



Cite this: *Phys. Chem. Chem. Phys.*, 2022, 24, 197

# Automated assessment of redox potentials for dyes in dye-sensitized photoelectrochemical cells†

Jelena Belić, <sup>a</sup> Arno Förster, <sup>a</sup> Jan Paul Menzel, <sup>b</sup> Francesco Buda <sup>b</sup> and Lucas Visscher <sup>\*a</sup>

Sustainable solutions for hydrogen production, such as dye-sensitized photoelectrochemical cells (DS-PEC), rely on the fundamental properties of its components whose modularity allows for their separate investigation. In this work, we design and execute a high-throughput scheme to tune the ground state oxidation potential (GSOP) of perylene-type dyes by functionalizing them with different ligands. This allows us to identify promising candidates which can then be used to improve the cell's efficiency. First, we investigate the accuracy of different theoretical approaches by benchmarking them against experimentally determined GSOPs. We test different methods to calculate the vertical oxidation potential, including GW with different levels of self-consistency, Kohn–Sham (KS) orbital energies and total energy differences. We find that there is little difference in the performance of these methods. However, we show that it is crucial to take into account solvent effects as well as the structural relaxation of the dye after oxidation. Other thermodynamic contributions are negligible. Based on this benchmark, we decide on an optimal strategy, balancing computational cost and accuracy, to screen more than 1000 dyes and identify promising candidates which could be used to construct more robust DS-PECs.

Received 15th September 2021,  
 Accepted 30th November 2021

DOI: 10.1039/d1cp04218a

[rsc.li/pccp](http://rsc.li/pccp)

## 1 Introduction

Hydrogen is seen as a symbol of sustainable energy source, even though its production currently mostly relies on fossil fuels.<sup>1,2</sup> To make production of this energy carrier more sustainable,<sup>3–5</sup> a promising route is to split water with a dye-sensitized photoelectrochemical cell (DS-PEC) driven by solar energy.<sup>6,7</sup> Such developments capitalize on the advantages of dye-sensitization, in particular the possibility to tune the dye's properties by small structural adjustments.<sup>7</sup>

In the present work we will focus on the dye on the photoanode of DS-PECs, for which the absorbed sunlight creates a photo-excited electron that is injected into the photo anode (usually TiO<sub>2</sub>). The thus oxidized dye is restored to its initial state by accepting an electron from a water oxidation catalysts (WOC), which then starts the water oxidation cycle.<sup>7</sup> To make

this process efficient the dye has to fulfill several requirements. Prerequisite for the chain of electron transfers (from the dye to the semiconductor and from the WOC to the dye) to proceed is a suitable alignment of the dye's redox potentials with the neighbouring components: (1) the ground state oxidation potential (GSOP) has to be higher than the highest oxidation potential (HOP) of the WOC and (2) the excited state oxidation potential (ESOP) has to be lower than the edge of the anode's conduction band (CB) (Fig. 1). While these can be considered as minimal requirements, one may further narrow the search for optimal dyes by estimating that for reducing the energy loss and for an optimal rate of the electron transfer the potential of the ESOP state needs to be just slightly higher ( $\sim 0.3$  eV)<sup>8</sup> than the semiconductor's CB.

A further boundary condition for sustainable deployment of DS-PECs is the use of dyes made entirely from abundant material, such as fully organic dyes. Among these, perylene diimides (PDI) have been recognised as very promising for use in photovoltaic devices<sup>7,9</sup> due to their favourable optical and electrochemical properties, as well as their stability, relatively cheap synthesis and tunability. They have found a place in different types of photo-electrochemical devices, both at the anode<sup>10–13</sup> and at the cathode.<sup>14</sup> Also other dyes related to the perylene diimide have proven to be efficient components of such systems, like naphthalene diimide (NDI)<sup>15</sup> or perylene triimides (PTI).<sup>16</sup>

<sup>a</sup> Department of Chemistry and Pharmaceutical Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV, Amsterdam, The Netherlands.  
 E-mail: [L.visscher@vu.nl](mailto:L.visscher@vu.nl)

<sup>b</sup> Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, P.O. Box 9502, 2300 RA, Leiden, The Netherlands

† Electronic supplementary information (ESI) available: Experimental data set (S1), *R*<sup>2</sup> sensitivity (S2), calculated values of GSOP (S3), PDI-0000 outlier analysis (S4) and different contributions to GSOP (S5), screening results (S6). See DOI: 10.1039/d1cp04218a



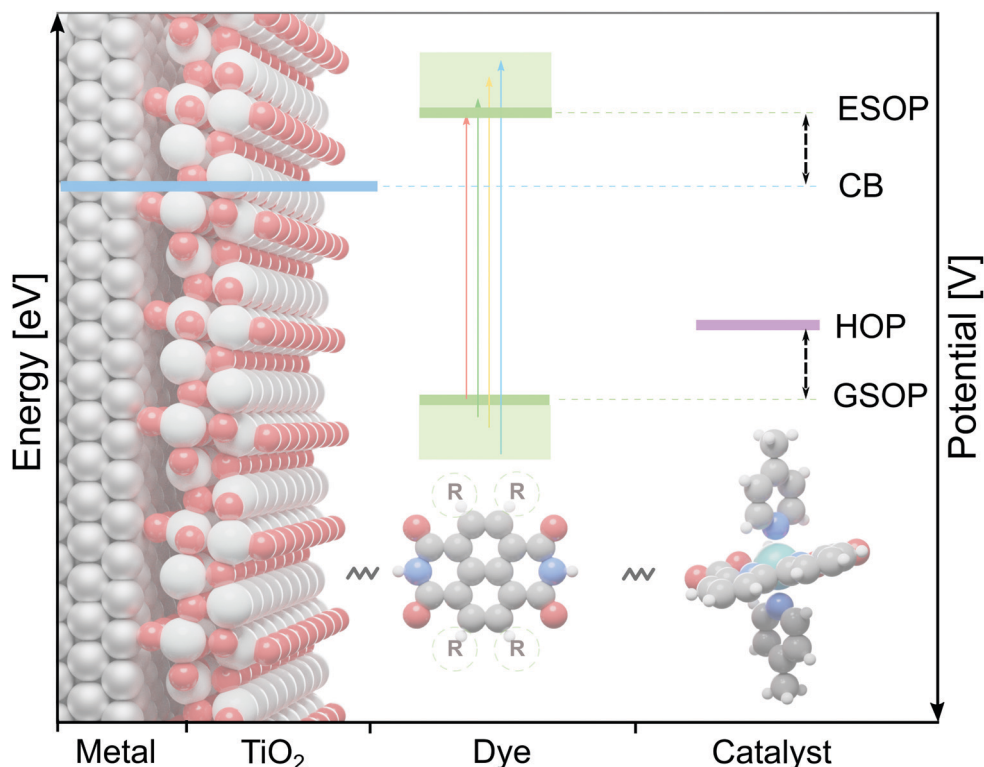


Fig. 1 The alignment between redox potentials of the DS-PEC's components: conduction band (CB) edge of the anode (blue), ground state oxidation potential (GSOP) and excited state oxidation potential (ESOP) of the dye (green) and highest oxidation potential (HOP) of the water oxidation catalyst (purple).

Given this versatility of the perylene group of dyes, there are thousands of dyes that could be synthesized and tested for their performance. To better narrow down the search for optimal performers, the use of computational modelling is of interest. With a clear idea and well-defined objectives, such a computational screening of the chemical space can strongly reduce the amount of trial experiments. Important is thereby to take into account the uncertainties that the construction of model and the choice of theoretical methods bring. In a complex system there are many factors that affect the final performance and theory should merely narrow down the search rather than finding the perfect candidate. A great advantage is the modularity of the DS-PEC system, thus one may study its molecular components separately and define suitable components before constructing an overall model. Studies on molecular components therefore encompass dyes with favourable light absorption properties,<sup>17</sup> anchoring groups between the dye and the semiconductor for ultrafast electron injection<sup>18</sup> and the WOC-dye complex for driving the water splitting process.<sup>19,20</sup> Analysis and screening of the individual components will help to define optimal DS-PECs and to improve upon the hydrogen production efficiency. In this work we focus on finding dyes that fit the system's thermodynamic requirements, in particular those that have a suitable GSOP.

The GSOP can be predicted by calculating the Gibbs free energy of the oxidation reaction,<sup>21</sup> the difference between the solution-phase Gibbs free energies of a molecule in neutral and

oxidized form. We can access these thermodynamic properties using generalized<sup>22,23</sup> Kohn-Sham (KS)<sup>24</sup> density functional theory (DFT).<sup>25</sup> As reactions take place in a medium, solvent effects need to be accounted for which is most efficiently done with continuum solvation (CS) models.<sup>26</sup> Furthermore we may distinguish between two major pathways that are commonly used to calculate solution-phase Gibbs free energy: use of a thermodynamic cycle (TC) and direct computation (DC) within a CS model.

Simpler than these two thermodynamical approaches is to approximate the GSOP by the vertical ionization energy of the solvated dye. Such a calculation can be done in three ways, one may compute the electronic energy difference by two separate KS-DFT calculations of the molecule in its two forms (before and after oxidation), or one may take the GSOP as the negative of the highest occupied molecular orbital (HOMO) energy of the neutral molecule. For the latter approach, it is important to take into account that semi-local or hybrid approximations to the exact functional of DFT do not describe the physics of addition or removal of electrons fully correctly.<sup>27-30</sup> A theoretically more rigorous alternative is many-body perturbation theory (MBPT) in the *GW* approximation (GWA) to the electronic self-energy,<sup>31</sup> where *G* denotes the single-particle Green's function and *W* is the screened Coulomb interaction. As in approaches based on total energy differences, the GWA accounts for the systems response to the hole created by ionization.<sup>32</sup> It yields QP energies which can be directly identified with vertical



electron addition and removal energies<sup>33</sup> and is therefore increasingly used for energy level alignment in photo-catalytic interfaces<sup>34,35</sup> or dye-sensitized solar cells.<sup>36–39</sup>

The most commonly used way to validate theoretical predictions of the GSOP is a comparison to the oxidation potential measured by cyclic voltammetry (CV). The half-reaction (Dye → Dye<sup>+</sup> + e<sub>anode</sub><sup>-</sup>) taking place at the surface of the anode, with the dye redox couple [Dye/Dye<sup>+</sup>] rapidly exchanging electrons with the electrode, is thereby considered to be electrochemically reversible. The measured half-wave potential  $E_{1/2}$  of the reaction is, at standard conditions, a very good estimate for the standard electrode potential:  $E_{1/2} \approx E^\circ$ ,<sup>40</sup> with  $E^\circ$  being the driving force for electrochemical work.

As we are primarily interested in the GSOP here, the measured  $E_{1/2}$  potentials can directly be used as experimental reference data to validate the accuracies of the different methods to calculate the GSOPs. Treatment of the ESOP is a bit more involved as this depends on the GSOP as well as on the energy difference between the ground and excited state. The latter can be computed as the adiabatic energy difference between the ground state and the geometry-relaxed excited state, but also as a vertical excitation energy under the assumption that the electron transfer is much faster than the nuclear relaxation (Franck–Condon principle). Which picture is the most appropriate depends on the dye-semiconductor system; it has been observed for some semiconductors that the injection happens only after the relaxation.<sup>8</sup> One may also note that for certain dyes the excited-state reorganisation energy is very small, making the differences between the schemes less important.<sup>41</sup>

In this work we aim to construct and validate a fast procedure for the calculation of GSOPs which can be used for the automated screening of the GSOPs for different sets of dyes and their derivatives. In this construction step we consider accuracy as well as computational efficiency and discuss the effect of different approximations on these factors. This work represents a step towards full automation for the screening of dyes for use in DS-PECs – an automated strategy to obtain accurate energetics of the system's interfaces.

## 2 Methods

In this section we will discuss the possible computational strategies for GSOP determination and the experimental data that can be used to validate them. We will start with the two adiabatic strategies directed towards full determination of the Gibbs free energy, consider in the second section approaches that use the vertical ionization potential, and end with the selection of the experimental reference data.

### 2.1 GSOP – adiabatic approach

In the adiabatic approach, the GSOP is defined as the absolute difference between the solution-phase Gibbs free energies of the products and reactants, Dye and Dye<sup>+</sup> + e<sup>-</sup>. It describes the energy change between the molecular species in their respective solvation and thermal equilibrium conditions. The Gibbs

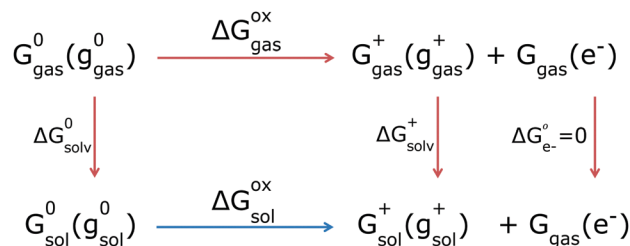


Fig. 2 The thermodynamic cycle. The molecule's total Gibbs free energies are denoted with  $G$  and molecule's geometry with  $g$  in brackets; the subscript is the phase and in superscript is the oxidation state of the species. Solvation processes are the vertical and oxidation processes are the horizontal reactions. Next to the arrows are changes in Gibbs free energies for each process. The free electron is by electrochemical convention always in gas phase.

free energy of the electron depends on the choice of the statistical mechanical formalism – which does not matter as long as it is consistent with the formalism of the reference electrode, which we will address later. For the Fermi–Dirac statistics this value is  $-3.62 \text{ kJ mol}^{-1}$  ( $-0.0375 \text{ eV}$ ).<sup>42,43</sup> Solution-phase Gibbs free energies  $G_{\text{sol}}^i(g_{\text{sol}}^i)$ , where the index  $i$  labels the neutral (0) or oxidized (+) forms of the dye and  $g$  stands for the optimized geometry of the species, can be calculated using DFT. We thereby distinguish between the dye's energy  $E_{\text{sol}}^i(g_{\text{sol}}^i)$  and thermal contributions  $G_{\text{therm}}^i(g_{\text{sol}}^i, T)$  at the temperature  $T$  of interest:

$$G_{\text{sol}}^i(g_{\text{sol}}^i) = E_{\text{sol}}^i(g_{\text{sol}}^i) + G_{\text{therm}}^i(g_{\text{sol}}^i, T) \quad (1)$$

The subscripts hereby indicate the phase of the species: solvated (sol) or in gas phase.

The thermal contribution is the most demanding term in the calculation as it requires determination of the vibrational frequencies of the molecule. The  $G_{\text{therm}}^i(g_{\text{sol}}^i, T)$  is computed within an ideal gas model and includes, besides zero-point vibrations, the translational and rotational contributions to nuclear kinetic energy. In the computation of the entropy term the quasi-rigid-rotor-harmonic-oscillator approximation model<sup>44</sup> for weak (less than  $20 \text{ cm}^{-1}$ ) vibrational modes is employed to avoid artifacts due to inaccuracies in these modes.

Given these individual Gibbs free energies, the GSOP can then be calculated as the difference ( $\Delta G_{\text{sol}}^{\text{ox}}$ ) between the solution-phase Gibbs free energies of a neutral and charged species (Fig. 2, bottom reaction, blue arrow). We will refer to this way of calculation as direct computation (DC). This procedure is rather straightforward and the final value of the adiabatic GSOP with DC and COSMO<sup>45</sup> as CS model,  $\Delta G_{\text{COSMO}}^{\text{DC}}$  is calculated using eqn (2):

$$\Delta G_{\text{COSMO}}^{\text{DC}} = G_{\text{sol}}^+(g_{\text{sol}}^+) + G_{\text{gas}}(e^-) - G_{\text{sol}}^0(g_{\text{sol}}^0) \quad (2)$$

$G_{\text{sol}}^+(g_{\text{sol}}^+)$  and  $G_{\text{sol}}^0(g_{\text{sol}}^0)$  are solution-phase Gibbs free energies of the oxidised and neutral molecular species, respectively.

An alternative is the use of a thermodynamic cycle (TC) method (Fig. 2, top reactions, red arrows). In the TC method, a gas-phase Gibbs free energy is augmented by solvation contributions to yield a solution-phase Gibbs free energy. For each



oxidation state (neutral, 0, or ionized, +) three separate contributions are computed: the gas-phase Gibbs free energy  $G_{\text{gas}}^i(g_{\text{gas}}^i)$  (following eqn (1) for gas phase), the Gibbs free energy of solvation computed at the gas phase geometry  $g_{\text{gas}}^i$ , and the energy difference between the solvated and gas phase structure:

$$G_{\text{sol}}^i(g_{\text{gas}}^i) = G_{\text{gas}}^i(g_{\text{gas}}^i) + \Delta G_{\text{solv}}^i(g_{\text{gas}}^i) + [E_{\text{gas}}(g_{\text{sol}}^i) - E_{\text{gas}}(g_{\text{gas}}^i)] \quad (3)$$

The equation for the adiabatic GSOP computed with the TC method and COSMO,  $\Delta G_{\text{COSMO}}^{\text{TC}}$  is then similar to eqn (2) and reads:

$$\Delta G_{\text{COSMO}}^{\text{TC}} = G_{\text{sol}}^+(g_{\text{gas}}^+) + G_{\text{gas}}(e^-) - G_{\text{sol}}^0(g_{\text{gas}}^0) \quad (4)$$

The DC and TC strategies can be expected to give close agreement if the structure and vibrational frequencies of the dye are only weakly affected by solvation. If the stationary points in solution strongly differ from the ones in the gas phase, the DC and TC approach will deviate from each other.<sup>46,47</sup> For the thermal correction the use of a thermodynamic cycle is then often considered to be more robust<sup>48</sup> as CS models are parametrised to yield accurate Gibbs free energies of solvation (Fig. 2, vertical red lines).<sup>46,49–52</sup> This implies that the thermal contributions calculated in a dielectric continuum in the (DC strategy) can contain corrections that are already incorporated in the CS model (*via* parametrization) which might result in double counting for some solvent effects. Ho<sup>47</sup> compared the thermal contributions (particularly vibrational corrections) obtained by the two strategies, TC and DC, and showed that mean absolute errors in the redox potential for the two strategies differ by 10 mV from each other. In this work he also considered introducing a higher-level electronic structure theory, but found that this did not necessarily lead to better agreement with experiment, indicating that the description of solvation *via* a CS model is the main source of errors. Regarding the choice between the DC and TC approach, we furthermore note that in the TC approach evaluation of multiple solvents is more economical as the most demanding step of the calculation, determination of vibrational frequencies, is independent of the chosen solvent. If only one solvent is to be considered, the TC and DC strategies are similar in terms of computational costs. In DC we avoid the separate calculations for the gas and solution phase that are needed in the TC approach, but the effort needed for frequency calculations in solution is typically higher than in gas phase.

As a cheaper alternative in the TC pathway we can also employ the conductor-like screening model for realistic solvents (COSMO-RS),<sup>53,54</sup> which performs statistical thermodynamics for the molecular surface interactions after the quantum chemical calculation. While COSMO provides the Gibbs free energy of solvation only (vertical red reaction), COSMO-RS provides the chemical potential of a solvated molecule and allows calculating all fluid phase equilibrium thermodynamic properties which includes the solution-phase Gibbs free energy as well.<sup>50,55</sup> The low cost of this model lies in statistical thermodynamics part whose cost is negligible compared to quantum

chemical calculations and which makes the explicit calculation of thermal contributions redundant, as it directly provides the solution-phase Gibbs free energy. The solution-phase Gibbs free energy thus reduces to a sum of the gas-phase energy and a COSMO-RS Gibbs free energy correction. The geometry changes due to solvation are ignored,

$$G_{\text{sol,CRS}}^i(g_{\text{gas}}^i) = E_{\text{gas}}^i(g_{\text{gas}}^i) + \Delta G_{\text{CRS,solv}}^i(g_{\text{gas}}^i) \quad (5)$$

Inserting the solution-phase Gibbs free energy defined in eqn (5) into eqn (4), the adiabatic GSOP using the TC pathway in combination with COSMO-RS is obtained.

Finally, to be able to compute a large number of molecules, we can reduce the time spent for geometry optimization. A workflow combining semiempirical quantum mechanical (SQM) techniques for geometry optimisations and frequency calculations and DFT electronic energy with COSMO is shown to provide reliable predictions of redox potentials.<sup>56,57</sup> Similarly, we can employ the SQM method instead of DFT for geometry optimisation and combine it with DFT electronic energies and the COSMO-RS solvation model. The combination of SQM methods with a COSMO-RS thermodynamic calculation is very efficient and the limiting step in this workflow will be the single point DFT calculation including COSMO. A dedicated COSMO calculation is needed to define the molecule's reference state that is used in further COSMO-RS statistical thermodynamics calculations. As COSMO-RS provides the solution-phase Gibbs free energy, no frequency calculations are needed. We will denote the result of this composite method as  $\Delta G_{\text{COSMO-RS}}^{\text{screening}}$ .

## 2.2 GSOP – vertical approach

The Gibbs free energy difference should correspond exactly to the GSOP measured in the CV experiments as the time scale of these experiments is long enough to allow for full relaxation. Computing the GSOP in the vertical approximation and neglecting vibrational effects is computationally attractive and can be considered a valid approximation if the system is not in thermodynamical equilibrium, which will be the case when the electron transfer process is fast enough. Physically, the relaxation of the neutral geometry after electron removal is then neglected and only relaxation of the electrons as a response to the removal of an electron to the system is accounted for ref. 32. In this regime one can consider three different strategies.

**2.2.1 Calculation as total electronic energy difference.** The first approach is to compute total electronic energy for both the neutral and oxidized species at the optimized geometry of neutral molecule in solvent. The total energy difference

$$\Delta E^{\text{ox}} = E_{\text{sol}}^+(g_{\text{sol}}^0) - E_{\text{sol}}^0(g_{\text{sol}}^0) \quad (6)$$

is then taken as approximation to the GSOP.

**2.2.2 Calculation as KS-DFT HOMO energy.** The simplest approach would be to identify the negative of the HOMO energy,  $-\varepsilon_{\text{HOMO}}$ , from a KS-DFT calculation with the vertical ionisation energy in exact KS-DFT. While for any finite system  $-\varepsilon_{\text{HOMO}}$  exactly equals the ionisation energy in exact KS-DFT,<sup>58</sup> errors obtained from common approximations to the exact





exchange–correlation potential  $v_{xc}$  can be of the order of several eV.<sup>29,30,59</sup> This failure is a direct consequence of the incorrect long-range behaviour of the  $v_{xc}(r)$  (exponentially instead of  $1/|r|$ ) in the local density approximation (LDA) and generalized gradient approximations (GGA).<sup>28</sup> On the other hand, the exact exchange (eex) potential shows the correct long-range behaviour<sup>60</sup> and hybrid functionals, combining GGAs with a fraction of eex give typically much better ionisation energies. Despite these obvious shortcomings, the KS-DFT approach is computationally very efficient since it only requires to perform a single geometry optimization in solution.

**2.2.3 Calculation as GW HOMO energy.** In a single-particle energy framework, one can also go beyond DFT and calculate a vertical approximation to GSOP using MBPT. Central to MBPT is Dyson's equation<sup>61</sup>

$$\sum_j \Sigma_{ij}(\omega_i) \phi_j(\omega_i) = [\omega_i - \varepsilon_i] \phi_i(\omega_i), \quad (7)$$

where  $\varepsilon_i$  is a KS eigenvalue,  $\phi_i$  denotes a molecular orbital,  $\omega_i$  is the exact one-electron energy and  $\Sigma$  is the so-called self-energy. In practice, the GWA to the self-energy<sup>31</sup> is often used. In the GWA,  $\Sigma$  is calculated as the convolution (in frequency space) of the interacting Green's function  $G$  and the dynamically screened interaction  $W$ , which is related to the unscreened Coulomb interaction  $v_c$  by

$$W_0(\omega) = [\varepsilon_{\text{RPA}}(\omega)]^{-1} v_c, \quad \varepsilon_{\text{RPA}} = 1 + v_c \pi_0(\omega), \quad (8)$$

where  $\pi_0$  is the independent particle polarizability in the RPA. Eqn (7) is usually simplified by evaluating  $\Sigma$  with a non-interacting Green's function,  $G_0$ , instead of the interacting one. In the  $G_0W_0$  approximation<sup>62</sup>  $\Sigma$  is additionally approximated as diagonal, so that the eqn (7) reduces to a set of independent non-linear equations,

$$\Sigma_{ii}(\omega_i) = \omega_i - \varepsilon_i, \quad (9)$$

with  $\Sigma_{ii} = \phi_i | \Sigma(\omega) | \phi_i$ . Thus, the QP energies are obtained as a perturbative correction to the KS eigenvalues. In eigenvalue-only self-consistent GW (evGW), the QP energies are additionally updated until self-consistency is reached.<sup>62</sup> Yet another option is to map  $\Sigma$  to a static, non-local and Hermitian

exchange–correlation potential,<sup>63,64</sup> which then defines a non-linear eigenproblem, much like in KS-DFT, with the only difference that the potential is a functional of  $G_0$ . This approach is referred to as QP self-consistent GW (qsGW). The GW approximation can be implemented with almost quadratic scaling using localized atomic orbitals,<sup>65–67</sup> which makes at least  $G_0W_0$  competitive with hybrid functional calculations in terms of computational cost in the same basis set.<sup>67</sup> Self-consistent variants are more expensive, but the calculations typically converge in 5–10 iterations.<sup>68</sup> However, GW calculations converge much slower to the complete basis set (CBS) limit.<sup>33,69,70</sup> For accurate QP energies, it is usually necessary to perform calculations using correlation consistent basis sets of at least triple- and quadruple- $\zeta$  quality and to extrapolate to the CBS limit.<sup>69–71</sup>

### 2.3 Experimental data set

To validate the methods discussed above we compared the calculated GSOP values with experimental values for  $E_{1/2}$ . The set used for validation is a subset of all considered dye cores (Fig. 3) for which we only selected those GSOP measurements that were done under similar experimental conditions. The substituted naphthalene-diiimides (NDI),<sup>72–75</sup> perylene-diiimide (PDI)<sup>76</sup> core and substituted PDI cores<sup>76,77</sup> with different number and different type of substituents are used for validation of theoretical predictions.

Each substituent is given a single character for identification to shorten the notation for the complete dye in which the substituents are attached to the aromatic core, as shown in Fig. 4. The substituents containing electron donating atoms and groups are numbered from 1 to 7. The substituents 8, 9 and 0 are introduced only to enlarge the experimental data set available for validation and they are not used in the final screening procedure. As shown in Fig. 4, side chains of the substituents and cores were sometimes shortened for computational convenience. We thereby checked that the consequence of such replacements on the computed GSOPs are minor, and moreover, since this is done systematically for all the dyes this should give at most a consistent shift in oxidation potential value (ESI,† S1). For comparison of electrochemical

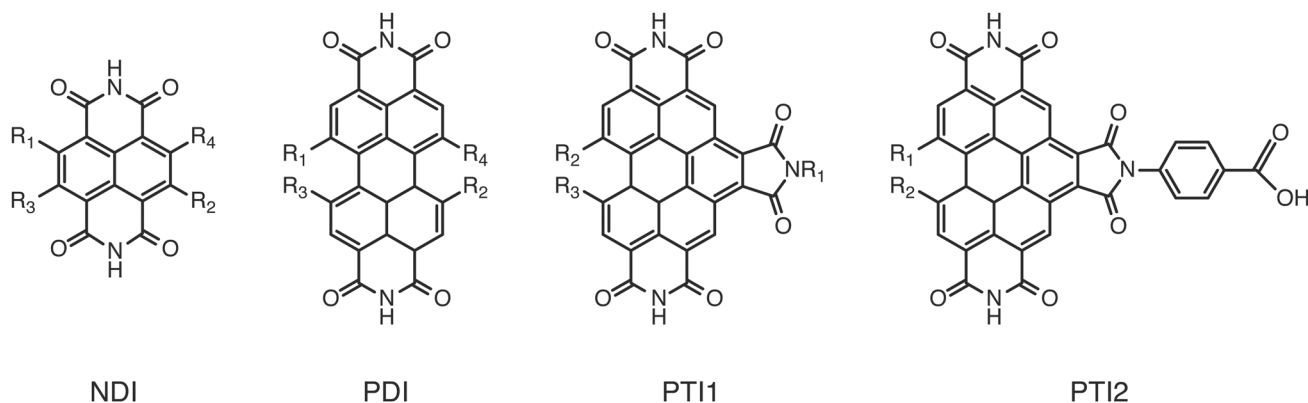


Fig. 3 Structures of the unsubstituted cores considered for screening.



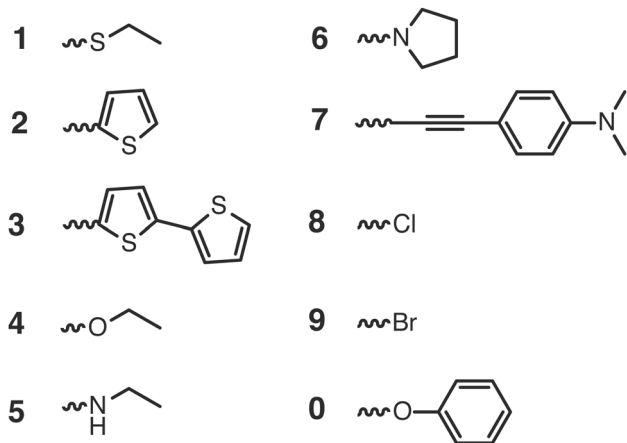


Fig. 4 Structures of the substituents attached to the cores.

properties from different experiments or for comparison of computational results with experiment we need to convert all values to a common scale, the absolute scale (against vacuum). Choosing Fermi–Dirac statistics as statistical mechanical formalism for the electron we set the absolute potentials for  $F_c/F_c^+$  couple as 4.98 V and for SCE as 4.52 V (ESI,† S1).<sup>43</sup> The complete set of dyes with experimental oxidation potentials can be found in the ESI,† (Table S1).

To quantify the predicting ability of the methods we will use mean deviations and squared correlation –  $R^2$ . The magnitude of mean deviations mainly depends on the choice of the formalism to relate the reference electrode scale to the absolute scale. This choice might yield higher or lower mean deviations, but the shift will be consistent throughout the data set. For this reason we will rely on the  $R^2$  – strength of the linear correlation – for the evaluation of the performance. Having in mind possible errors from experimental and theoretical methods we monitored how sensitive  $R^2$  is to small variations in the reference values. Considering the errors that might occur as a consequence of side chain adjustments we tested how much  $R^2$  deviates from its initial value if we introduce a random error in the (+0.05, –0.05) eV interval. We found that the mean absolute deviation of the  $R^2$  converges to 0.01 (ESI,† S2). In conclusion, the methods for which the correlation with the experiment differs in  $\pm 0.01$  would be considered to perform equally well.

### 3 Computational details

All computations were performed with the Amsterdam Modeling Suite (AMS) from Software for Chemistry and Materials (SCM),<sup>78</sup> versions 2019.302 and 2020.1, except for the  $GW$  calculations which were performed with a modified development version of 2020.1. Initial geometries for the core and the library of substituents were prepared with the SCM Graphical User Interface (GUI). The long alkyl chains were replaced with ethyl groups. Generation of NDI and PDI derivative structures was done with the Compound Attachment Tool (CAT).<sup>79</sup> CAT is a Python code employing, among others, the Python Library for Automating

Molecular Simulation (PLAMS)<sup>80</sup> for generating structures. The initial molecular structures were pre-optimized with the DFTB3<sup>81</sup> method using the 3-ob parameter set.<sup>82–85</sup> Convergence to a stable minimum of the nuclear potential energy surface was checked by calculating only the lowest frequency normal modes.<sup>86</sup> We allowed for some numerical noise and considered structures with imaginary frequencies above  $-20\text{ cm}^{-1}$  for further characterisation. The workflow for the strategies above in general consists of geometry optimization (GO) in some cases followed by frequencies (normal modes calculation) and single point (SP) calculation(s). SP calculations were done with DFT or GW. Depending on the strategy, solvation effects were included in the geometry optimization calculation or in the following single point calculation (ESI,† S3). The DFT calculations (GO and SP) were done using the B3LYP functional,<sup>87</sup> with the TZ2P basis set.<sup>88</sup>

The quality of density fitting and numerical integration were set to *Good*, except for the part with Hartree Fock exchange where quality was set to *Normal*. The SQM technique for GO in the screening workflow was GFN1-xTB.<sup>89</sup> For the dyes that showed too large imaginary frequencies the GO calculation was repeated after tightening the convergence criteria for energy and nuclear gradients by a factor of 10 relative to the default values. The frequency calculations for hybrids were done numerically; their outputs contain all the thermodynamic properties, electronic bonding energy and nuclear kinetic energies, at room temperature. If present, negative frequencies were re-scanned to assure that they were not spurious (tolerating frequencies above  $-20\text{ cm}^{-1}$ ).

Solvation effects were incorporated with the COSMO<sup>90</sup> or COSMO-RS<sup>91</sup> models, at a pressure of 1 atm and temperature of 298.15 K. All the calculations including the COSMO model were performed in dichloromethane, with a dielectric constant of 8.9. The default Amsterdam density functional (ADF) 2019.302<sup>92</sup> settings were used: the cavity construction is Delley type,<sup>93</sup> atomic radii are the corresponding van der Waals radii from the MM3 method by Allinger<sup>94</sup> (with widely accepted increase of 120%<sup>54</sup>). COSMO-RS requires an input from a COSMO calculation that defines the reference state which is the molecule in a perfect conductor, with infinite dielectric constant. In the reference COSMO calculation, the cavity construction is of Delley type and atomic radii are the Klamt atomic radii.<sup>54</sup> The subsequent statistical thermodynamic calculation describes the effect of dichloromethane as a solvent.

The  $GW$  calculations were based on the solvated geometries optimized with B3LYP/TZ2P. We considered different variants of non self-consistent  $GW$  ( $G_0W_0$ ) and partially self-consistent  $GW$ :  $G_0W_0@PBE0(HF = 0.4)$  ( $G_0W_0$  performed on top of a PBE0<sup>95,96</sup> calculations with 40% exact exchange),  $G_0W_0@LRC-\omega PBEH$ ,<sup>97</sup> eigenvalue-only self-consistent  $GW$  (ev $GW$ ) based on the same two functionals (ev $GW@PBE0(HF = 0.4)$  and ev $GW@LRC-\omega PBEH$ ), QP self-consistent  $GW$  (qs  $GW$ ) as well as QP self-consistent  $GW$  with self-consistency only in  $G$ , but not in  $W$  based on a PBE starting point (qs $GW_0@PBE$ ). All  $GW$  calculations were performed using the TZ2P and TZ3P<sup>71</sup> basis sets. For  $G_0W_0@PBE(HF = 0.4)$  we additionally used the QZ6P



**Table 1** The strategies considered in this publication together with required calculations, different methods, and corresponding computation times. Method performance was accessed for DFT<sup>c</sup> with B3LYP functional in combination with other methods. Timings denote elapsed times measured in CPU hours and are for unsubstituted PDI with 40 atoms

Approach	Calculations <sup>a</sup>	Method	Elapsed (step) <sup>b</sup>	Total (core h) <sup>b</sup>
$\Delta G_{\text{COSMO}}^{\text{DC}}$	GO(n,s) + f	DFT	390	1066
	GO(o,s) + f		676	
$\Delta G_{\text{COSMO}}^{\text{TC}}$	GO(n,v) + f	DFT	381	1109
	+SP(n,s)		1.4	
	GO(n,s)		9.6	
	+SP(n,v)		1.3	
	GO(o,v) + f		692	
	+SP(o,s)		2.5	
	+SP(n,v)		1.3	
	GO(o,s)		18.3	
	+SP(n,v)		1.3	
$\Delta G_{\text{COSMO-RS}}^{\text{TC}}$	GO(n,v)	DFT	9.6	32.7
	+SP(n,s)		1.4	
	GO(o,v)		19.2	
	+SP(o,s)		2.5	
$\Delta E_{\text{ox}}^{\text{DFT}}$	GO(n,s)	DFT	9.6	12.1
	+SP(o,s)		2.5	
$\epsilon_{\text{HOMO}}^{\text{DFT}}$	GO(n,s)	DFT	9.6	9.6
	+SP(o,s)		2.5	
$\epsilon_{\text{HOMO}}^{\text{GW}}$	GO(n,s)	DFT	9.6	47.7
	+SP(o,s)	DFT	2.5	
$\Delta G_{\text{COSMO-RS}}^{\text{screening}}$	GO(n,v)	$G_0W_0^d$	29 + 6.6	4.1
	+SP(n,s)	DFT	1.4	
	GO(o,v)	GFN-xTB	0.1	
	+SP(o,s)	DFT	2.5	

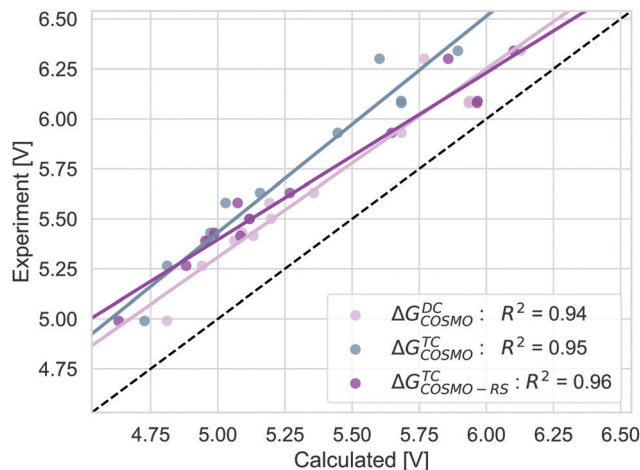
<sup>a</sup> GO denotes geometry optimization of the neutral (n) or oxidized (o) molecule in vacuum (v) or in solution (s); f denotes frequencies; SP denotes single point calculation using the preceding geometry. All geometry calculations are started from the same initial geometry. <sup>b</sup> All calculations have been performed on a single 2.2 GHz intel Xeon (E5-2650 v4) node (broadwell architecture) with 24 cores and 128 GB RAM. <sup>c</sup> The TZ2P basis set and B3LYP-D3(BJ) functional has been used for all DFT calculations. *Good* numerical quality has been used throughout except for the Hartree-Fock part, where the *Normal* fit set has been used. <sup>d</sup> (QZ6P + TZ3P).

basis set<sup>71</sup> and extrapolated the IP to the CBS limit using

$$\epsilon_n^{\text{CBS}} = \epsilon_n^{\text{QZ}} - \frac{1}{N_{\text{bas}}^{\text{QZ}}} \frac{\epsilon_n^{\text{QZ}} - \epsilon_n^{\text{TZ}}}{\frac{1}{N_{\text{bas}}^{\text{QZ}}} - \frac{1}{N_{\text{bas}}^{\text{TZ}}}}, \quad (10)$$

where  $\epsilon_n^{\text{QZ}}$  ( $\epsilon_n^{\text{TZ}}$ ) denotes the value of the QP energy using QZ6P (TZ3P) and  $N_{\text{bas}}^{\text{QZ}}$  and  $N_{\text{bas}}^{\text{TZ}}$  denote the respective numbers of basis functions (in spherical harmonics so that there are 5d and 7f functions).<sup>71</sup>

The *GW* implementation in ADF follows the space-time method and is outlined in ref. 67 and 71 as well as in ref. 68 for  $qsGW_0$  and  $qsGW$  and the numerical settings chosen for all the *GW* calculations in this work follow the recommendations therein: We used 20 grid points in imaginary time and imaginary frequency each. For the  $G_0W_0$  and  $evGW$  we set NumericalQuality *Good* and for the  $qsGW_0$  calculations we used *veryGood* quality combined with the *Good* fit sets and thresholds for the HF part of



**Fig. 5** The correlation of adiabatic GSOP computed with  $\Delta G_{\text{COSMO}}^{\text{TC}}$  (pink),  $\Delta G_{\text{COSMO}}^{\text{DC}}$  (grey) and  $\Delta G_{\text{COSMO-RS}}^{\text{TC}}$  (purple) methods to the experimental oxidation potential vs. vacuum (dashed line).

the *gKS* and for the *GW* part of the calculations. For reasons outlined in ref. 67 we set *Dependency Bas* =  $5 \times 10^{-4}$  in the AMS input for  $G_0W_0$  and  $evGW$  calculations and *Dependency Bas* =  $5 \times 10^{-3}$  for  $qsGW_0$ .

## 4 Results

### 4.1 Adiabatic GSOP

We compare the strategies for obtaining the Gibbs free energy of the oxidation reaction with two pathways: TC and DC with COSMO. For the TC strategy we additionally consider the COSMO-RS method. The performance of these methods is determined for 13 out of 14 molecules from the initial data set. We excluded the PDI-0000 molecule as it turned out to heavily influence the correlation with the experimental reference data and thus presents a clear outlier. The unexpectedly large errors seen for this molecule may be caused by uncertainties in selecting its most appropriate conformation<sup>98</sup> or by its sensitivity to CS model parameters (ESI,† S4). Both aspects make this molecule unsuited for getting accurate results with the automated procedure for initial structure generation and subsequent calculations that we currently employ. Table 1 shows that calculation timings are similar for the  $\Delta G_{\text{COSMO}}^{\text{TC}}$  and  $\Delta G_{\text{COSMO}}^{\text{DC}}$  method. Both methods perform equally well for the given set of experimental results, Fig. 5. The TC method has a slightly higher squared correlation coefficient ( $R^2 = 0.95$ ) than the DC methods ( $R^2 = 0.94$ ). The values from the TC method have an almost consistent shift from the corresponding experimental values, with a mean absolute deviation of 0.45 eV (RMSD = 0.46 eV). The DC method is a bit closer to the absolute experimental values with a MAD of 0.28 eV (RMSD = 0.30 eV). The  $\Delta G_{\text{COSMO-RS}}^{\text{TC}}$  absolute deviation of 0.34 eV (RMSD = 0.36 eV) is as large as for other two methods, while the squared correlation is higher than both methods employing only COSMO,  $R^2 = 0.96$ . The  $\Delta G_{\text{COSMO-RS}}^{\text{screening}}$ , where the geometry optimisation is performed using the SQM method (GFN1-xTB),



performs as well as  $\Delta G_{\text{COSMO-RS}}^{\text{TC}}$ . The correlation between these two methods is almost 1 and for both methods the correlation with experiment is 0.96.

As the GSOP represents the difference between the two different oxidation states of the same molecule, some contributions will cancel out. Here we elaborate on the effective contributions to the GSOP originating from solvation, geometry relaxation due to oxidation and due to solvation and thermal contribution. The energy contribution due to geometry relaxation related to the solvation in the TC method is added separately from electronic solvation effects, as shown in eqn (3).

We found that this energy difference for both, neutral or oxidized molecules is negligible, being on average 0.02 eV and independent from type or number of ligands (ESI,† S5). Therefore, for this type of molecules solvated in dichloromethane the energy contribution coming from the change in geometry due to the solvation is negligible. In the TC method, this means that evaluation of these contributions can be omitted to reduce the total computational cost of the  $\Delta G_{\text{COSMO}}^{\text{TC}}$  strategy.

On the other hand, the electronic solvent contributions are crucial to the GSOP evaluation. The effective contribution to the GSOP is the difference between the solvation effects on neutral and oxidized molecules, eqn (11). For the TC method these values are calculated in the gas-phase. In the DC method this solvent contribution is contained in the  $E_{\text{sol}}^i(g_{\text{sol}}^i)$  energy term in eqn (1). On average the value of  $\Delta\Delta G_{\text{soliv}}$  is  $-1.61$  eV, with a maximum value of  $-2.00$  eV for NDI-58. This contribution does not depend largely on the method, TC or DC (ESI,† S5)

$$\Delta\Delta G_{\text{soliv}} = \Delta G_{\text{soliv}}^+ - \Delta G_{\text{soliv}}^0 \quad (11)$$

We also evaluated the effective contribution of the thermal effects on the GSOP. Depending on the method, the thermal contribution is evaluated in solution or gas-phase for the DC or TC method respectively, for the temperature  $T = 298.15$  K.

$$\Delta G_{\text{therm}}^{\text{DC(TC)}}(T) = G_{\text{therm}}^+(g_{\text{sol/gas}}^+, T) - G_{\text{therm}}^0(g_{\text{sol/gas}}^0, T) \quad (12)$$

The effective  $\Delta G_{\text{therm}}^{\text{DC(TC)}}(T)$  contribution for DC and TC, has an average of 0.04 eV and 0.03 eV respectively (for the absolute value difference). So, frequencies calculated in the same environment are very close in value for neutral and oxidized dyes and their contribution cancel out. Considering the cost of frequency calculations listed in Table 1, for GO(n,s) + f and just GO(n,s), the frequencies are 97% of GO(n,s) + f time. We therefore conclude that by neglecting this small thermal contribution, workflows for this type of dyes can be sped up significantly.

Evaluation of the contribution to GSOP due to relaxation in the oxidized state, the main difference between the adiabatic and vertical approach, will follow in the next section.

## 4.2 Vertical approximations to GSOP

The alternative to the full adiabatic GSOP is to approximate it by a vertical oxidation potential. Apart from experimental considerations, such an approximation can always be justified

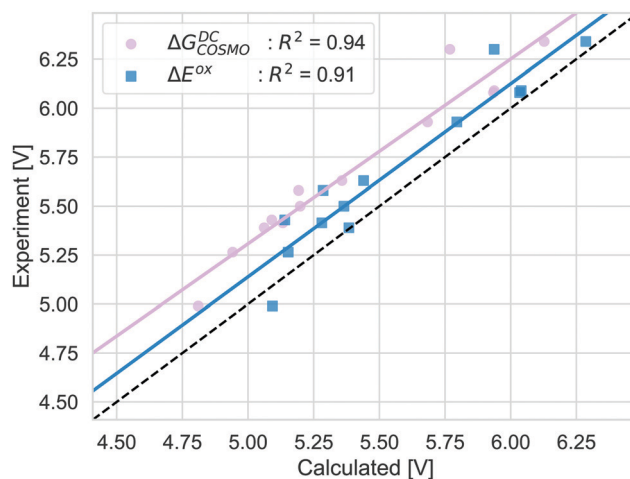


Fig. 6 The GSOP computed as Gibbs free energy of the oxidation using the DC pathway (pink) and the vertical approach employing total electronic energy difference (blue) compared to the experimental oxidation potential (dashed line) vs. vacuum.

if the effect of the geometry relaxation as well as other thermal contributions are negligible. The validity of this approximation can be assessed directly by comparing the Gibbs free energy for the oxidation reaction obtained with DC,  $\Delta G_{\text{COSMO}}^{\text{DC}}$  to the electronic energy difference  $\Delta E^{\text{ox}}$ : the calculations are performed in the same manner and we can focus on the physical approximations. The comparison is shown in Fig. 6.

$\Delta E^{\text{ox}}$  does include solvation effects (relaxation due to solvation as well) but it neglects the thermal contribution and the relaxation from the neutral to the oxidized geometry. The  $\Delta E_{\text{ox}}$  data are close to the absolute experimental values with the MAD being (0.05 (RMSD = 0.18) eV). For  $\Delta E^{\text{ox}}$  with  $R^2 = 0.91$ , the strength of the linear correlation is slightly weaker than for the adiabatic approach. Comparing the differences between the two approaches, as discussed above, the thermal contributions (ESI,† S5) are rather small and consistent and should not affect the correlation strongly. In contrast, the difference coming from the relaxation due to oxidation is less consistent. On average, the energy difference due to using the oxidized or neutral geometry, both optimized in solution (gas) is 0.12 eV (0.15 eV) (ESI,† S5). The lower correlation for  $\Delta E_{\text{ox}}$  is caused by a few dyes that show a larger change in energy between two oxidation states, especially NDI-5555 and NDI-555 for which the contributions due to this relaxation are 0.36 eV and 0.25 eV, respectively (ESI,† S5). However, the effect of the geometry relaxation after electron removal appears to be crucial.

Different strategies to calculate the vertical oxidation potential of the solvated dye are shown in Fig. 7.  $\Delta E^{\text{ox}}$  (blue) and  $-e_{\text{HOMO}}^{\text{DFT}}$  give very similar results and also  $-e_{\text{HOMO}}^{\text{GW,solv}}$  gives a similar squared correlation coefficient. Compared to the other approaches,  $-e_{\text{HOMO}}^{\text{GW,solv}}$  overestimates the GSOP considerably, while  $\Delta E_{\text{ox}}^{\text{DFT}}$  (blue) and  $-e_{\text{HOMO}}^{\text{DFT}}$  underestimate it. This depends of course on the choice of the formalism to relate the reference electrode scale to the absolute scale. However, it is known that for the gas phase, partially self-consistent GW as well as  $G_0W_0$





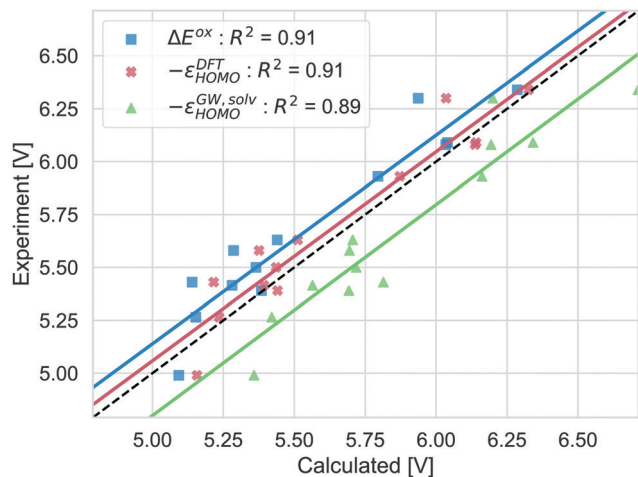


Fig. 7 Computed vertical GSOPs with  $\Delta E^{\text{ox}}$  (blue),  $-\epsilon_{\text{HOMO}}^{\text{DFT}}$  (red) and  $-\epsilon_{\text{HOMO}}^{\text{GW,solv}}$  (green) compared to the experimental oxidation potential (dashed line) vs. vacuum.

based on a functional with a high percentage of Hartree-Fock exchange often leads to overestimated IPs for organic molecules compared to experiment.<sup>99</sup> On the other hand, using  $-\epsilon_{\text{HOMO}}^{\text{DFT}}$  from B3LYP, due to lower percentage of Hartree-Fock exchange should rather lead to underestimation of the oxidation potential.

The correlation with experiment becomes considerably better when the missing thermodynamic contributions are considered. Adding the energy lowering due to geometry relaxation to  $-\epsilon_{\text{HOMO}}^{\text{GW,solv}}$  significantly increases correlation with experiment, from  $R^2 = 0.89$  to  $R^2 = 0.93$  and for  $-\epsilon_{\text{HOMO}}^{\text{DFT}}$  and  $\Delta E^{\text{ox}}$ , the squared correlation coefficients increase from  $R^2 = 0.91$  to  $R^2 = 0.96$ . These numbers should always be interpreted with the error bars in the experimental data in mind: differences in  $R^2$  of 0.01 (as for the difference between  $-\epsilon_{\text{HOMO}}^{\text{DFT}}$  and  $-\epsilon_{\text{HOMO}}^{\text{GW,solv}}$ ) are negligible, but the effect of structural relaxation is significant.

In summary, the data presented in this section suggests that the vertical approximation to GSOP is not reliable. It is irrelevant what particular method is used to calculate the vertical GSOP of the solvated dye. Reliable correlation with experiment is only achieved when the relaxation of the geometry after oxidation is accounted for. This is due to the fact that this contribution can be quite large for certain systems, but almost negligible for others, as illustrated for the example of NDI-5555. However, investigating the effect of thermal contributions, we found that they are very small and that their influence is negligible. In practice, this is very important since the frequency calculations required to calculate these contributions are computationally very demanding.

### 4.3 Timings

Since we want to find a method which is suitable for large-scale screening for potentially thousands of compounds, the CPU times of the different approaches are an important consideration. We use the unsubstituted PDI with 40 atoms as a representative example. Some of the systems considered here

Table 2 Statistical analysis<sup>a</sup> of the considered strategies compared with cyclic voltammetry measurements in dichloromethane

Approach	MD	MAD	RMSD	$R^2$
$\Delta G_{\text{COSMO}}^{\text{DC}}$	-0.28	0.28	0.30	0.94
$\Delta G_{\text{COSMO}}^{\text{TC}}$	-0.45	0.45	0.46	0.95
$\Delta G_{\text{COSMO-RS}}^{\text{TC}}$	-0.34	0.34	0.36	0.96
$\Delta E^{\text{ox}}$	-0.13	0.15	0.18	0.91
$-\epsilon_{\text{HOMO}}^{\text{DFT}}$	-0.05	0.1	0.13	0.91
$-\epsilon_{\text{HOMO}}^{\text{GW,solv}}$	0.20	0.22	0.24	0.89
$-\epsilon_{\text{HOMO}}^{\text{GW,solv,geo}}$	0.03	0.08	0.12	0.93
$\Delta G_{\text{COSMO-RS}}^{\text{screening}}$	-0.34	0.34	0.37	0.96

<sup>a</sup> MD stands for the mean deviation; MAD stands for the mean absolute deviation, RMSD stands for the root mean squared deviation;  $R^2$  is squared correlation.

are about twice as large, but our conclusions do not change for these systems. All calculations presented in this subsection have been performed on a single 2.2 GHz intel Xeon (E5-2650 v4) node (broadwell architecture) with 24 cores and 128 GB RAM. Timings are shown in Table 1.

As already alluded to,  $\Delta G_{\text{COSMO}}^{\text{DC}}$  and  $\Delta G_{\text{COSMO}}^{\text{TC}}$  are by far the computational most demanding strategies, which is due to the expensive frequency calculations. Despite their good accuracy, they are clearly not suitable for large-scale compound screening. Calculation of the  $G_0W_0$  QP energies is much cheaper but still more demanding than all other DFT-based strategies without frequency calculations. However, most of the time is spent on the QZ6P calculation. As explained in appendix 7, the GW calculations can also be performed with a TZ3P calculation only, without impairing the accuracy significantly. With the TZ3P calculation only,  $-\epsilon_{\text{HOMO}}^{\text{GW}}$  becomes competitive with  $\Delta E_{\text{ox}}^{\text{DFT}}$  and  $\Delta G_{\text{COSMO-RS}}^{\text{TC}}$  and all of these strategies could be considered for large scale-screening as well. Finally,  $\Delta G_{\text{COSMO-RS}}^{\text{screening}}$  is the computationally cheapest approach which is due to the fact that the geometry optimizations are performed on the GFN-xTB level. Since it provides an accuracy comparable to the involved  $\Delta G_{\text{COSMO}}^{\text{DC}}$  and  $\Delta G_{\text{COSMO}}^{\text{TC}}$  strategies, it is our method of choice for the screening of the dyes, described in the next subsection (Table 2).

### 4.4 Screening

We have employed the  $\Delta G_{\text{COSMO-RS}}^{\text{screening}}$  method to calculate the GSOP potentials of large number of dyes. We have applied it to set of dyes that we determined previously<sup>17</sup> to have desirable optical properties. Among almost 2500 NDIs, PDIs, PTI1s and PTI2s derivatives around 1400 dyes fulfilled the criteria to be used as a photosensitizers in DS-PECs. All these dyes have an intense, lowest transition in the desired energy range, between the 1.35 eV which is the minimal thermodynamic requirement for water oxidation and 3.20 eV which is roughly the boundary between the visible and the UV energy range. During the water oxidation cycle the WOC goes through different oxidation states. For the GSOP of the dye the most important criteria is that it should lie higher than the WOC's highest oxidation potential (Fig. 1) to assure a favourable potential gradient for electron transfer. We thereby assume 0.1 V potential difference to be sufficient. As a trial WOC we will take the Ru-based





**Table 3** GOSP [V vs. vacuum], lowest most intense excitation  $\lambda_{\max}$  [eV] and oscillation strength (osc.str) values for 10 suitable dyes with the highest oscillation strengths<sup>17</sup>

Name	GSOP	$\lambda_{\max}$	osc.str.
PTI1-5	6.00	2.90	0.66
PTI1-144	5.88	2.67	0.58
PTI1-444	5.81	2.67	0.56
PTI1-544	5.91	2.67	0.55
PTI2-4	6.03	2.70	0.55
PTI2-44	5.82	2.64	0.54
PTI1-54	6.02	2.72	0.53
NDI-125	5.78	2.57	0.45
NDI-225	5.84	2.49	0.44
PTI1-15	5.73	2.49	0.44

experimental ones, is to shift the criteria for the value of MD. The dotted line, (parallel to the purple horizontal line) is the limit for GSOP including the underestimations of experimental oxidation potentials, the MD of about 0.3 V. The dashed line (parallel to the blue diagonal line) takes into account also the overestimation of the experimental absorption properties, which has a MD of about 0.4 V. The dyes in between the purple line and dashed line therefore represent the set of dyes that are suitable to this particular example system, taking into account errors in the computations. This reduces the set of suitable dyes to 66 dyes. From this set we show the dyes that have the highest oscillator strength in Table 3. The full list of these dyes is also indicated in the ESI,† S6.

## 5 Conclusions

We have computed the solution-phase Gibbs free energy, with adiabatic and vertical approaches, employing different methods. First we have built the experimental data set from the CV measurements for the derivatives of the NDI and PDI dyes to validate the computational results. Calculations of solution-phase Gibbs free energy are done at the DFT level using different CSM pathways, the DC and TC pathways, and using different CSM models – COSMO and COSMO-RS. We find that, to calculate the ground state oxidation potential for these dyes, both pathways using the COSMO model perform well. The TC path shows a slightly higher value of squared correlation with the experiment, but a higher MAD value as well. The TC pathway with the COSMO-RS model has the highest correlation with the experiment with a MAD lower than the TC pathway with COSMO. Comparison to different vertical approximation,  $\Delta E^{\text{ox}}$  and one electron energy strategies, the  $-\epsilon_{\text{HOMO}}^{\text{DFT}}$  and  $-\epsilon_{\text{HOMO}}^{\text{GW}}$ , shows the importance of taking into account geometry relaxation after oxidation, as well as the inclusion of electronic solvent effects. However, other thermal effects do not play an important role when it comes to this set of molecules. Therefore, the adiabatic approach appears more suitable for screening purposes, but one should be aware that this procedure might be unfavourable for the molecules with large number of conformational isomers that exist in a small energy range and that are highly affected by solvent, such as PDI-0000. Replacing DFT geometry optimization by SQM optimization has not

significantly affected the correlation with experiment for the adiabatic approach combined with the COSMO-RS model, while it reduced the cost of the strategy by a factor of eight. Therefore, the  $\Delta G_{\text{COSMO-RS}}^{\text{screening}}$  is found to be a suitable strategy for screening on a desired GSOP range for derivatives of the NDI and PDI dyes.

The dyes that have been proposed as suitable for panchromatic sensitization of the photoelectrode in DS-PECs have been further characterised for their redox properties, in particular GSOP. Using the  $\Delta G_{\text{COSMO-RS}}^{\text{screening}}$  strategy, the GSOP is evaluated for 1340 dyes. For the system of choice, which is a  $\text{TiO}_2$  based photoelectrode and Ru-based catalyst<sup>100</sup> as a WOC, there are 118 dyes that fulfil the given criteria.

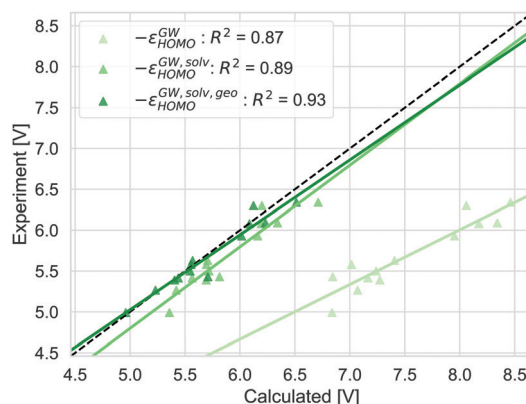
## Conflicts of interest

There are no conflicts to declare.

## Appendix

### A Effect of solvent and structural relaxation on GW

The output of the GW calculations are QP energies. They correspond to the vertical electron addition and removal energies in vacuum and can not be compared to the GSOP directly. This is shown here for  $G_0W_0@PBE0(\text{HF} = 0.4)$  in Fig. 10, the same is observed for all other tested GW approaches. Identifying  $-\epsilon_{\text{HOMO}}^{\text{GW}}$  systematically overestimates the GSOP by several eV on average. Also the squared correlation is with only  $R^2 = 0.87$  worse than for  $\Delta G$ . Accounting for solvent effects removes most of the overestimation and improves  $R^2$  to 0.89. Taking into account the structural relaxation after electron removal improves the correlation to  $R^2 = 0.93$ , which is now competitive with  $\Delta G$ . Note, that this does not include other thermal contributions. As discussed above, they are very small and can be



**Fig. 10** The GSOP calculated with the GW approaches successively including more physical effects (solvation effects and geometry relaxation due to oxidation):  $-\epsilon_{\text{HOMO}}^{\text{GW}}$ ,  $-\epsilon_{\text{HOMO}}^{\text{GW,solv}}$  and  $-\epsilon_{\text{HOMO}}^{\text{GW,solv,geo}}$  (lightest to darkest shade of green) compared to the experimental oxidation potential (dashed line) vs. vacuum.





**Table 4** The squared correlation coefficients  $R^2$  compared to experiment for different  $GW$  approaches for different basis sets. Solvent corrections and the effects of geometry relaxation are included

Method	TZ2P	TZ3P
$G_0W_0@PBE0(HF = 0.4)$	0.91	0.92
$G_0W_0@LRC \omega PBEH$	0.92	0.91
$evGW@PBE0(HF = 0.4)$	0.87	0.92
$evGW@LRC\omega PBEH$	0.89	0.90
$qsGW_0@PBE$	0.88	0.85
$qsGW$	0.86	0.91

neglected. Our results are consistent with earlier work by Umari *et al.*<sup>37</sup> who compared  $G_0W_0@BLYP$  vertical oxidation potentials to experimental data from differential pulse voltammetry in acetonitrile.<sup>101</sup> They observed that their  $GW$  approach overestimated the experimental values by nearly one eV on average, even though it is known, that  $G_0W_0$  based on a GGA starting point systematically underestimates the gas phase oxidation potential of organic molecules.<sup>99</sup>

## B Comparison and validation of $GW$ methods

Taking into account these considerations, we can compare the different  $GW$  methods to experiment. We tested all approaches using the TZ2P and TZ3P basis sets and calculated the  $R^2$  values with respect to experiment. The result of our comparison is shown in Table 4. The correlation of  $G_0W_0$  was found to be more or less independent of the starting point. Eigenvalues-only self-consistency does not improve the results compared to  $G_0W_0$  for TZ3P and worsens them for TZ2P. The same is observed for  $qsGW$ .  $qsGW_0$  gives much worse correlation coefficient on the TZ3P level ( $R^2 = 0.85$ ) than the other methods. For this reason, using one of the more expensive partially self-consistent approaches is not justified and we used the  $G_0W_0PBE0(HF = 0.4)$  method for comparison to the other computational methods in the main body of the text. Our conclusion that self-consistency is not needed here is of course not valid in general and in other cases it might be possible that updates in the wavefunction are needed. Reordering of orbitals during the  $GW$  calculation can be an indicator for this and in such cases it might be necessary to use some kind of self-consistency in  $GW$ .<sup>36</sup>

For optimal agreement with experiment, the basis set error needs to be removed from the QP energies as well, using eqn (10). However, for the molecules considered here, basis set limit extrapolation increases the correlation of  $G_0W_0PBE0(HF = 0.4)$  with experiment only marginally from  $R^2 = 0.92$  to  $R^2 = 0.93$ . This is due to the fact that CBS limit extrapolation results in a more or less constant shift of the QP energies, which then leaves the correlation coefficient mainly unaffected.

## Acknowledgements

This research has been financially supported by the NWO Solar to Products program (project number 733.000.007). AF

acknowledges funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the Innovation Fund for Chemistry and from the Ministry of Economic Affairs in the framework of the *TKI/PPS-Toeslagregeling*. We acknowledge the use of supercomputer facilities at SURFsara sponsored by NWO Physical Sciences, with financial support from The Netherlands Organization for Scientific Research (NWO).

## References

- 1 The Future of Hydrogen: Seizing today's opportunities, IEA Technical Report June, 2019.
- 2 Hydrogen Roadmap Europe: A sustainable pathway for the European Energy Transition, Fuel cells and hydrogen joint undertaking technical report, 2019.
- 3 K. Fujishima and A. Honda, *Nature*, 1972, **238**, 37–38.
- 4 C. Acar and I. Dincer, *Chem. Eng. Sci.*, 2019, **197**, 74–86.
- 5 C. Decavoli, C. L. Boldrini, N. Manfredi and A. Abboto, *Rend. Lincei*, 2019, **30**, 469–483.
- 6 W. Song, Z. Chen, M. Kyle Brennaman, J. J. Concepcion, A. O. T. Patrocinio, N. Y. Murakami Iha and T. J. Meyer, *Pure Appl. Chem.*, 2011, **83**, 749–768.
- 7 C. Decavoli, C. L. Boldrini, N. Manfredi and A. Abboto, *Eur. J. Inorg. Chem.*, 2020, 978–999.
- 8 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.
- 9 C. Li and H. Wonneberger, *Adv. Mater.*, 2012, **24**, 613–636.
- 10 J. T. Kirner and R. G. Finke, *J. Mater. Chem. A*, 2017, **5**, 19560–19592.
- 11 J. T. Kirner, J. J. Stracke, B. A. Gregg and R. G. Finke, *ACS Appl. Mater. Interfaces*, 2014, **6**, 13367–13377.
- 12 F. Ronconi, Z. Syrgiannis, A. Bonasera, M. Prato, R. Argazzi, S. Caramori, V. Cristino and C. A. Bignozzi, *J. Am. Chem. Soc.*, 2015, **137**, 4630–4633.
- 13 M. Bonchio, Z. Syrgiannis, M. Burian, N. Marino, E. Pizzolato, K. Dirian, F. Rigodanza, G. A. Volpato, G. La Ganga, N. Demitri, S. Berardi, H. Amenitsch, D. M. Guldi, S. Caramori, C. A. Bignozzi, A. Sartorel and M. Prato, *Nat. Chem.*, 2019, **11**, 146–153.
- 14 C. E. Creissen, J. Warnan, D. Antón-García, Y. Farré, F. Odobel and E. Reisner, *ACS Catal.*, 2019, **9**, 9530–9538.
- 15 B. Van Den Bosch, J. A. Rombouts, R. V. Orru, J. N. Reek and R. J. Detz, *ChemCatChem*, 2016, **8**, 1392–1398.
- 16 H. C. Chen, R. M. Williams, J. N. Reek and A. M. Brouwer, *Chem. – Eur. J.*, 2016, **22**, 5489–5493.
- 17 J. Belić, B. van Beek, J. P. Menzel, F. Buda and L. Visscher, *J. Phys. Chem. A*, 2020, **124**, 6380–6388.
- 18 J. P. Menzel, A. Papadopoulos, J. Belić, H. J. de Groot, L. Visscher and F. Buda, *J. Phys. Chem. C*, 2020, **124**, 27965–27976.
- 19 Y. Shao, H. J. de Groot and F. Buda, *ChemSusChem*, 2021, **14**, 479–486.
- 20 Y. Shao, J. M. De Ruiter, H. J. De Groot and F. Buda, *J. Phys. Chem. C*, 2019, **123**, 21403–21414.





- 21 F. De Angelis, S. Fantacci and A. Selloni, *Nanotechnology*, 2008, **19**, 424002.
- 22 A. Seidl, A. Görling, P. Vogl, J. Majewski and M. Levy, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **53**, 3764–3774.
- 23 A. Görling and M. Levy, *J. Chem. Phys.*, 1997, **106**, 2675–2680.
- 24 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133.
- 25 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, 864–871.
- 26 R. E. Skyner, J. L. McDonagh, C. R. Groom, T. Van Mourik and J. B. Mitchell, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6174–6191.
- 27 R. W. Godby, M. Schlüter and L. J. Sham, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 10159–10175.
- 28 E. Engel, J. A. Chevary, L. D. Macdonald and S. H. Vosko, *Zeitschrift für Phys. D Atoms, Mol. Clust.*, 1992, **23**, 7–14.
- 29 E. J. Baerends, O. V. Gritsenko and R. Van Meer, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16408–16425.
- 30 E. J. Baerends, *Phys. Chem. Chem. Phys.*, 2017, **19**, 15639–15656.
- 31 L. Hedin, *Phys. Rev.*, 1965, **139**, A796–A823.
- 32 G. Onida, I. Nazionale, R. T. Vergata, R. Scientifica and I. Roma, *Rev. Mod. Phys.*, 2002, **74**, 601.
- 33 D. Golze, M. Dvorak and P. Rinke, *Front. Chem.*, 2019, **7**, 1–66.
- 34 A. Migani, D. J. Mowbray, A. Iacomino, J. Zhao, H. Petek and A. Rubio, *J. Am. Chem. Soc.*, 2013, **135**, 11429–11432.
- 35 A. Migani, D. J. Mowbray, J. Zhao, H. Petek and A. Rubio, *J. Chem. Theory Comput.*, 2014, **10**, 2103–2113.
- 36 N. Marom, J. E. Moussa, X. Ren, A. Tkatchenko and J. R. Chelikowsky, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 1–15.
- 37 P. Umari, L. Giacomazzi, F. De Angelis, M. Pastore and S. Baroni, *J. Chem. Phys.*, 2013, **139**, 0147091–0147098.
- 38 N. Marom, T. Körzdörfer, X. Ren, A. Tkatchenko and J. R. Chelikowsky, *J. Phys. Chem. Lett.*, 2014, **5**, 2395–2401.
- 39 D. J. Mowbray and A. Migani, *J. Phys. Chem. C*, 2015, **119**, 19634–19641.
- 40 A. J. Bard and L. R. Faulkner, *ELECTROCHEMICAL METHODS Fundamentals and Applications*, John Wiley & Sons, Inc, New York, 2nd edn, 2001.
- 41 J. Preat, C. Michaux, D. Jacquemin and E. A. Perpète, *J. Phys. Chem. C*, 2009, **113**, 16821–16833.
- 42 J. E. Bartmess, *J. Phys. Chem.*, 1994, **98**, 6420–6424.
- 43 A. V. Marenich, J. Ho, M. L. Coote, C. J. Cramer and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15068–15106.
- 44 S. Grimme, *Chem. – Eur. J.*, 2012, **18**, 9955–9964.
- 45 A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.
- 46 R. F. Ribeiro, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2011, **115**, 14556–14562.
- 47 J. Ho, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2859–2868.
- 48 J. Ho, M. L. Coote, C. J. Cramer and D. G. Truhlar, in *Org. Electrochem.*, ed. O. Hammerich and B. Speiser, CRC Press, Boca Raton, FL, 5th edn, 2016, ch. 6, pp. 229–259.
- 49 C. J. Cramer and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 760–768.
- 50 A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco and F. J. Luque, *Acc. Chem. Res.*, 2009, **42**, 489–492.
- 51 J. Ho and M. L. Coote, *Theor. Chem. Acc.*, 2009, **125**, 3–21.
- 52 J. Ho, A. Klamt and M. L. Coote, *J. Phys. Chem. A*, 2010, **114**, 13442–13444.
- 53 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.
- 54 A. Klamt, V. Jonas, T. Bürger and J. C. Lohrenz, *J. Phys. Chem. A*, 1998, **102**, 5074–5085.
- 55 A. Klamt, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 699–709.
- 56 H. Neugebauer, F. Bohle, M. Bursch, A. Hansen and S. Grimme, *J. Phys. Chem. A*, 2020, **124**, 7166–7176.
- 57 J. Menzel, M. Kloppenburg, J. Belić, H. de Groot, L. Visscher and F. Buda, *J. Comput. Chem.*, 2021, **42**, 1–10.
- 58 C. O. Almbladh and U. Von Barth, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1985, **31**, 3231–3244.
- 59 E. J. Baerends, O. V. Gritsenko and R. Van Meer, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16408–16425.
- 60 J. D. Talman and W. F. Shadwick, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1976, **14**, 36–40.
- 61 F. J. Dyson, *Phys. Rev.*, 1949, **75**, 1736–1755.
- 62 M. S. Hybertsen and S. G. Louie, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **34**, 5390–5413.
- 63 M. Van Schilfgaarde, T. Kotani and S. Faleev, *Phys. Rev. Lett.*, 2006, **96**, 1–4.
- 64 T. Kotani, M. Van Schilfgaarde and S. V. Faleev, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2007, **76**, 1–24.
- 65 J. Wilhelm, D. Golze, L. Talirz, J. Hutter and C. A. Pignedoli, *J. Phys. Chem. Lett.*, 2018, **9**, 306–312.
- 66 J. Wilhelm, P. Seewald and D. Golze, *J. Chem. Theory Comput.*, 2021, **17**, 1662–1677.
- 67 A. Förster and L. Visscher, *J. Chem. Theory Comput.*, 2020, **16**, 7381–7399.
- 68 A. Förster and L. Visscher, *Front. Chem.*, 2021, **9**, 736591.
- 69 M. J. Van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers and P. Rinke, *J. Chem. Theory Comput.*, 2015, **11**, 5665–5687.
- 70 A. Stuke, C. Kunkel, D. Golze, M. Todorović, J. T. Margraf, K. Reuter, P. Rinke and H. Oberhofer, *Sci. Data*, 2020, **7**, 1–11.
- 71 A. Förster and L. Visscher, *J. Chem. Theory Comput.*, 2021, **17**, 5080–5097.
- 72 C. Röger and F. Würthner, *J. Org. Chem.*, 2007, **72**, 8070–8075.
- 73 B. A. Jones, A. Facchetti, M. R. Wasielewski and T. J. Marks, *J. Am. Chem. Soc.*, 2007, **129**, 15259–15278.
- 74 S. Bhosale, A. L. Sisson, N. Sakai and S. Matile, *Org. Biomol. Chem.*, 2006, **4**, 3031–3039.
- 75 R. S. K. Kishore, O. Kel, N. Banerji, D. Emery, G. Bollot, J. Mareda, A. Gomez-Casado, P. Jonkheijm, J. Huskens, P. Maroni, M. Borkovec, E. Vauthey, N. Sakai and S. Matile, *J. Am. Chem. Soc.*, 2009, **131**, 11106–11116.
- 76 P. Leowanawat, A. Nowak-Król and F. Würthner, *Org. Chem. Front.*, 2016, **3**, 537–544.



- 77 M. J. Ahrens, M. J. Tauber and M. R. Wasielewski, *J. Org. Chem.*, 2006, **71**, 2107–2114.
- 78 SCM, <http://www.scm.com>.
- 79 B. van Beek and J. Belić, *CAT version 0.8.7*, <https://github.com/nlesc-nano/CAT>, DOI: 10.5281/zenodo.3832787, 2020.
- 80 PLAMS, <https://github.com/SCM-NV/PLAMS>; <http://www.scm.com>.
- 81 M. Gaus, Q. Cui and M. Elstner, *J. Chem. Theory Comput.*, 2011, **7**, 931–948.
- 82 M. Kubillus, T. Kubař, M. Gaus, J. Řezáč and M. Elstner, *J. Chem. Theory Comput.*, 2015, **11**, 332–342.
- 83 X. Lu, M. Gaus, M. Elstner and Q. Cui, *J. Phys. Chem. B*, 2015, **119**, 1062–1082.
- 84 M. Gaus, A. Goetz and M. Elstner, *J. Chem. Theory Comput.*, 2013, **9**, 338–354.
- 85 M. Gaus, X. Lu, M. Elstner and Q. Cui, *J. Chem. Theory Comput.*, 2014, **10**, 1518–1537.
- 86 P. Deglmann and F. Furche, *J. Chem. Phys.*, 2002, **117**, 9535–9538.
- 87 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 88 E. Van Lenthe and E. J. Baerends, *J. Comput. Chem.*, 2003, **24**, 1142–1156.
- 89 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.
- 90 C. Pye and T. Ziegler, *Theor. Chem. Acc.*, 1999, **101**, 396.
- 91 C. Pye, T. Ziegler, E. van Lenthe and J. Louwen, *Can. J. Chem.*, 2009, **87**, 790.
- 92 ADF 2019.3, <http://www.scm.com>.
- 93 B. Delley, *Mol. Simul.*, 2006, **32**, 117–123.
- 94 N. L. Allinger, X. Zhou and J. Bergsma, *J. Mol. Struct.: THEOCHEM*, 1994, **312**, 69–83.
- 95 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
- 96 M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.*, 1999, **110**, 5029–5036.
- 97 M. A. Rohrdanz, K. M. Martins and J. M. Herbert, *J. Chem. Phys.*, 2009, **130**, 054112-1-8.
- 98 F. Würthner, *Chem. Commun.*, 2004, 1564–1579.
- 99 J. W. Knight, X. Wang, L. Gallandi, O. Dolgounitcheva, X. Ren, J. V. Ortiz, P. Rinke, T. Körzdörfer and N. Marom, *J. Chem. Theory Comput.*, 2016, **12**, 615–626.
- 100 L. Duan, F. Bozoglian, S. Mandal, B. Stewart, T. Privalov, A. Llobet and L. Sun, *Nat. Chem.*, 2012, **4**, 418–423.
- 101 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.

