new and promising coumarin dye sensitizers with improved properties. Underlying the approach is a QSPR model built from experimental data which enables fast estimation of the property being optimized i.e. $J_{sc} \times V_{oc}$. The scheme has been able to propose a diverse set of dyes with different molecular architectures (D- π -A, D-A- π -A, D-D- π -A etc.) and anchoring groups. The analyzed dyes contained substituents such as fluorene and tetraphenylethylene which not only limit aggregation but also act as additional electron donors. DFT verification of the new coumarin dyes makes them promising candidates for future experimental testing.

Acknowledgements

The Norwegian Research Council (NFR) is acknowledged for financial support from the eVITA (Grant No. 205273) and the CLIMIT (Grant No. 233776) and for CPU resources granted through the NOTUR supercomputing programme. We also thank ChemAxon (<http://www.chemaxon.com>) for free academic use of the Marvin package.

References

- 1 J. Jean, P. R. Brown, R. L. Jaffe, T. Buonassisi and V. Bulovic, *Energy Environ. Sci.*, 2015, **8**, 1200–1219.
- 2 K. G. Reddy, T. G. Deepak, G. S. Anjusree, S. Thomas, S. Vadukumpully, K. R. V. Subramanian, S. V. Nair and A. S. Nair, *Phys. Chem. Chem. Phys.*, 2014, **16**, 6838–6858.
- 3 A. Fakharuddin, R. Jose, T. M. Brown, F. Fabregat-Santiago and J. Bisquert, *Energy Environ. Sci.*, 2014, **7**, 3952–3981.
- 4 S. Zhang, X. Yang, Y. Numata and L. Han, *Energy Environ. Sci.*, 2013, **6**, 1443–1464.
- 5 T. W. Hamann, R. A. Jensen, A. B. F. Martinson, H. Van Ryswyk and J. T. Hupp, *Energy Environ. Sci.*, 2008, **1**, 66–78.
- 6 A. Yella, H.-W. Lee, H. N. Tsao, C. Yi, A. K. Chandiran, M. Nazeeruddin, E. W.-G. Diau, C.-Y. Yeh, S. M. Zakeeruddin and M. Grätzel, *Science*, 2011, **334**, 629–634.
- 7 X. Shun-Jian, L. Yu-Feng, Z. Wei, X. Zong-Hu, L. Yong-Ping and O. Hui, *J. Inorg. Mater.*, 2015, **30**, 29–34.
- 8 S. Ma, M. Shang, L. Yu and L. Dong, *J. Mater. Chem. A*, 2015, **3**, 1222–1229.
- 9 O. Byrne, I. Ahmad, P. K. Surolia, Y. K. Gun'ko and K. R. Thampi, *Sol. Energy*, 2014, **110**, 239–246.
- 10 S. H. Park, Y.-H. Cho, M. Choi, H. Choi, J. S. Kang, J. H. Um, J.-W. Choi, H. Choe and Y.-E. Sung, *Surf. Coat. Technol.*, 2014, **259**, 560–569.
- 11 M. A. K. L. Dissanayake, W. N. S. Rupasinghe, V. A. Seneviratne, C. A. Thotawatthage and G. K. R. Senadeera, *Electrochim. Acta*, 2014, **145**, 319–326.
- 12 A. Chiappone, F. Bella, J. R. Nair, G. Meligrana, R. Bongiovanni and C. Gerbaldi, *ChemElectroChem*, 2014, **1**, 1350–1358.
- 13 F. Bella, C. Gerbaldi, C. Barolo and M. Grätzel, *Chem. Soc. Rev.*, 2015, **44**, 3431–3473.

- 14 C.-P. Lee, R. Y.-Y. Lin, L.-Y. Lin, C.-T. Li, T.-C. Chu, S.-S. Sun, J. T. Lin and K.-C. Ho, *RSC Adv.*, 2015, **5**, 23810–23825.
- 15 M. Liang and J. Chen, *Chem. Soc. Rev.*, 2013, **42**, 3453–3488.
- 16 J.-H. Yum, E. Baranoff, S. Wenger, M. K. Nazeeruddin and M. Grätzel, *Energy Environ. Sci.*, 2011, **4**, 842–857.
- 17 S. Mathew and H. Imahori, *J. Mater. Chem.*, 2011, **21**, 7166–7174.
- 18 J. Yin, M. Velayudham, D. Bhattacharya, H. Lin and H. Lu, *Coord. Chem. Rev.*, 2012, **256**, 3008–3035.
- 19 L.-L. Li and E. W.-G. Diau, *Chem. Soc. Rev.*, 2013, **42**, 291–304.
- 20 A. Hagfeldt, G. Boschloo, L. Sun, L. Kloo and H. Pettersson, *Chem. Rev.*, 2010, **110**, 6595–6663.
- 21 L. Han, A. Islam, H. Chen, C. Malapaka, B. Chiranjeevi, S. Zhang, X. Yang and M. Yanagida, *Energy Environ. Sci.*, 2012, **5**, 6057–6060.
- 22 K. Kakiage, Y. Aoyama, T. Yano, T. Otsuka, T. Kyomen, M. Unno and M. Hanaya, *Chem. Commun.*, 2014, **50**, 6379–6381.
- 23 D. Joly, L. Pellejá, S. Narbey, F. Oswald, J. Chiron, J. N. Clifford, E. Palomares and R. Demadrille, *Sci. Rep.*, 2014, **4**.
- 24 W. Zeng, Y. Cao, Y. Bai, Y. Wang, Y. Shi, M. Zhang, F. Wang, C. Pan and P. Wang, *Chem. Mater.*, 2010, **22**, 1915–1925.
- 25 A. Mishra, M. Fischer and P. Bäuerle, *Angew. Chem. Int. Edit.*, 2009, **48**, 2474–2499.
- 26 Y. Wu and W. Zhu, *Chem. Soc. Rev.*, 2013, **42**, 2039–2058.
- 27 L. Zhang and J. M. Cole, *ACS Appl. Mater. Inter.*, 2015, **7**, 3427–3455.
- 28 T. Le, V. C. Epa, F. R. Burden and D. A. Winkler, *Chem. Rev.*, 2012, **112**, 2889–2919.
- 29 V. Venkatraman, P.-O. Åstrand and B. K. Alsberg, *J. Comput. Chem.*, 2014, **35**, 214–226.
- 30 V. Venkatraman and B. K. Alsberg, *Dyes Pigments*, 2015, **114**, 69–77.
- 31 C. M. Ip, A. Eleuteri and A. Troisi, *Phys. Chem. Chem. Phys.*, 2014, **16**, 19106–19110.
- 32 S. Tortorella, G. Marotta, G. Cruciani and F. De Angelis, *RSC Adv.*, 2015, **5**, 23865–23873.
- 33 H. Li, Z. Zhong, L. Li, R. Gao, J. Cui, T. Gao, L. H. Hu, Y. Lu, Z.-M. Su and H. Li, *J. Comp. Chem.*, 2015, **36**, 1036–1046.
- 34 A. Yosipof, O. E. Nahum, A. Y. Anderson, H.-N. Barad, A. Zaban and H. Senderowitz, *Molecular Informatics*, 2015, n/a–n/a.
- 35 A. Mellit and S. A. Kalogirou, *Progress in Energy and Combustion Science*, 2008, **34**, 574–632.
- 36 V. Venkatraman, M. Foscato, V. R. Jensen and B. K. Alsberg, *J. Mater. Chem. A*, 2015, **3**, 9851–9860.
- 37 K. Hara, Z.-S. Wang, T. Sato, A. Furube, R. Katoh, H. Sugihara, Y. Dan-oh, C. Kasada, A. Shinpo and S. Suga, *J. Phys. Chem. B*, 2005, **109**, 15476–15482.
- 38 J. Wang, M. Li, D. Qi, W. Shen, R. He and S. H. Lin, *RSC Adv.*, 2014, **4**, 53927–53938.
- 39 R. Sánchez-de Armas, M. A. San-Miguel, J. Oviedo and J. F. Sanz, *J. Chem. Phys.*, 2012, **136**, –.
- 40 X. Zhang, J.-J. Zhang and Y.-Y. Xia, *Journal of Photochemistry and Photobiology A: Chemistry*, 2008, **194**, 167–172.
- 41 Y. Chu, W. Heyndrickx, G. Occhipinti, V. R. Jensen and B. K. Alsberg, *J. Am. Chem. Soc.*, 2012, **134**, 8885–8895.
- 42 N. Brown, B. McKay, F. Gilardoni and J. Gasteiger, *J. Chem. Inf. Model.*, 2004, **44**, 1079–1087.
- 43 B. P. Hay and T. K. Firman, *Inorg. Chem.*, 2002, **41**, 5502–5512.
- 44 M. Foscato, G. Occhipinti, V. Venkatraman, B. K. Alsberg and V. R. Jensen, *J. Chem. Inf. Model.*, 2014, **54**, 767–780.
- 45 M. Foscato, V. Venkatraman, G. Occhipinti, B. K. Alsberg and V. R. Jensen, *J. Chem. Inf. Model.*, 2014, **54**, 1919–1931.
- 46 N. Martsinovich and A. Troisi, *Energy Environ. Sci.*, 2011, **4**, 4473–4495.
- 47 R. Ramakrishnan and O. A. von Lilienfeld, *CHIMIA International Journal for Chemistry*, 2015, **69**, 182–186.
- 48 K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, 2013, **9**, 3404–3419.
- 49 G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *New Journal of Physics*, 2013, **15**, 095003.
- 50 C. Lopez-Lopez, S. Colodrero, M. E. Calvo and H. Miguez, *Energy Environ. Sci.*, 2013, **6**, 1260–1266.
- 51 K. Hara, Y. Tachibana, Y. Ohga, A. Shinpo, S. Suga, K. Sayama, H. Sugihara and H. Arakawa, *Sol. Energy Mater. Sol. Cells*, 2003, **77**, 89–103.
- 52 S. Verma and H. N. Ghosh, *J. Phys. Chem. C*, 2014, **118**, 10661–10669.
- 53 V. Kandavelu, H.-S. Huang, J.-L. Jian, T. C.-K. Yang, K.-L. Wang and S.-T. Huang, *Solar Energy*, 2009, **83**, 574–581.
- 54 H. Choi, C. Baik, H. J. Kim, J.-J. Kim, K. Song, S. O. Kang and J. Ko, *Bull. Korean Chem. Soc.*, 2007, **28**, 1973–1979.
- 55 B. Liu, R. Wang, W. Mi, X. Li and H. Yu, *J. Mater. Chem.*, 2012, **22**, 15379–15387.
- 56 K. Hara, T. Sato, R. Katoh, A. Furube, Y. Ohga, A. Shinpo, S. Suga, K. Sayama, H. Sugihara and H. Arakawa, *J. Phys. Chem. B*, 2003, **107**, 597–606.
- 57 Z.-S. Wang, Y. Cui, Y. Dan-oh, C. Kasada, A. Shinpo and K. Hara, *J. Phys. Chem. C*, 2007, **111**, 7224–7230.
- 58 Z.-S. Wang, K. Hara, Y. Dan-oh, C. Kasada, A. Shinpo, S. Suga, H. Arakawa and H. Sugihara, *J. Phys. Chem. B*, 2005, **109**, 3907–3914.
- 59 L. Alibabaei, J.-H. Kim, M. Wang, N. Pootrakulchote, J. Teuscher, D. Di Censo, R. Humphry-Baker, J.-E. Moser, Y.-J. Yu, K.-Y. Kay, S. M. Zakeeruddin and M. Grätzel, *Energy Environ. Sci.*, 2010, **3**, 1757–1764.
- 60 L. Han, H. Wu, Y. Cui, X. Zu, Q. Ye and J. Gao, *J. Photoch. Photobiol., A*, 2014, **290**, 54–62.
- 61 C. Zhong, J. Gao, Y. Cui, T. Li and L. Han, *J. Power Sources*, 2015, **273**, 831–838.
- 62 K. Hara, K. Sayama, Y. Ohga, A. Shinpo, S. Suga and H. Arakawa, *Chem. Commun.*, 2001, 569–570.
- 63 K. Hara, T. Sato, R. Katoh, A. Furube, T. Yoshihara, M. Murai, M. Kurashige, S. Ito, A. Shinpo, S. Suga and H. Arakawa, *Adv. Funct. Mater.*, 2005, **15**, 246–252.
- 64 T. Heritage, A. Ferguson, D. Turner and P. Willett, *Perspect. Drug Discov.*, 1998, **9-11**, 381–398.
- 65 R. D. Cramer, D. E. Patterson and J. D. Bunce, *J. Am. Chem. Soc.*, 1988, **110**, 5959–5967.
- 66 K. Tuppurainen, *SAR QSAR Environ. Res.*, 1999, **10**, 39–46.
- 67 Marvin 5.9.3, 2012, ChemAxon (<http://www.chemaxon.com>).
- 68 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
- 69 J. J. P. Stewart, *MOPAC2012*, 2012, Stewart Computational Chemistry, Colorado Springs, CO, USA, (<http://OpenMOPAC.net>).
- 70 H. Martens and T. Naes, *Multivariate Calibration*, John Wiley & Sons, New York, 1989.
- 71 R. W. Kennard and L. A. Stone, *Technometrics*, 1969, **11**, 137–148.
- 72 I.-G. Chong and C.-H. Jun, *Chemometr. Intell. Lab.*, 2005, **78**, 103–112.
- 73 S. Wiklund, D. Nilsson, L. Eriksson, M. Sjöström, S. Wold and K. Faber, *J. Chemom.*, 2007, **21**, 427–439.
- 74 B.-H. Mevik and R. Wehrens, *J. Stat. Softw.*, 2007, **18**, 1–24.
- 75 R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012.
- 76 H. Martens and T. Naes, *Multivariate Calibration*, John Wiley & Sons, New York, 1989.
- 77 Y. An, W. Sherman and S. L. Dixon, *J. Chem. Inf. Model.*, 2013, **53**, 2312–2321.
- 78 K. D. Seo, H. M. Song, M. J. Lee, M. Pastore, C. Anselmi, F. D. Angelis, M. K. Nazeeruddin, M. Grätzel and H. K. Kim, *Dyes Pigments*, 2011,

- 90, 304–310.
- 79 K. D. Seo, I. T. Choi, Y. G. Park, S. Kang, J. Y. Lee and H. K. Kim, *Dyes Pigments*, 2012, **94**, 469–474.
 - 80 K. Hara, M. Kurashige, Y. Dan-oh, C. Kasada, A. Shinpo, S. Suga, K. Sayama and H. Arakawa, *New J. Chem.*, 2003, **27**, 783–785.
 - 81 Z.-S. Wang, Y. Cui, Y. Dan-oh, C. Kasada, A. Shinpo and K. Hara, *J. Phys. Chem. C*, 2008, **112**, 17011–17017.
 - 82 G. Marzari, J. Durantini, D. Minudri, M. Gervaldo, L. Otero, F. Fungo, G. Pozzi, M. Cavazzini, S. Orlandi and S. Quici, *J. Phys. Chem. C*, 2012, **116**, 21190–21200.
 - 83 C. Chen, J.-Y. Liao, Z. Chi, B. Xu, X. Zhang, D.-B. Kuang, Y. Zhang, S. Liu and J. Xu, *J. Mater. Chem.*, 2012, **22**, 8994–9005.
 - 84 C. Chen, J.-Y. Liao, Z. Chi, B. Xu, X. Zhang, D.-B. Kuang, Y. Zhang, S. Liu and J. Xu, *RSC Adv.*, 2012, **2**, 7788–7797.
 - 85 J. Shi, J. Huang, R. Tang, Z. Chai, J. Hua, J. Qin, Q. Li and Z. Li, *Eur. J. Org. Chem.*, 2012, **2012**, 5248–5255.
 - 86 X. Jiang, T. Marinado, E. Gabrielsson, D. P. Hagberg, L. Sun and A. Hagfeldt, *J. Phys. Chem. C*, 2010, **114**, 2799–2805.
 - 87 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
 - 88 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.
 - 89 D. P. Hagberg, T. Marinado, K. M. Karlsson, K. Nonomura, P. Qin, G. Boschloo, T. Brinck, A. Hagfeldt and L. Sun, *J. Org. Chem.*, 2007, **72**, 9550–9556.
 - 90 J. Zhang, H.-B. Li, S.-L. Sun, Y. Geng, Y. Wu and Z.-M. Su, *J. Mater. Chem.*, 2012, **22**, 568–576.
 - 91 J. Preat, D. Jacquemin and E. A. Perpete, *Energy Environ. Sci.*, 2010, **3**, 891–904.
 - 92 C. Adamo and D. Jacquemin, *Chem. Soc. Rev.*, 2013, **42**, 845–856.
 - 93 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
 - 94 N. Godbout, D. R. Salahub, J. Andzelm and E. Wimmer, *Can. J. Chem.*, 1992, **70**, 560–571.
 - 95 N. M. O’boyle, A. L. Tenderholt and K. M. Langner, *J. Comput. Chem.*, 2008, **29**, 839–845.
 - 96 R. Sanchez-de Armas, M. A. San Miguel, J. Oviedo and J. F. Sanz, *Phys. Chem. Chem. Phys.*, 2012, **14**, 225–233.
 - 97 R. Sánchez-de Armas, M. A. San-Miguel, J. Oviedo and J. Fdez Sanz, *J. Chem. Phys.*, 2012, **136**, 194702.
 - 98 J. Wang, F.-Q. Bai, B.-H. Xia, L. Feng, H.-X. Zhang and Q.-J. Pan, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2206–2213.
 - 99 R. Katoh, A. Furube, T. Yoshihara, K. Hara, G. Fujihashi, S. Takano, S. Murata, H. Arakawa and M. Tachiya, *J. Phys. Chem. B*, 2004, **108**, 4818–4822.
 - 100 J. W. Ondersma and T. W. Hamann, *Coord. Chem. Rev.*, 2013, **257**, 1533–1543.
 - 101 V. Vaissier, P. Barnes, J. Kirkpatrick and J. Nelson, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4804–4814.
 - 102 T. Marinado, D. P. Hagberg, M. Hedlund, T. Edvinsson, E. M. J. Johansson, G. Boschloo, H. Rensmo, T. Brinck, L. Sun and A. Hagfeldt, *Phys. Chem. Chem. Phys.*, 2009, **11**, 133–141.
 - 103 Z. Wan, C. Jia, Y. Duan, J. Zhang, Y. Lin and Y. Shi, *Dyes Pigments*, 2012, **94**, 150–155.
 - 104 J. Cong, X. Yang, J. Liu, J. Zhao, Y. Hao, Y. Wang and L. Sun, *Chem. Commun.*, 2012, **48**, 6663–6665.
 - 105 E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre and A. Aspuru-Guzik, *Ann. Rev. Mater. Res.*, 2015, **45**, 195–216.
 - 106 M. Pastore, E. Mosconi and F. De Angelis, *J. Phys. Chem. C*, 2012, **116**, 5965–5973.
 - 107 M. Pastore and F. De Angelis, *J. Phys. Chem. Lett.*, 2013, **4**, 956–974.
 - 108 E. Ronca, M. Pastore, L. Belpassi, F. Tarantelli and F. De Angelis, *Energy Environ. Sci.*, 2013, **6**, 183–193.
 - 109 H. J. Snaith, *Nat Photon.*, 2012, **6**, 337–340.
 - 110 H. J. Snaith, *Energy Environ. Sci.*, 2012, **5**, 6513–6520.
 - 111 S. Fuse, S. Sugiyama, M. M. Maitani, Y. Wada, Y. Ogomi, S. Hayase, R. Katoh, T. Kaiho and T. Takahashi, *Chem. Eur. J.*, 2014, **20**, 10685–10694.

Table 2 Chemical structures of the most promising (large $J_{sc} \times V_{oc}$ values) dyes emerging from different *de novo* runs. For each dye, computed isodensity surfaces (0.02 a.u.) of the HOMO and LUMO orbitals (based on B3LYP/6-31G(d,p) geometries) are shown alongside.

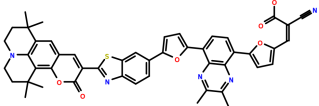
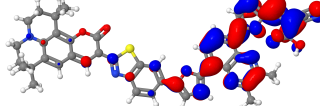
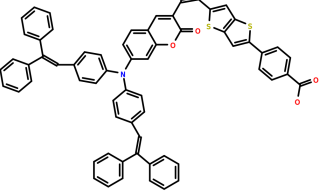
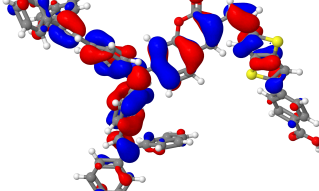
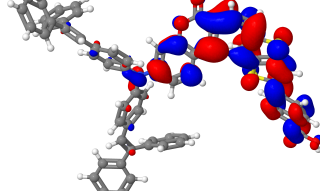
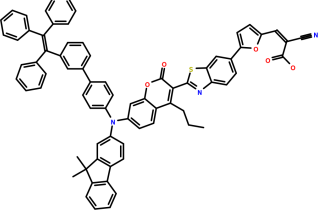
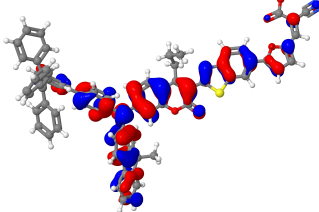
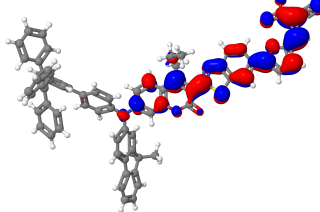
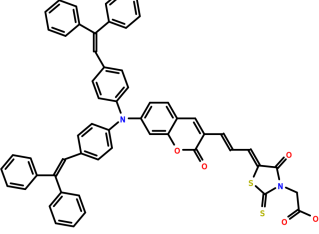
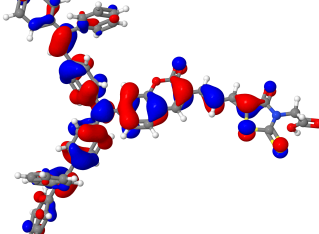
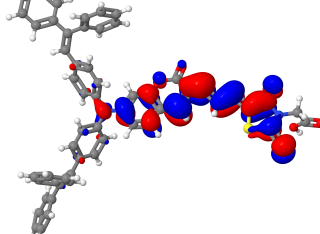
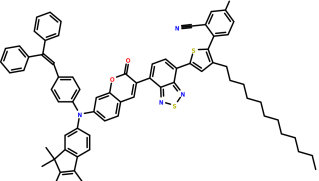
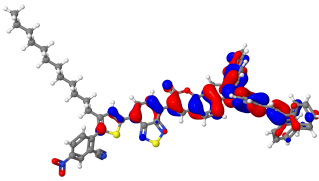
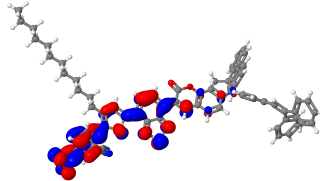
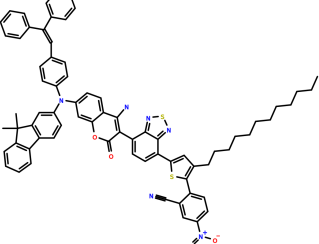
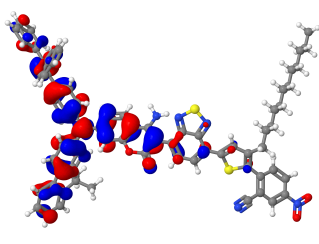
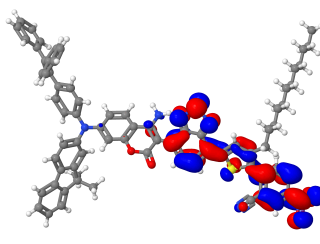
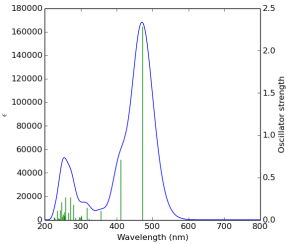
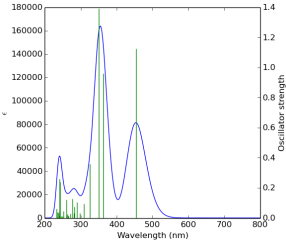
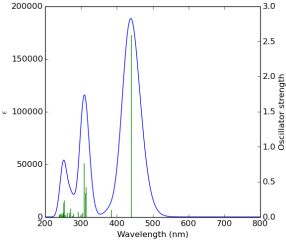
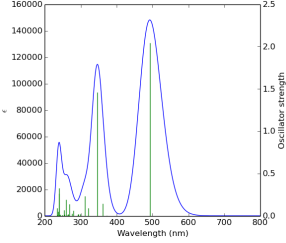
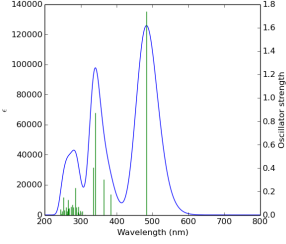
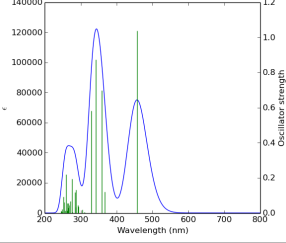
Dye	$J_{sc} \times V_{oc}$	Structure	HOMO	LUMO
M01	12.09 ± 2.43			
M02	11.98 ± 3.47			
M03	11.12 ± 3.24			
M04	11.02 ± 3.16			
M05	11.68 ± 4.00			
M06	11.51 ± 4.00			

Table 3 Calculated UV/Vis spectra corresponding to the S_0 - S_1 transition in ethanol using the CAM-B3LYP functional and DGDZVP basis set. The spectra are obtained using Gausssum⁹⁵. The letters "H" and "L" in the column "Major transitions" correspond to the HOMO and LUMO orbitals.

Dye	Absorption Spectra	λ_{max} (nm/eV)	f	LHE	Major Transitions
M01		471 (2.63)	2.28	0.994	H-1→L (26%), H→L (50%), H→L+1 (10%)
M02		454 (2.73)	1.12	0.924	H-1→L (20%), H→L (64%)
M03		439 (2.49)	2.59	0.997	H-2→L (28%), H→L (42%), H→L+1 (18%)
M04		492 (2.51)	2.04	0.990	H-1→L (19%), H→L (68%)
M05		483 (2.57)	1.73	0.981	H-1→L (27%), H→L (40%), H→L+1 (16%)
M06		457 (2.71)	1.04	0.908	H-1→L (37%), H-1→L+1 (24%), H→L (17%)