PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2024, **26**, 16847

Screening potential dye sensitizers for water splitting photocatalysts using a genetic algorithm†

Tao Liu, *[a] Linjiang Chen, *[b] Xiaoyan Wang [a] and Andrew I. Cooper *[a]

Addressing the global fossil energy crisis necessitates the efficient utilization of sustainable energy sources. Hydrogen, a green fuel, can be generated using sunlight, water, and a photocatalyst. Employing sensitizers holds promise for enhancing photocatalyst performance, enabling high rates of hydrogen evolution through increased visible light absorption. However, sifting through millions of diverse molecules to identify suitable dyes for specific photocatalysts poses a significant challenge. In this study, we integrate genetic algorithm and geometry-frequency-noncovalent extended tight binding methods to efficiently screen 2.6 million potential sensitizers with a D–π–A–π–AA structure within a short timeframe. Subsequently, these optimized sensitizers are rigorously reassessed by using DFT/TDDFT methods, elucidating why they may serve as superior dyes compared to the reference dye WS5F, particularly in terms of light absorption, driving force, binding energy, *etc.* Additionally, our methodology uncovers molecular motifs of particular interest, including the furan π-bridge and the double cyano anchoring acceptor, which are prevalent in the most promising set of molecules. The developed genetic algorithm workflow and dye design principles can be extended to various compelling projects, such as dye-sensitized solar cells, organic photovoltaics, photo-induced redox reactions, pharmaceuticals, and beyond.

Received 11th April 2024,
Accepted 28th May 2024

DOI: 10.1039/d4cp01487a

rsc.li/pccp

Introduction

Hydrogen is considered a promising and environmentally benign energy source; however, its natural abundance on Earth is low. One viable approach for hydrogen generation involves the utilization of sunlight and photocatalysts to facilitate water splitting. Since the initial discovery of TiO_2[1] as a photocatalyst for light-driven water splitting, a plethora of inorganic photocatalysts has been investigated, including $SrTiO_3$,[2,3] $BiVO_4$,[4,5] ZrO and CdS.[6,7] One notable challenge associated with these inorganic photocatalysts is their limited capacity to absorb visible light. Consequently, various sensitizers have been explored to broaden the light absorption range, akin to the strategy employed in dye-sensitized solar cells.[8,9] It is worth noting that these sensitizing dyes predominantly interface with the surface of the photocatalysts, thereby constraining the quantity of dye molecules capable of injecting electrons into the semiconductor photocatalyst. Recently, purely organic photocatalysts, such as carbon nitride,[10,11] conjugated microporous polymers (CMPs),[12] linear conjugated polymers,[13] covalent triazine-based frameworks (CTFs),[14,15] and covalent organic frameworks (COFs)[16–18] have attracted significant attention. Among these, COFs[19] have emerged as a focal point of research for water splitting[20] and other applications due to their commendable thermal stability, molecular tunability, and extensive surface areas.[21] Although conjugated COFs can exhibit broader visible light absorption compared to many inorganic photocatalysts, their light absorption profiles generally remain insufficient for harnessing energy in the near-infrared (NIR) region, which constitutes a substantial portion of the solar spectrum.[22,23] Notably, Wang *et al.*[16] previously reported hydrogen evolution rates (HERs) of up to 10.1 mmol g^{-1} h^{-1} using a dibenzosulfone photocatalyst, FS-COF. Furthermore, the HER of this COF could be augmented by over 60% to 16.3 mmol g^{-1} h^{-1} through the incorporation of a NIR absorbing dye, WS5F, as a sensitizer. However, WS5F exhibits limited solubility in water and relatively weak intermolecular interaction with the FS-COF surface, necessitating preloading into the pores of FS-COF using acetone prior to photocatalysis. Given the expansive chemical space of organic dyes, it is improbable that WS5F represents the optimal sensitizer for FS-COF. However, *a priori* design of superior dye sensitizers poses a non-trivial challenge.

In this context, the primary function of a sensitizer, as depicted in Fig. 1, is to capture incident photons and facilitate the injection of photoexcited electrons into the conduction

[a] *Department of Chemistry and Materials Innovation Factory, Leverhulme Research Centre for Functional Materials Design, University of Liverpool, 51 Oxford Street, Liverpool, L7 3NY, UK. E-mail: tao.liu@liverpool.ac.uk, aicooper@liverpool.ac.uk*
[b] *School of Chemistry and School of Computer Science, University of Birmingham, Birmingham, B15 2TT, UK. E-mail: l.j.chen@bham.ac.uk*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4cp01487a

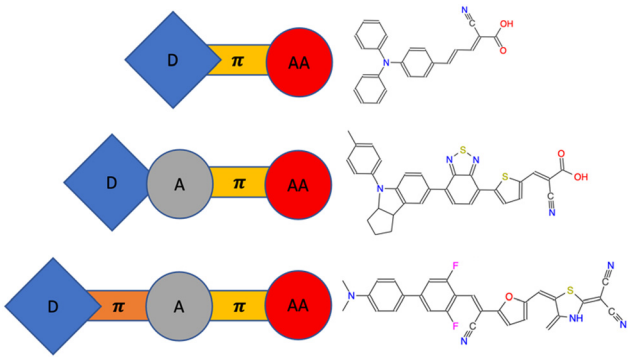

Fig. 1 Typical dye sensitizers with D–π–AA/D–A–π–AA structure and new structure D–π–A–π–AA (D: donor, π: π-bridge, A: acceptor, AA: anchoring acceptor).

band of the photocatalyst, thereby augmenting the performance of water splitting. Several crucial properties must be met for a sensitizer to be deemed effective.[23,24] Firstly, it must exhibit optical absorption characteristics that encompass substantial portions of the visible spectrum, while also extending into the NIR range. Secondly, the presence of an anchoring acceptor group, which may be established through either covalent or non-covalent means, is imperative. This group serves the purpose of affixing the sensitizer onto the surface of the photocatalyst in an orientation conducive to charge injection. Thirdly, the electron affinity (EA) of the dye, denoted by its exciton ionization potential (IP*), and the ionization potential (IP), reflected by its exciton electron affinity (EA*), should straddle the potentials associated with proton reduction and the oxidation of water or sacrificial electron donors (SED), as illustrated in Fig. 2. While other properties relating to dynamics of excitons are very important, they were not included in this study due to their time-consuming nature.[25]

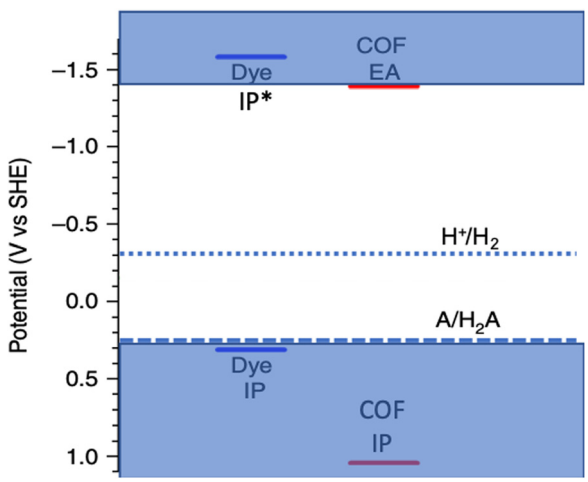

Fig. 2 Scheme showing idealized EA/IP* and IP of a COF photocatalyst and a dye sensitizer. The hydrogen reduction potential (H^+/H_2) and the sacrificial donor oxidation potential (A/H_2A) are marked with dashed lines. The IP/IP* of good candidate dye should be located, broadly speaking, in the blue shaded regions of this diagram.

Many dye sensitizers exhibit a fundamental linear D–π–AA (D: donor, π: π-bridge, AA: anchoring acceptor) configuration, with or without a π-bridge, as exemplified in Fig. 1. This structure is adept at enhancing charge separation and impeding charge recombination. The extent of conjugation in the donor, π-bridge, and acceptor moieties within the sensitizer, coupled with the particular functional groups attached, dictates the light absorption characteristics. The π-bridge connecting the donor and acceptor is pivotal in facilitating efficient charge separation. Moreover, the anchoring acceptor group plays a crucial role in determining the strength of the interaction between the sensitizer and the photocatalyst, a factor that can significantly influence the charge injection process. Additionally, the specific attributes of the donor and acceptor entities will influence their respective ionization potentials (IP and IP*), which in turn have a profound impact on charge injection and the regeneration of the dye.

As depicted in Fig. 1, four common building blocks—donor (D), π-bridge (π), acceptor (A) and anchoring acceptor (AA)—can be systematically combined to create a vast array of potential sensitizers with diverse structures, including D–π–AA,[26,27] D–A–π–AA,[28–31] and D–π–A–π–AA.[32] D–π–AA architecture was initially introduced as a sensitizer due to its adaptable modulation of intramolecular charge-transfer characteristics.[33] Subsequently, D–A–π–AA configurations, incorporating various additional electron-withdrawing groups between the donor and anchoring acceptor (such as benzothiadiazole, quinoxaline, thiazole, triazine, cyanovinyl, substituted phenyl, *etc.*), were introduced and widely applied, demonstrating superior performance compared to D–π–AA. This is attributed to the several advantages of the D–A–π–AA structure, including facile adjustment of molecular energy levels, broader light absorption range, and enhanced photovoltaic efficiency and stability. D–π–A–π–AA structure combines the advantages of both of these aforementioned structures. However, there are over three million potential dye sensitizers in D–π–A–π–AA configuration when each of the four building block libraries contains only twenty molecular fragments. Clearly, it is impractical to experimentally synthesize and test more than a minuscule fraction of these candidates. Furthermore, it is cost-prohibitive to computationally predict the properties of all 3 million candidate sensitizer molecules using density functional theory (DFT), despite its relatively moderate computational expense. Therefore, we have developed an algorithmic approach to systematically explore this vast molecular dye sensitizer space.

A genetic algorithm (GA)[34–36] draw inspiration from Darwin's theory of natural selection, wherein individuals that are better adapted to their environment have a higher likelihood of survival and passing on their genetic traits. In the realm of chemistry and materials science, GA approaches prove invaluable for addressing various challenges, including copolymer design and stable phase searching,[37,38] doping and defects in crystals,[2,39,40] perovskites for water splitting,[41] and optimization of organic photovoltaic materials.[42] GA offers distinct advantages over machine learning (ML) methods for several specific problems. First, GA can initiate the optimization process without the need for pre-existing data, whereas machine

learning methods often necessitate a substantial database for model training and testing. Second, GA can yield satisfactory solutions within relatively short time frames. Last but not least, GA constitute a direct optimization process wherein solutions progressively improve with each generation, while the quality of solutions derived from machine learning hinges on the accuracy of the trained model. However, there is no doubt that ML will benefit material searching combining with GA, but it is not the scope of this research.[40,43]

The objective of this study was to employ GA for the optimization of a set of theoretical molecules constructed from predefined building block libraries. The aim was to identify potential dye sensitizers for the FS-COF photocatalyst that could potentially surpass the original sensitizer, WS5F.[14] To accomplish this, we integrated the recently developed geometry, frequency, noncovalent, eXtended tight binding (GFN-xTB) method implemented in xTB[44] with calibration to properties obtained by DFT/TD-DFT methods using a linear model (Fig. S1, ESI†). This approach enabled us to conduct *in silico* optimization of dyes to sensitize the specific photocatalyst, FS-COF, with DFT/TD-DFT level of accuracy. Previously, Wilbraham *et al.*[45] and Bai *et al.*[46] combined DFT (TDDFT) with GFN-xTB, employing a tight-binding-based simplified Tamm–Dancoff approximation (sTDA-xTB) and reported that this hybrid approach could be employed to efficiently screen thousands of polymers for targeted applications. Additionally, the screened dyes were re-evaluated using DFT/TD-DFT methods. Properties including ionization potential, exciton ionization potential, light absorption characteristics, solubility, binding energy, and other excited state properties were calculated with reference to the 'starting' dye, WS5F.[14] Consequently, we proposed three candidate dye structures that hold potential for sensitizing FS-COF for enhanced photocatalytic hydrogen evolution compared to WS5F.

Methods

Dye building blocks and architectures

The building blocks featured in our libraries, as illustrated in Fig. 3, comprise commonplace organic fragments capable of interconnecting to form various dyes. Through the random combination of five building blocks derived from these four distinct libraries, namely donor (D), π-bridge (π), acceptor (A), and anchoring acceptor (AA) (refer to Fig. 1 and 3), we have the capacity to generate over 2.6 million unique dye sensitizers, all adhering to a D–π–A–π–AA template structure.

Methods and properties

xTB and sTDA-xTB methods. The computationally efficient GFN-xTB method, implemented within the xTB package, was

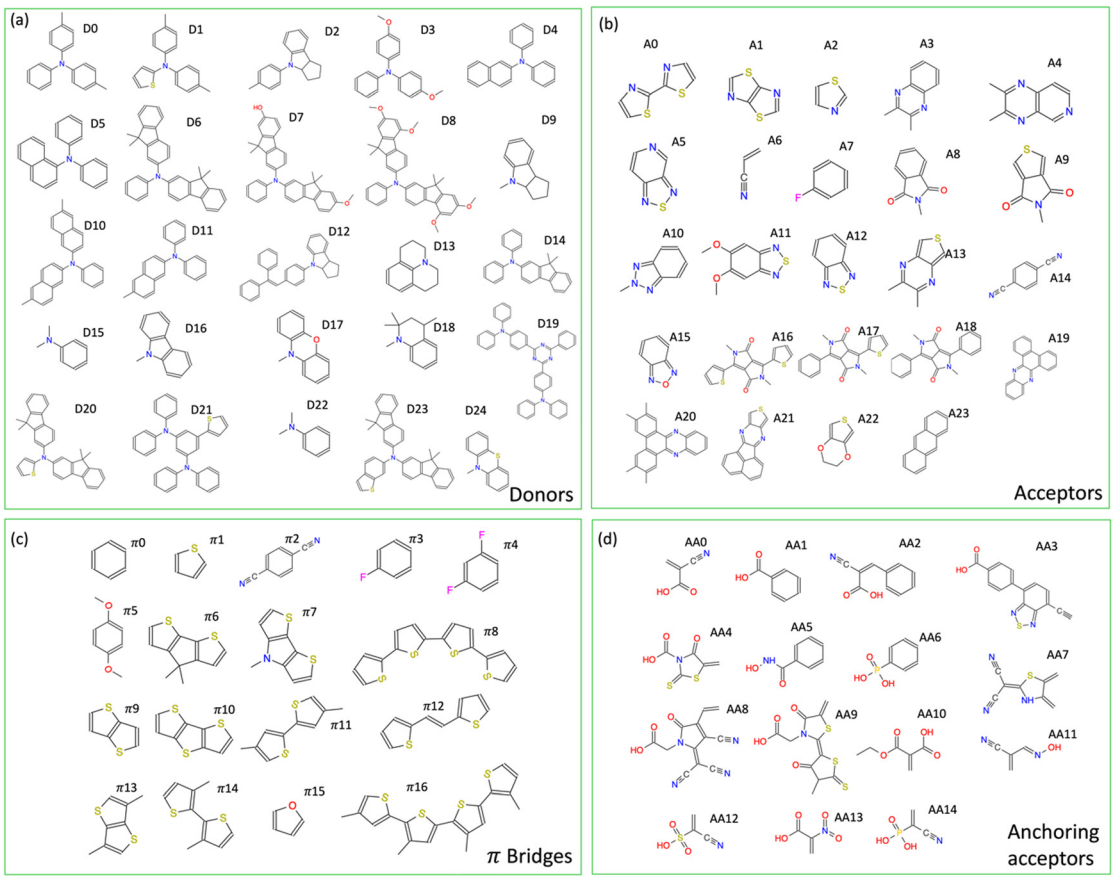

Fig. 3 Building blocks with IDs for D–π–A–π–AA dye sensitizer structures. (a) donors, (b) acceptors (c) π-bridges, (d) anchoring acceptors.

employed for optimizing the neutral molecular structures.[44] Additionally, for determining the vertical ionization potentials and electron affinities of the dyes, we utilized IPEA-xTB,[47] an extension of GFN-xTB with distinct parameterization. All xTB calculations, encompassing both optimization and IP/EA assessments, were conducted using the generalized Born surface area solvation model, with default parameters for water provided by the xTB code. To evaluate dye solubility, we computed the energy difference between molecules with and without solvation effects. For rapid and accurate determination of the lowest energy absorption of dyes, we applied the xTB-based tight binding simplified Tamm–Dancoff approach (sTDA),[48,49] a method implemented in sTDA-xTB.[50]

Genetic algorithm

Workflow. Fig. 4 and 5 show the workflow of the GA built by C^{++}, which could be downloaded from **https://github.com/taoliuliverpool/GA**. Initially, a population of dyes (100) was generated by combining the simplified molecular-input line-entry system (SMILES)[51] notation from five building block libraries (donor, π-bridge, acceptor, π-bridge, anchoring acceptor). These notations were then transformed into 3D structures using the MMFF94 force field, a process facilitated by Open-Babel[52] followed by a conformer search. Subsequently, the fitness function of each dye was computed. These dyes, along with their respective fitness functions, were subjected to three key genetic operators: selection, crossover, and mutation. These operations produced an equal number of offspring, which then replaced the previous generation, thus initiating a new iteration of the GA cycle. The GA process was set to terminate either after 50 generations (dependent on the number of solutions) or when the total fitness difference between the new generation and its parents' generation fell below 0.1.

Fitness function

The properties of the dyes were computed by xTB and sTDA-xTB in the following sequence: IP, IP*, driving force ($E_d = IP^*_{dye} - EA_{COF}$), solvation energy ($E_{sol}$), light absorption ($E_{abs}$) and then binding energy ($E_b = E_{dye\_COF} - E_{COF} - E_{dye}$, defined in ESI†)[53,54] as depicted in Fig. 4. If any of the dye's properties fail to meet its predefined working criteria (IP $>$ 0.2 eV, IP* $<$ −1.6 eV, $E_{abs} < 2.8$ eV, $E_{sol} < 0.0$ eV, $E_b < 0.0$ eV), the fitness function of the dye is set to 0 and the remaining properties are not computed to conserve computational resources. Should all properties of a given dye satisfy their working criteria, the fitness of the dye is then calculated and proceeded to the GA process to do the optimization and generate the next generation. To address various interests, three distinct fitness functions are defined, each covering different aspects: (a) E_{abs} (50%) + E_b (50%); (b) E_{abs} (100%); (c) E_d (100%). (a) and (b) are designed to ensure that the optimized dyes surpass the reference dye (WS5F) in terms of light absorption (E_{abs}) and binding energy (E_b). (c) is formulated to guarantee that the dyes outperform WS5F in terms of charge injection driving force (E_d) into FS-COF.

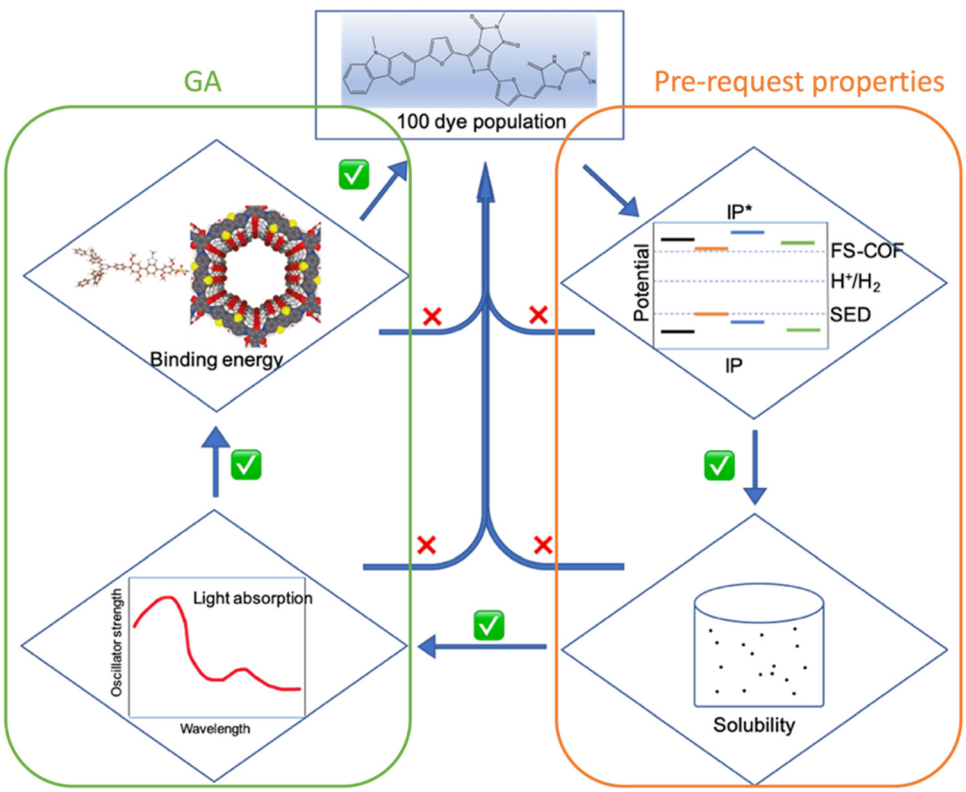

Fig. 4 Workflow for calculating pre-request properties (IP, IP*, E_{sol}) and the fitness function (E_{abs}, E_b, E_d) in the GA process.

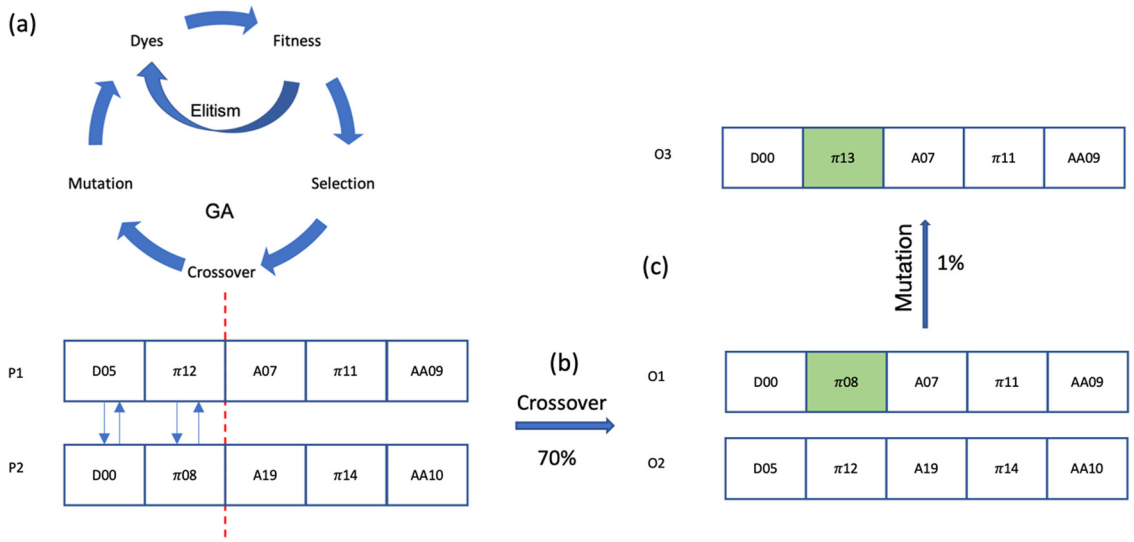

Fig. 5 GA workflow (a), crossover (b) and mutation (c). P1 and P2 are the selected dyes for crossover, O1 and O2 are the offspring after crossover, O3 is the offspring after mutation on O1. There are five building blocks in each dye representing by the squares, the number in each building block is the ID in the library of this building block listed in Fig. 3. Red dash line is the random single locus for crossover. The green building block is the building block undergoing a mutation.

Operators in GA

The genetic algorithm employed here utilizes the simplest form, incorporating three fundamental operators: selection, crossover (specifically, random single-point crossover), and mutation (also employing random single-point mutation), together with Elitism (Only the best parent with the highest fitness score will enter Elitism) as shown in Fig. 5.

1. Selection: this process involves the selection of two dyes from the population pool for the purpose of generating offspring. Dyes that exhibit higher fitness score are more likely to be chosen for the subsequent step, which is crossover.

2. Crossover: this step entails the selection of a random locus, where the subsequence before and after the locus are exchanged between two dyes. This operation results in the creation of two offspring. The crossover rate is set at 0.7.

3. Mutation: in this phase, one building block of the dye is replaced by another building block chosen at random from its respective library. The mutation rate is set at 0.01.

Re-evaluation by DFT/TDDFT

The DFT (B3LYP) and TDDFT (CAM-B3LYP) methodologies[55–58] are employed for re-evaluating the optimized dyes. The IP and IP* are computed using an adiabatic DFT (TDDFT) approach, relying on the DFT (TDDFT) energies of the neutral ground state and excited state dyes, as well as their respective relaxed cationic counterparts.[59] The values of IP and IP* are reported on the standard hydrogen electrode (SHE) scale.

The solvation effect was incorporated using the SMD model with water as the solvent.[60] The DFT calculations employed the B3LYP density functional in conjunction with the Def2-SVP basis-set,[61] consistent with our prior work.[16] To account for long-range effects during excitation, the absorption spectra of the dye were computed *via* vertical singlet excitations utilizing the TD-CAM-B3LYP method. All DFT/TDDFT calculations were conducted by Gaussian16.[62] Properties such as electron–hole overlap, coulombic attraction energy, and dipole moment were derived from TD-CAM-B3LYP results by Multiwfn.[63]

Several similar research studies have been conducted by Fan *et al.*[64–66] on optimizing substitutions of a main structure for solar cells. However, our approach offers several advantages: we use xTB instead of DFTB/DFTB+ to reduce element restrictions, provide calibrated TB results for fitness scores to minimize deviations and adopt a selection operator for faster and better offspring generation.

Results and discussion

The initial step involves selecting the first generation to serve as the starting population. One hundred distinct dyes, meeting the criteria for IP, IP*, E_{abs}, E_{sol}, E_d and E_b as discussed earlier, are constructed by randomly selecting five building blocks from their respective libraries. To ensure the initial population is as diverse as possible, GA optimizations with 20 different initial populations have been successfully conducted. The comparison of GA performance with a random selection process is presented in Fig. 6. It is evident that GA optimizes the best fitness function of the lowest-energy absorption to 0.69 by the 7th generation. Additionally, the average fitness function exhibits a steady increase over successive generations. In contrast, the best fitness function attained through random selection only reaches 0.56 by the 23rd generation, with no further improvement thereafter.

The performance of the sensitizer is determined by a multitude of parameters, which encompass IP, IP*, E_d, E_{abs} and E_b to the photocatalyst, as well as properties of exciton (electron–hole pair). These crucial factors are visualized in Fig. 7 and 8.

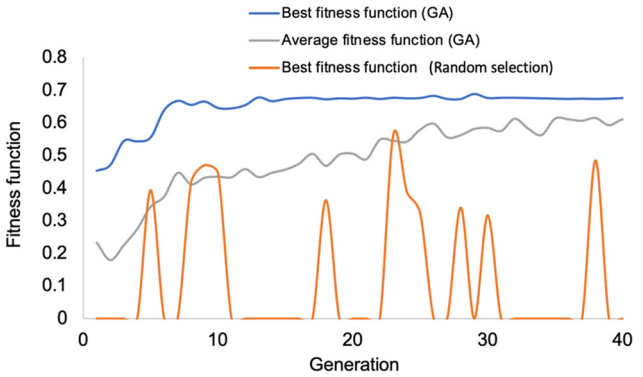

Fig. 6 Comparison of fitness function of GA with random selection as a function of generation. Best fitness function represents the maximum fitness function (F_{max}) of the population both in GA optimization and random selection. Average fitness function represents the average fitness function of all 100 individuals in the population, which can be calculated by $(F_{average} = 1 \Big/ 100 \sum_{i}^{100} F_i)$.

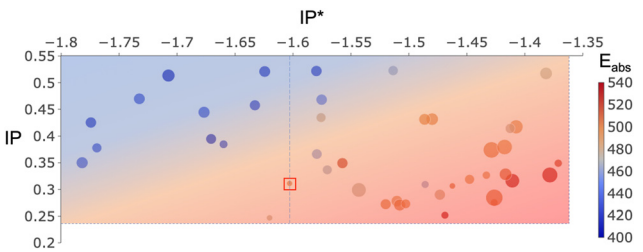

Fig. 7 5D plot for all optimized sensitizers. *x* axis is IP* (V), *y* axis is IP (V), colour coded by E_{abs} (nm), the size of the points is proportional to E_b of the dye to FS-COF (kcal mol^{-1}), distance to right-hand side dash line is E_d for charge injection to photocatalyst FS-COF. Data of WS5F as a reference dye is in red box. The data on the left-hand side of WS5F blue dash line is obtained by setting driving force as fitness in GA and the data on the right of WS5F blue dash line is obtained by setting absorption and/or binding energy as fitness function in GA. Black dash line at the bottom is the oxidation potential of sacrificial electron donor, while the black dash line on the right-hand side is the EA of photocatalyst FS-COF. The data of this plot are also listed in Table S1 (ESI†). The structures of these optimized dyes are listed in Fig. S2 (ESI†).

The IP values of all the optimized dyes (Fig. 7) are found to be higher than the potential of A/H2A (+0.246 V, indicated by the dashed line at the bottom). This indicates that all of them can effectively undergo regeneration by a sacrificial electron donor like ascorbic acid. Furthermore, the IP* values of all the selected dyes are more negative than the EA of FS-COF (−1.366 V, denoted by the dashed line on the right-hand side of Fig. 7), suggesting that these dyes possess the capability to inject hot electrons into the photocatalyst. The net potential or driving force ($IP_{dye} - E_{ox}$ and $EA_{FS\text{-}COF} - IP^*_{dye}$) is a significant area of discussion.[16,67–69] Previous research by Wang *et al.*[16] demonstrated a 60% enhancement in HER of FS-COF by incorporating WS5F as a sensitizer, with a difference between IP_{dye} and E_{ox} of less than 0.1 V, while the difference between IP^*_{dye} and E_{red} was larger than 0.2 V. Conversely, Swierk *et al.*[69] working on TiO_2, concluded that the rutile polymorph outperformed anatase TiO_2 in water splitting with the aid of a sensitizer, attributing this to the conduction band edge of rutile being 0.2 eV below that of anatase, providing an increased driving force for electron injection from dye to TiO_2.

Fig. 7 illustrates that the reference dye WS5F (dash line in the middle) separates our optimized dyes into two distinct groups. Those on the right-hand side of WS5F, optimized by GA using absorption and/or binding energy as the fitness function, exhibit driving forces similar to WS5F, approximately 0.1 V, on the oxidation half-reaction, and a smaller driving force than WS5F in the range of 0 to 0.25 V for charge injection into FS-COF. However, most of them are capable of absorbing lower-energy light and binding to FS-COF more strongly than WS5F. Additionally, the data on the left-hand side of WS5F, optimized using charge injection driving force as the fitness function, possess significantly larger driving forces than WS5F for both dye regeneration (0.1 to 0.3 V) and charge injection (0.3 to 0.5 V). Nevertheless, they are unable to capture lower-energy light as efficiently as WS5F.

Light absorption is undeniably one of the pivotal properties for designing effective sensitizers. It is imperative that the dye's absorption spectrum spans the entire sunlight region and, even more crucially, extends into the visible to near-infrared region to harness a greater portion of light with lower energy.[70,71] Fig. 7 is color-coded to represent the lowest energy absorption of the optimized dyes, where red indicates absorption in the low energy range, while blue signifies absorption in the high energy range.

On the right-hand side of WS5F, our findings indicate that:

– 31% of the screened dyes exhibit superior absorption compared to WS5F (absorption falls within the range of 503 to 531 nm).

– 38% of them demonstrate similar light absorption as WS5F (absorption within the range of 493 to 501 nm).

– 31% of the optimized dyes have moderate light absorption relative to WS5F (absorption within the range of 457 to 490 nm); nevertheless, they outperform the light absorption of FS-COF (403 nm).

It remains plausible that dyes with modest light absorption may outperform WS5F due to other favourable parameters, such as higher binding energy, smaller exciton energy, larger dipole moment, and so on.

On the left-hand side of WS5F, these dyes absorb light with higher energy compared to WS5F, yet their light absorption still surpasses that of FS-COF. As discussed earlier, these dyes boast a greater charge injection driving force towards FS-COF than WS5F.

Fig. 7 illustrates that a superior charge injection driving force tends to accompany a less efficient light absorption. This implies that it is possible to fine-tune light absorption (to lower energies) by compromising on the driving force, effectively shifting the dye towards the bottom right of Fig. 7. This trade-off between light absorption and driving force is crucial. To achieve a substantial driving force, one could either lower the IP or elevate EA(IP*). This would result in a wider physical gap between the highest occupied molecular orbital

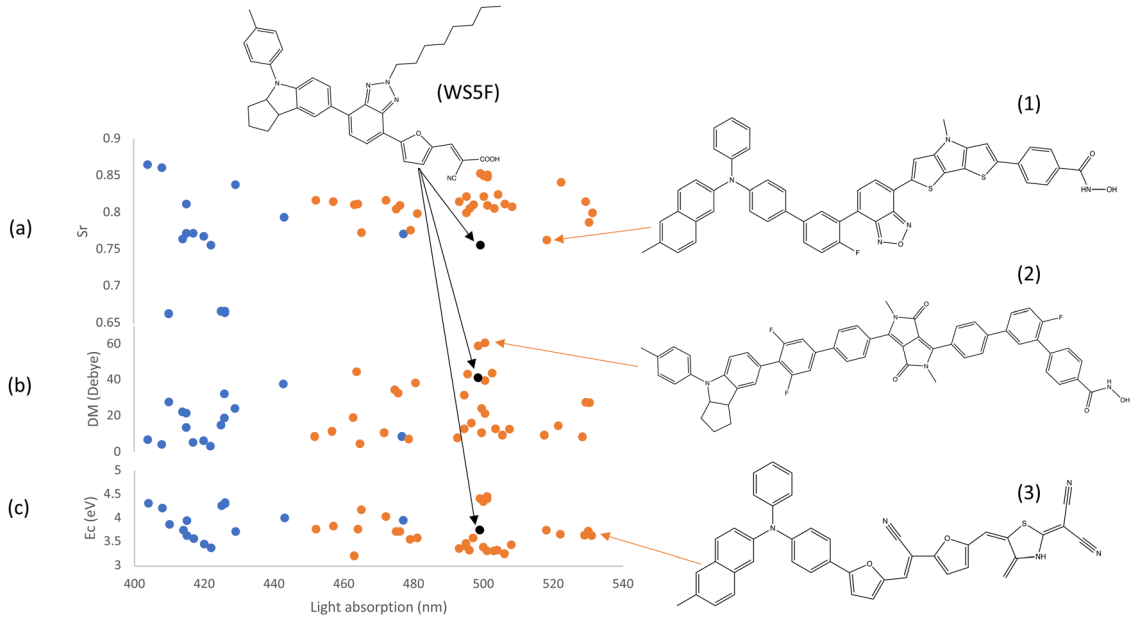

Fig. 8 The properties of the optimized dyes including S_r (a), DM (b) and E_c of electron and hole (c) in the first excited state as a function of E_{abs}. Data in orange is obtained by setting E_{abs} and/or E_b as a fitness function, data in blue is obtained by setting E_d as a fitness function, data in black is the referencing dye WS5F. Promising structures (1)/(2)/(3) are selected by considering both E_{abs} and S_r/DM/E_c. The data in this plot is listed in Table S1 (ESI†).

(HOMO, IP) and lowest unoccupied molecular orbital (LUMO, IP*/EA). Consequently, the optical gap also expands, causing the absorption spectrum to shift towards the high-energy (blue) region, and *vice versa*. Our findings suggest that there might not be a single dye that excels in both light absorption and driving force. Instead, there is likely a balancing point or a narrow region that optimally considers both driving force and absorption.

It has been previously established that in both dye-sensitized photocatalytic hydrogen evolution and dye-sensitized solar cells (DSSCs), the strong chemical binding between the dye and the semiconductor surface plays a crucial role in maximizing the electronic coupling between the excited dye molecular orbitals and the conduction band of the semiconductor. Consequently, sensitizers are firmly anchored to the semiconductor surface through anchoring acceptor groups (*e.g.*, carboxylate and phosphonate).[23,72–74] This strong binding is an additional factor, aside from the driving force, that significantly influences the rate of charge injection from the dye to the photocatalyst. Research by Ambrosio *et al.*[75] demonstrated that a faster injection rate consistently corresponds to a stronger binding energy. This conclusion was drawn from a study involving different anchoring acceptor groups and TiO_2 in the context of DSSCs. In Fig. 7, the size of the data points represents the binding energy between the optimized dyes and FS-COF. Notably, all of the optimized dyes exhibit a stronger connection with FS-COF (E_b = −1 to −16 kcal mol^{-1}) than WS5F (E_b = −0.43 kcal mol^{-1}). Consequently, the enhanced binding of these dyes to FS-COF results in larger electronic coupling and faster charge injection into the photocatalyst compared to WS5F.

Electron–hole overlap (S_r), shown in Fig. 8(a), provides a clear representation of electron and hole distribution as well as their separation. Smaller values are preferable (ranging from 0 to 1; 0.0 indicates no overlap at all, while 1.0 signifies perfect electron–hole overlap). It's worth noting that this calculation is carried out for the first excited state (also with the highest intensity), considering all excitation contributions with coefficients larger than 10^{-4} to ensure accuracy in determining the overlap parameter. Fig. 8(a) reveals that the reference dye WS5F exhibits a relatively small electron–hole overlap (0.76), while the optimized dyes demonstrate a similar or slightly larger overlap, falling within the range of 0.76 to 0.85. However, the difference between WS5F and the other screened dyes is less than 10%. As an illustration, one of the best dyes, denoted as structure (1), showcases an S_r of 0.76, which is slightly larger than that of WS5F but with a better light absorption. The smallest S_r value of screened dye is 0.66, which is much smaller than that of WS5F, but with a poor light absorption. Consequently, relying solely on the electron–hole overlap is insufficient in determining the best dye. As mentioned earlier, the overall performance of the dye is influenced by various parameters, in addition to the electron–hole overlap.

While electron–hole overlap offers insights into the spatial distribution of electrons and holes, examining the dipole moment of the first excited state provides information on the strength of the separation of electrons and holes. This dipole moment arises from the excited electron and hole, taking into account not only their distance but also their respective charges. In Fig. 8(b), it's evident that only 12 of the optimized dyes exhibit a dipole moment equal to or better than that of the reference dye, falling within the range of 30 to 61 Debye.

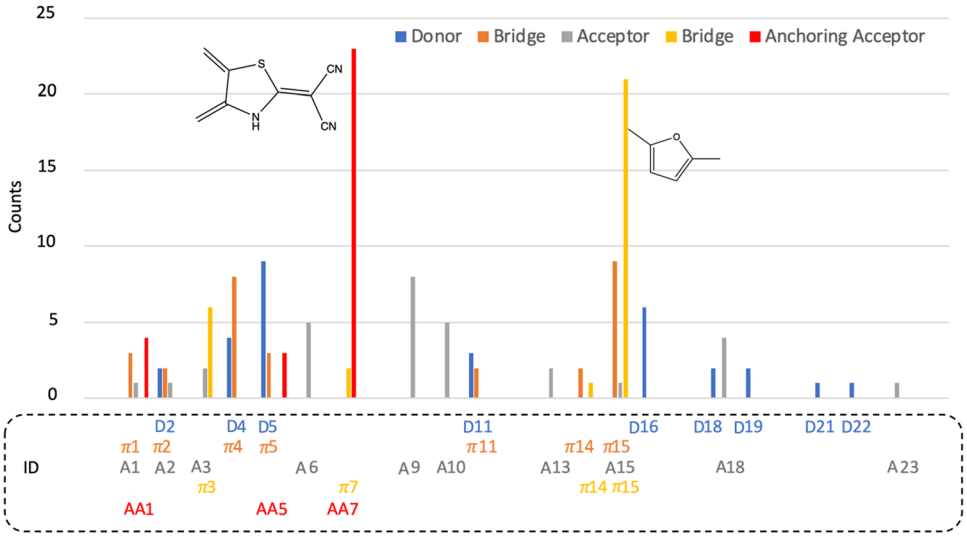

Fig. 9 Statistics of donor (D) (blue), π-bridge (π) (orange) between donor and acceptor, acceptor (A) (grey), π-bridge (π) (yellow) between acceptor and anchoring acceptor, anchoring acceptor (AA) (red) of optimized dyes. The missing IDs are the fragments not selected in our GA process.

This indicates that these dyes have a higher likelihood of keeping the electron and hole separated for an extended period compared to WS5F. This prolonged separation benefits the hot electrons in their journey to be injected into the conduction band of the photocatalyst, rather than recombining back to the ground state. As an illustrative example, one of the best dyes, denoted as structure (2), is shown in Fig. 8(b).

Once a dye reaches an excited state, it becomes crucial for the hot electron to be injected into the conduction band of the photocatalyst rather than recombine with holes, which can lead to a return to the ground state and is one of a major loss mechanism. E_c between the electron and hole is an important parameter to assess the strength of this attraction, with smaller values being more favourable. In Fig. 8(c), it is evident that approximately 65% of the optimized dyes exhibit a smaller electron–hole attraction energy compared to the reference dye WS5F. This suggests that electron–hole separation is likely to be more successful, with a reduced chance of undergoing recombination compared to WS5F in the complex process of water splitting. As an illustrative example, one of the best dyes, denoted as structure (3), is provided.

Fig. 9 provides a statistical overview of the building blocks present in the optimized dyes. Notably, 9 out of 26 donors (blue), 10 out of 25 acceptors (grey), and 7 out of 17 π-bridges (orange) between donors and acceptors are observed in the optimized dyes. However, none of them stand out significantly when compared to others. This is attributed to the fact that several parameters influenced by these three blocks are in competition, impacting the overall performance of the dye. Flattening the donor, acceptor, and the π-bridge between them enhances charge separation (beneficial for charge injection) and conjugated length (favourable for low-energy light absorption). Simultaneously, this strengthens charge recombination, and *vice versa*.

In contrast, for the anchoring acceptor (red) and the π-bridge (yellow) between the acceptor and anchoring acceptor, as shown in Fig. 9, it is evident that π-bridge fragment number π15 (furan) and anchoring acceptor fragment number AA7 (structure with dicyanomethylene) are notably more prevalent among the optimized dyes. In experimental work, Mao *et al.*[76] have previously reported that a stable dye with the acceptor 2-(1,1-dicyanomethylene)rhodamine (lacking COOH) is advantageous for DSSCs efficiency and stability. This structure closely resembles anchoring acceptor fragment number AA7 in our case. This anchoring acceptor comprises two acceptors (a five-membered ring and double cyano groups), ensuring a robust electron-accepting capability. Moreover, the heteroatom is easily able to chelate to the photocatalyst surface, facilitating a strong electron coupling between the dye and the photocatalyst FS-COF. Additionally, the anchor unit is positioned in the middle of the anchoring acceptor block, which minimizes the electron transfer distance from donor to photocatalyst. The use of the π-bridge fragment furan reduces the dihedral angle between the acceptor and anchoring acceptor to maximize the π-conjugated effect, enhancing electron mixing between the acceptor, π-bridge, and anchoring acceptor. This facilitates electron diffusion to the anchoring acceptor, resulting in a substantial electronic coupling between the anchoring acceptor and the photocatalyst. It is intriguing to note that WS5F (as shown in Fig. 8) exhibits a similar structure, with a short furan π-bridge between the acceptor and anchoring acceptor, and these three groups lie on the same plane.

Conclusion

In this study, a combination of genetic algorithm and GNF-xTB methods was employed to systematically screen a vast dataset of 2.6 million dye molecules. The objective was to sensitize the FS-COF photocatalyst for efficient water splitting. Subsequently, crucial parameters like ionization potential, electron affinity,

and light absorption characteristics of the optimized dyes were rigorously re-evaluated at the DFT/TDDFT level. Additionally, properties of the first excited state, including electron–hole overlap, dipole moment, and coulombic attraction energy, were computed and compared against the reference dye WS5F and the photocatalyst FS-COF.

Two distinct groups of dye molecules exhibiting superior properties compared to WS5F were identified. The first group was optimized with a focus on absorption and/or binding energy, resulting in dyes with comparable or enhanced absorption compared to WS5F. The second group was optimized with an emphasis on charge injection driving force, leading to dyes with substantially greater driving forces than WS5F. Furthermore, dyes in both groups demonstrated superior electron–hole overlap, coulombic attraction energy, and/or dipole moment of the excited state, all of which are critical factors in charge separation and recombination processes.

The analysis of building block statistics revealed that the relationship between donor, acceptor, and the π-bridge connecting them is more intricate compared to the anchoring acceptor and the π-bridge between the acceptor and anchoring acceptor. It was observed that the most prevalent anchoring acceptor molecule features two cyano groups in its library, while the most frequently employed π-bridge between the anchoring acceptor and the acceptor is a single furan ring. This structural arrangement proved beneficial for facilitating charge separation/injection and enhancing light absorption.

This work not only establishes a robust GA workflow coupled with GNF-xTB and (TD-)DFT methods for the screening and optimization of a large number of candidate molecules, but it also provides valuable insights into the design principles for anchoring acceptors and the π-bridge between the acceptor and anchoring acceptor. These findings have far-reaching implications and can be broadly applied in various materials science fields, especially in the design of sensitizers, material screening, and high-throughput applications, with particular relevance to dye-sensitized solar cells, organic photovoltaics, and drug design.

Author contributions

T. L. and L. C. conceived the project. T. L. built the screening workflow and performed all the simulations and calculations. T. L, and L. C. interpreted the data. X. W. provided experimental support. T. L., L. C. and A. I. C. wrote the paper.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC) (EP/N004884/1), and the Leverhulme Trust *via* the Leverhulme Research Centre for Functional Materials Design. Computing support from HPC in the University of Liverpool is gratefully acknowledged.

References

1 A. Fujishima and K. Honda, Electrochemical Photolysis of Water at a Semiconductor Electrode, *Nature*, 1972, **238**, 37–38.
2 D. Wang, J. Ye, T. Kako and T. Kimura, Photophysical and Photocatalytic Properties of $SrTiO_3$ Doped with Cr Cations on Different Sites, *J. Phys. Chem. B*, 2006, **110**, 15824–15830.
3 H. Yu, S. Ouyang, S. Yan, Z. Li, T. Yu and Z. Zou, Sol–gel hydrothermal synthesis of visible-light-driven Cr-doped $SrTiO_3$ for efficient hydrogen production, *J. Mater. Chem.*, 2011, **21**, 11347–11351.
4 B. S. Kalanoor, H. Seo and S. S. Kalanur, Recent developments in photoelectrochemical water-splitting using WO_3/$BiVO_4$ heterojunction photoanode: A review, *Mater. Sci. Energy Technol.*, 2018, **1**, 49–62.
5 J. Wiktor, F. Ambrosio and A. Pasquarello, Role of Polarons in Water Splitting: The Case of $BiVO_4$, *ACS Energy Lett.*, 2018, **3**, 1693–1697.
6 J. Xing, W. Q. Fang, H. J. Zhao and H. G. Yang, Inorganic Photocatalysts for Overall Water Splitting, *Chem. – Asian J.*, 2012, **7**, 642–657.
7 S. Y. Tee, K. Y. Win, W. S. Teo, L.-D. Koh, S. Liu, C. P. Teng and M.-Y. Han, Recent Progress in Energy-Driven Water Splitting, *Adv. Sci.*, 2017, **4**, 1600337.
8 K. Sharma, V. Sharma and S. S. Sharma, Dye-Sensitized Solar Cells: Fundamentals and Current Status, *Nanoscale Res. Lett.*, 2018, **13**, 381.
9 B. O'Regan and M. Grätzel, A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films, *Nature*, 1991, **353**, 737–740.
10 X. Wang, S. Blechert and M. Antonietti, Polymeric Graphitic Carbon Nitride for Heterogeneous Photocatalysis, *ACS Catal.*, 2012, **2**, 1596–1606.
11 J. Liu, Y. Liu, N. Liu, Y. Han, X. Zhang, H. Huang, Y. Lifshitz, S.-T. Lee, J. Zhong and Z. Kang, Metal-free efficient photocatalyst for stable visible water splitting via a two-electron pathway, *Science*, 2015, **347**, 970.
12 R. S. Sprick, J.-X. Jiang, B. Bonillo, S. Ren, T. Ratvijitvech, P. Guiglion, M. A. Zwijnenburg, D. J. Adams and A. I. Cooper, Tunable Organic Photocatalysts for Visible-Light-Driven Hydrogen Evolution, *J. Am. Chem. Soc.*, 2015, **137**, 3265–3270.
13 C. M. Aitchison, R. S. Sprick and A. I. Cooper, Emulsion polymerization derived organic photocatalysts for improved light-driven hydrogen evolution, *J. Mater. Chem. A*, 2019, **7**, 2490–2496.
14 J. Bi, W. Fang, L. Li, J. Wang, S. Liang, Y. He, M. Liu and L. Wu, Covalent Triazine-Based Frameworks as Visible Light Photocatalysts for the Splitting of Water, *Macromol. Rapid Commun.*, 2015, **36**, 1799–1805.

15 K. Schwinghammer, S. Hug, M. B. Mesch, J. Senker and B. V. Lotsch, Phenyl-triazine oligomers for light-driven hydrogen evolution, *Energy Environ. Sci.*, 2015, **8**, 3345–3353.
16 X. Wang, L. Chen, S. Y. Chong, M. A. Little, Y. Wu, W.-H. Zhu, R. Clowes, Y. Yan, M. A. Zwijnenburg, R. S. Sprick and A. I. Cooper, Sulfone-containing covalent organic frameworks for photocatalytic hydrogen evolution from water, *Nat. Chem.*, 2018, **10**, 1180–1189.
17 H. M. El-Kaderi, J. R. Hunt, J. L. Mendoza-Cortés, A. P. Côté, R. E. Taylor, M. O'Keeffe and O. M. Yaghi, Designed Synthesis of 3D Covalent Organic Frameworks, *Science*, 2007, **316**, 268.
18 C. S. Diercks and O. M. Yaghi, The atom, the molecule, and the covalent organic framework, *Science*, 2017, **355**, eaal1585.
19 A. P. Côté, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger and O. M. Yaghi, Porous, Crystalline, Covalent Organic Frameworks, *Science*, 2005, **310**, 1166.
20 T. Banerjee, F. Haase, G. Savasci, K. Gottschling, C. Ochsenfeld and B. V. Lotsch, Single-Site Photocatalytic H2 Evolution from Covalent Organic Frameworks with Molecular Cobaloxime Co-Catalysts, *J. Am. Chem. Soc.*, 2017, **139**, 16228–16234.
21 Y. Song, Q. Sun, B. Aguila and S. Ma, Opportunities of Covalent Organic Frameworks for Advanced Applications, *Adv. Sci.*, 2019, **6**, 1801410.
22 T. Sick, A. G. Hufnagel, J. Kampmann, I. Kondofersky, M. Calik, J. M. Rotter, A. Evans, M. Döblinger, S. Herbert, K. Peters, D. Böhm, P. Knochel, D. D. Medina, D. Fattakhova-Rohlfing and T. Bein, Oriented Films of Conjugated 2D Covalent Organic Frameworks as Photocathodes for Water Splitting, *J. Am. Chem. Soc.*, 2018, **140**, 2085–2092.
23 X. Zhang, T. Peng and S. Song, Recent advances in dye-sensitized semiconductor systems for photocatalytic hydrogen production, *J. Mater. Chem. A*, 2016, **4**, 2365–2402.
24 A. Carella, F. Borbone and R. Centore, Research Progress on Photosensitizers for DSSC, *Front. Chem.*, 2018, **6**, 481.
25 T. Liu, L. Chen, X. Li and Andrew I. Cooper, Investigating the factors that influence sacrificial hydrogen evolution activity for three structurally-related molecular photocatalysts: thermodynamic driving force, excited-state dynamics, and surface interaction with cocatalysts, *Phys. Chem. Chem. Phys.*, 2023, **25**, 3494–3501.
26 P. Ganesan, A. Yella, T. W. Holcombe, P. Gao, R. Rajalingam, S. A. Al-Muhtaseb, M. Grätzel and M. K. Nazeeruddin, Unravel the Impact of Anchoring Groups on the Photovoltaic Performances of Diketopyrrolopyrrole Sensitizers for Dye-Sensitized Solar Cells, *ACS Sustainable Chem. Eng.*, 2015, **3**, 2389–2396.
27 M. Velusamy, K. R. Justin Thomas, J. T. Lin, Y.-C. Hsu and K.-C. Ho, Organic Dyes Incorporating Low-Band-Gap Chromophores for Dye-Sensitized Solar Cells, *Org. Lett.*, 2005, **7**, 1899–1902.
28 S. Chaurasia, J.-S. Ni, W.-I. Hung and J. T. Lin, 2*H*-[1,2,3]Triazolo[4,5-*c*]pyridine Cored Organic Dyes Achieving a High Efficiency: a Systematic Study of the Effect of Different Donors and π Spacers, *ACS Appl. Mater. Interfaces*, 2015, **7**, 22046–22057.
29 W. Zhu, Y. Wu, S. Wang, W. Li, X. Li, J. Chen, Z. Wang and H. Tian, Organic D–A–π–A Solar Cell Sensitizers with Improved Stability and Spectral Response, *Adv. Funct. Mater.*, 2011, **21**, 756–763.
30 Y. Wu and W. Zhu, Organic sensitizers from D–π–A to D–A–π–A: effect of the internal electron-withdrawing units on molecular absorption, energy levels and photovoltaic performances, *Chem. Soc. Rev.*, 2013, **42**, 2039–2058.
31 B.-G. Kim, K. Chung and J. Kim, Molecular Design Principle of All-organic Dyes for Dye-Sensitized Solar Cells, *Chem. – Eur. J.*, 2013, **19**, 5220–5230.
32 Y. Adachi, Y. Ooyama, N. Shibayama and J. Ohshita, Dithienogermole-containing D–π–A–π–A Photosensitizers for Dye-sensitized Solar Cells, *Chem. Lett.*, 2016, **46**, 310–312.
33 M. K. R. Fischer, S. Wenger, M. Wang, A. Mishra, S. M. Zakeeruddin, M. Grätzel and P. Bäuerle, D–π–A Sensitizers for Dye-Sensitized Solar Cells: Linear *vs.* Branched Oligothiophenes, *Chem. Mater.*, 2010, **22**, 1836–1845.
34 W. Banzhaf, F. D. Francone, R. E. Keller and P. Nordin, *Genetic Programming: An Introduction: on the Automatic Evolution of Computer Programs and Its Applications*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.
35 R. R. Bies, M. F. Muldoon, B. G. Pollock, S. Manuck, G. Smith and M. E. Sale, A genetic algorithm-based, hybrid machine learning approach to model selection, *J. Pharmacokinet. Pharmacodyn.*, 2006, **33**, 195–221.
36 T. C. Le and D. A. Winkler, Discovery and Optimization of Materials Using Evolutionary Approaches, *Chem. Rev.*, 2016, **116**, 6107–6132.
37 V. Meenakshisundaram, J.-H. Hung, T. K. Patra and D. S. Simmons, Designing Sequence-Specific Copolymer Compatibilizers Using a Molecular-Dynamics-Simulation-Based Genetic Algorithm, *Macromolecules*, 2017, **50**, 1155–1166.
38 C. L. Tsai, K. T. Delaney and G. H. Fredrickson, Genetic Algorithm for Discovery of Globally Stable Phases in Block Copolymers, *Macromolecules*, 2016, **49**, 6558–6567.
39 A. Kaczmarowski, S. Yang, I. Szlufarska and D. Morgan, Genetic algorithm optimization of defect clusters in crystalline materials, *Comput. Mater. Sci.*, 2015, **98**, 234–244.
40 M. P. Lourenço, J. Hostaš, L. B. Herrera, P. Calaminici, A. M. Köster, A. Tchagang and D. R. Salahub, GAMaterial—A genetic-algorithm software for material design and discovery, *J. Comput. Chem.*, 2023, **44**, 814–823.
41 A. Jain, I. E. Castelli, G. Hautier, D. H. Bailey and K. W. Jacobsen, Performance of genetic algorithms in search for water splitting perovskites, *J. Mater. Sci.*, 2013, **48**, 6519–6534.
42 N. M. O'Boyle, C. M. Campbell and G. R. Hutchison, Computational Design and Selection of Optimal Organic Photovoltaic Materials, *J. Phys. Chem. C*, 2011, **115**, 16200–16210.
43 P. C. Jennings, S. Lysgaard, J. S. Hummelshøj, T. Vegge and T. Bligaard, Genetic algorithms for computational materials discovery accelerated by machine learning, *npj Comput. Mater.*, 2019, **5**, 46.
44 S. Grimme, C. Bannwarth and P. Shushkov, A Robust and Accurate Tight-Binding Quantum Chemical Method for

Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86), *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.

45 L. Wilbraham, E. Berardo, L. Turcani, K. E. Jelfs and M. A. Zwijnenburg, High-Throughput Screening Approach for the Optoelectronic Properties of Conjugated Polymers, *J. Chem. Inf. Model.*, 2018, **58**, 2450–2459.

46 Y. Bai, L. Wilbraham, B. J. Slater, M. A. Zwijnenburg, R. S. Sprick and A. I. Cooper, Accelerated Discovery of Organic Polymer Photocatalysts for Hydrogen Evolution from Water through the Integration of Experiment and Theory, *J. Am. Chem. Soc.*, 2019, **141**, 9063–9071.

47 V. Ásgeirsson, C. A. Bauer and S. Grimme, Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules, *Chem. Sci.*, 2017, **8**, 4879–4895.

48 C. Bannwarth and S. Grimme, A simplified time-dependent density functional theory approach for electronic ultraviolet and circular dichroism spectra of very large molecules, *Comput. Theor. Chem.*, 2014, **1040–1041**, 45–53.

49 S. Grimme, A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules, *J. Chem. Phys.*, 2013, **138**, 244104.

50 S. Grimme and C. Bannwarth, Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB), *J. Chem. Phys.*, 2016, **145**, 054103.

51 D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.

52 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, Open Babel: An open chemical toolbox, *J. Cheminf.*, 2011, **3**, 33.

53 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.*, 1998, **19**, 1639–1662.

54 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J. Comput. Chem.*, 2009, **30**, 2785–2791.

55 E. Runge and E. K. U. Gross, Density-Functional Theory for Time-Dependent Systems, *Phys. Rev. Lett.*, 1984, **52**, 997–1000.

56 R. G. Parr, in *Horizons of Quantum Chemistry*, ed. K. Fukui and B. Pullman, Springer, Netherlands, 1980, pp.5–15.

57 A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

58 T. Yanai, D. P. Tew and N. C. Handy, A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP), *Chem. Phys. Lett.*, 2004, **393**, 51–57.

59 P. Guiglion, C. Butchosa and M. A. Zwijnenburg, Polymeric watersplitting photocatalysts; A computational perspective on the water oxidation conundrum, *J. Mater. Chem. A*, 2014, **2**, 11996–12004.

60 A. V. Marenich, C. J. Cramer and D. G. Truhlar, Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.

61 F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

62 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision C.01*, 2016.

63 T. Lu and F. Chen, Multiwfn: A multifunctional wavefunction analyzer, *J. Comput. Chem.*, 2012, **33**, 580–592.

64 C. Fan, M. Springborg and Y. Feng, Application of an inverse-design method to optimizing porphyrins in dye-sensitized solar cells, *Phys. Chem. Chem. Phys.*, 2019, **21**, 5834–5844.

65 C. Fan, M. Molayem, M. Springborg, M. Kick and Y. Feng, Role of the Backbone when Optimizing Functional Groups—A Theoretical Study Based on an Improved Inverse-Design Approach, *J. Phys. Chem. A*, 2022, **126**, 1289–1299.

66 A. S. Khazaal, M. Springborg, C. Fan and K. Huwig, Optimizing small conjugated molecules for solar-cell applications using an inverse-design method, *J. Mol. Graphics Modell.*, 2020, **100**, 107654.

67 P. Guiglion, C. Butchosa and M. A. Zwijnenburg, Polymer Photocatalysts for Water Splitting: Insights from Computational Modeling, *Macromol. Chem. Phys.*, 2016, **217**, 344–353.

68 M. Sachs, R. S. Sprick, D. Pearce, S. A. J. Hillman, A. Monti, A. A. Y. Guilbert, N. J. Brownbill, S. Dimitrov, X. Shi, F. Blanc, M. A. Zwijnenburg, J. Nelson, J. R. Durrant and A. I. Cooper, Understanding structure-activity relationships in linear polymer photocatalysts for hydrogen evolution, *Nat. Commun.*, 2018, **9**, 4968.

69 J. R. Swierk, K. P. Regan, J. Jiang, G. W. Brudvig and C. A. Schmuttenmaer, Rutile TiO_2 as an Anode Material

for Water-Splitting Dye-Sensitized Photoelectrochemical Cells, *ACS Energy Lett.*, 2016, **1**, 603–606.

70 O. Suryani, Y. Higashino, H. Sato and Y. Kubo, Visible-to-Near-Infrared Light-Driven Photocatalytic Hydrogen Production Using Dibenzo-BODIPY and Phenothiazine Conjugate as Organic Photosensitizer, *ACS Appl. Energy Mater.*, 2019, **2**, 448–458.

71 P. Acosta-Mora, K. Domen, T. Hisatomi, H. Lyu, J. Méndez-Ramos, J. C. Ruiz-Morales and N. M. Khaidukov, ''A bridge over troubled gaps'': up-conversion driven photocatalysis for hydrogen generation and pollutant degradation by near-infrared excitation, *Chem. Commun.*, 2018, **54**, 1905–1908.

72 N. Manfredi, M. Monai, T. Montini, F. Peri, F. De Angelis, P. Fornasiero and A. Abbotto, Dye-Sensitized Photocatalytic Hydrogen Generation: Efficiency Enhancement by Organic Photosensitizer–Coadsorbent Intermolecular Interaction, *ACS Energy Lett.*, 2018, **3**, 85–91.

73 N. Martsinovich and A. Troisi, Theoretical studies of dye-sensitised solar cells: from electronic structure to elementary processes, *Energy Environ. Sci.*, 2011, **4**, 4473–4495.

74 A. Hagfeldt, G. Boschloo, L. Sun, L. Kloo and H. Pettersson, Dye-Sensitized Solar Cells, *Chem. Rev.*, 2010, **110**, 6595–6663.

75 F. Ambrosio, N. Martsinovich and A. Troisi, What Is the Best Anchoring Group for a Dye in a Dye-Sensitized Solar Cell?, *J. Phys. Chem. Lett.*, 2012, **3**, 1531–1535.

76 J. Mao, N. He, Q. Zhang, F. Guo, L. Chen and W. Wu, Stable dyes containing double acceptors without COOH as anchors for highly efficient dye-sensitized solar cells, *Angew. Chem., Int. Ed.*, 2012, **51**, 9873–9876.

