In search of Covalent Organic Framework photocatalysts: A DFT-based screening approach

Beatriz Mourino, Kevin Maik Jablonka, Andres Ortega-Guerrero,* and Berend Smit*

Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Valais, Switzerland

E-mail: oandresg15@gmail.com; berend.smit@epfl.ch

Abstract

Covalent organic frameworks (COFs) stand out as prospective organic-based photocatalysts given their intriguing optoelectronic properties, such as visible light absorption and high charge-carrier mobility. The CURATED COFs is a database of reported experimental COFs that until now remained mostly unexplored for photocatalysis. In this work, we screen the CURATED COFs database for discovering potential photocatalysts using a set of DFT-based descriptors that cost-effectively assesses visible light absorption, preliminary thermodynamic feasibility of the desired pair of redox reactions, charge separation, and charge-carrier mobility. The workflow can shortlist 13 COFs as prospective candidates for water splitting, and identify materials (N_x-COF (x =0–3)) that have been reported as candidates for Hydrogen evolution reaction (HER). Overall, our strategy tackles the challenge of having too many COFs to explore by channeling the focus of future research to a selective group of COFs while elucidating insights into the structure design for a desired photocatalytic process.

Introduction

Photocatalysis offers a pathway for green energy and chemical industry alternatives such as sunlight-driven water splitting and CO_2 reduction. In such processes, solar-to-chemical energy conversion provides the driving force to generate renewable fuels and chemicals as a promising solution to the energy and environmental crisis. In short, vastly available sunlight in a photocatalytic system is absorbed, followed by photo-generation of separated charge carriers in the material valence (VB) and conduction (CB) bands. These charge carriers can participate in the desired surface reactions.[1,2]

Whether the promise of a photocatalysis-based sustainable future can reach industrial plants strongly depends on finding a material that can optimally fulfill requirements for each step of a photocatalytic process. An ideal photocatalyst should have valence and conduction band edges straddling the potentials of the targeted redox reactions, should absorb visible light, display high charge-carrier mobility, and have low detrimental electron-hole recombination. Covalent organic frameworks (COFs) stand out given their modular nature, optoelectronic properties, and stability. COFs are porous crystalline materials composed predominantly of earth-abundant, non-toxic elements and based on the connection of building blocks, *i.e.* linkers and linkages, through reversible covalent bonds.[3,4] Most COFs are layered materials of two-dimensional (2D) sheets, commonly referred to as 2D COFs,[5] displaying mechanical, optical, and electronic properties that can be extremely advantageous for photocatalytic properties, such as visible light absorption and high charge-carrier mobility.[6] Many COFs possess large surface areas, functional group tunability, and solution processability. The combination of such properties permits an unlimited number of functionality-based designed materials. Due to the above-mentioned advantages, there is an increasing interest in using COFs and their composites as photocatalysts in overall water splitting, H_2 generation, CO_2 reduction, and degradation of organic pollutants.[7,8]

Interestingly, a simple literature search in the Web of Science database (filtered for the keywords "Covalent organic framework" or "COF" and "photocatalysis" or "photocatalyst")

shows that to this moment, about 90% of the experimental COFs originally queried for this work remain unexplored for photocatalysis. Given a large number of already synthesized COFs,[9] computational screening approaches can aid the exploration of properties and structure-property relationships by quickly surveying many variables, thus redirecting experimental efforts.[10] Here, we developed a high-throughput density functional theory (DFT)-based workflow to discover photocatalytically active COFs. The strategy behind our workflow is to implement cost-effective DFT calculations enabling us to discover materials of interest given a high number of COFs structures. Our methodology defines a set of descriptors to evaluate the material's performance in the fundamental steps of photocatalysis. The descriptors are obtained through post-processing of DFT outputs to properly assess photocatalysis-specific features associated with the material's performance. The screening workflow takes as input the "Clean, Uniform, Refined with Automatic Tracking from Experimental Database" (CURATED) COFs,[11] a database of reported experimental COFs.

Screening strategy

With a defined set of materials to be evaluated, developing a thorough strategy is crucial as it will dictate the accuracy and feasibility of the screening protocol. The strategy employed in this work focuses on 1) determining DFT descriptors for the fundamental steps of a photocatalytic process and the approximations to be employed and 2) establishing a cost-effective way to compute those descriptors.

Photocatalytic DFT descriptors Firstly, we determine DFT-based descriptors of the photocatalytic steps considering a strategy that enables us to obtain satisfactory yet cost-effective results for a screening study. In this work, the selected descriptors are the following: 1) energy-based descriptors, *i.e.* band gap and band-edge alignment to the redox reactions to assess visible-light absorption and thermodynamics, respectively, 2) a charge separation descriptor, and 3) a charge-carrier mobility descriptor to assess the performance of the

material.

Energy-based descriptors The thermodynamic feasibility of a photo-redox reaction can be evaluated by ionization potential (IP) and electron affinity (EA) values. The IP can be associated with the absolute energy of the valence band edge after alignment with a reference level. The EA is taken to be the difference between the IP and the band gap.[12,13] This association is an approximation since Kohn-Sham energy levels are affected by the derivative discontinuity where common density functional approximations (Generalized gradient approximation (GGA), Meta-GGA, and Hybrids) differ from experimental values.[12,14] The reference level can be estimated by determining the vacuum potential in the pores where the variation of the electrostatic potential is the smallest, similar to what has been previously reported for MOFs and COFs.[15,16] The obtained absolute energies of IP and EA can then be used to select materials whose band edges align with the target redox potentials of the desired reactions, all referenced to vacuum. Hence, in this way, we ensure the thermodynamic feasibility of the desired reaction. In the case of hydrogen (HER) and oxygen (OER) evolution reactions, the redox potentials at pH 0 are $-4.4\,\mathrm{eV}$ and $-5.63\,\mathrm{eV}$ aligned to the vacuum level,[17] respectively.

To evaluate visible-light absorption, the Kohn-Sham band gap is calculated and empirically adjusted to PBE0 values, as discussed in the following section. Ideally, computational determination of the optical band gap should rely on methods that account for the excited states, especially for organic materials where excitonic effects are important.[12,18,19] However, some hybrid functionals, such as PBE0, can be reasonably accurate for some systems when calculating the optical band gap; and an empirical adjustment of PBE to PBE0 values employed here is justified by its cost-effectiveness. Hybrid functionals are particularly accurate in predicting optical band gaps when no charge transfer excitation is involved.[20]

Charge separation descriptor The exciton generated by light absorption can then undergo the process of charge separation (independent charge carriers, *i.e.* electron and

hole), which prevents electron-hole recombination and subsequent carrier loss for a higher conversion efficiency in photocatalysis. Charge separation can be calculated as the weighted average of the spatial overlap (Λ) between virtual and occupied Kohn-Sham orbitals in the lowest excited singlet state, which can be obtained from time-dependent DFT (TDDFT) calculations.[21] Aiming for cost-effectiveness, here we adopt a previously reported strategy that allows for using ground-state unrestricted Kohn-Sham (UKS) DFT calculations (PBE based and PBE0 adjusted, see Figure S1) for charged doublets, *i.e.* -1 for electron injection and +1 for hole injection.[21] In this case, the charge separation descriptor can be estimated similarly, but now considering the averaged spatial overlap between the highest occupied (HOMO(α)) and lowest unoccupied (LUMO(β)) molecular orbitals of the electron and the hole injection, respectively. When the photo-generated charges are spatially separated, this could favor long lifetimes and low recombination rates. Here we also refer to the overlap as the "charge recombination descriptor" to intuitively point out lower overlaps as optimal values for lower charge recombination and more separated charges.

Charge-carrier mobility descriptor Lastly, to qualitatively evaluate charge-carrier mobility, the carrier effective masses (m^*) were computed. According to the Bardeen-Schockley model,[21–23] m^* is inversely proportional to the electronic charge-carrier mobility. Therefore lower values of m^* are associated with higher carrier mobility. This strategy avoids the computation of the bulk modulus and the Bardeen-Shockley deformation potential for the band edges, which would be necessary for computing charge-carrier mobility requiring excessive computational resources for screening the entire database. In the parabolic approximation, the effective mass tensor can be reduced to a scalar quantity and can be computed directly from the band structure, taking into account the curvature of the VB (for m^* of holes) and CB (for m^* of electrons).[22]

PBE *vs* PBE0 A challenge with extended structures such as metal-organic frameworks (MOFs) and COFs is that their unit cell contains many atoms, which makes simulations at

high fidelity computationally prohibitive. One way to approach the cost *vs* accuracy issue, as previously reported for MOFs, is to employ a scheme to evaluate the DFT descriptors at a GGA-level (e.g., PBE[24]), but making use of a post-processing empirical adjustment to reproduce hybrid functional (e.g., PBE0) results.[15,21] The hypothesis that such an adjustment exists is based on the systematic employment of 25% of Hartree-Fock (HF) exact exchange in PBE0.[25,26] Moreover, the nature of the orbitals contributing to the band gap in COFs is mostly unchanged between PBE and PBE0. Unlike MOFs, most COFs do not contain transition and post-transition metals. Their organic element composition reduces the presence of the charge transfer mechanism in MOFs like ligand-to-metal and metal-to-ligand charge transfer, which would require higher levels of theory than hybrid functionals.

To apply this scheme to the database, we first tested the correlation between PBE and PBE0 descriptors values for a small subset. We manually selected 20 COFs to compose a representative subset where we could evaluate the above-mentioned PBE-based strategy to determine the validity of the strategy in our systems of interest. The dataset is chosen to contain 2D as well as 3D COFs, also including structures with heteroatoms, and is displayed in Table S1. Figure 1 shows that PBE and PBE0 display a linear relationship for ionization potential (IP), electron affinity (EA), and band gap energies (R^2 rounded values of 0.99, 1.00, and 0.99, respectively), even more than MOFs (R^2 values of 0.92, 0.98, and 0.93 respectively).[15] For band gap values, we observed the well-established band gap opening effect of hybrids compared to GGA functionals.[12] Similarly, the charge separation descriptor can be empirically adjusted to PBE0 values based on the correlation displayed in Figure S1.[21]

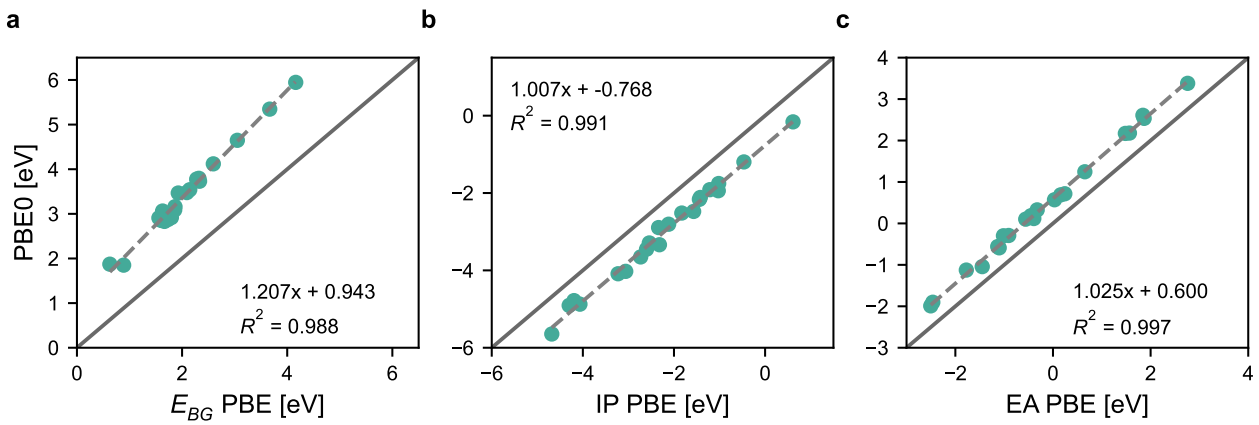


Figure 1: PBE and PBE0 correlation for ionization potential (IP), electron affinity (EA) and band gap values in a set of 20 selected CURATED COFs (see Table S1).

Figure 2 depicts the workflow we developed to compute the photocatalytic descriptors, with empirical adjustments performed during data treatment. More details regarding the decision-making process for the parameters in the workflow are displayed in the SI.

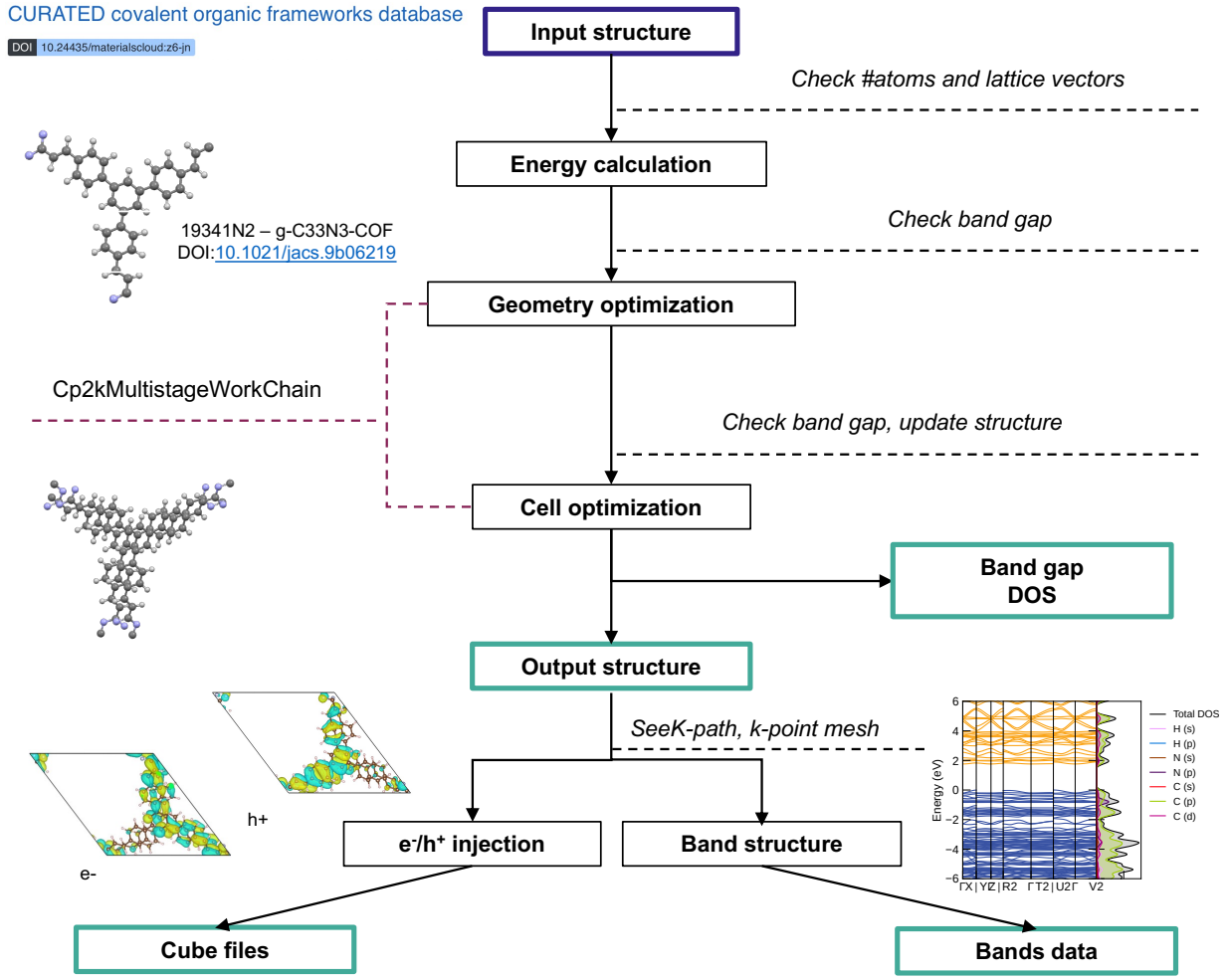


Figure 2: Flowchart of the workflow developed for this screening study. It takes as input the structures from the CURATED COFs database (box highlighted in Persian indigo), then checks the number of atoms and lattice parameters. The next step is a single-point energy calculation. If this calculation confirms that we have a closed-shell system with a semiconducting band, the structure is optimized. From the optimized structure, the workflow computes the main outputs (boxes highlighted in jungle green) utilized to calculate our photocatalytic DFT descriptors, namely, band gap and band alignment (with band gap and cube files printed after optimization), charge separation (with cube files from electron and hole injection) and charge carrier effective masses (with bands data).

Results and discussion

From now on, the case study of overall water splitting (OWS) is considered. The sunlight-driven overall splitting of water in H_2 and O_2 is often referred to as the "Holy Grail" in the energy landscape, as it is a renewable way of generating energy carriers in a simple but robust

strategy from clean and widely available energy sources.[27,28] Although OWS for COFs does not come without challenges, e.g., slow kinetics of OER and insufficient driving force from photogenerated charge carriers,[29,30] this case study can guide the discussion, especially for the preliminary evaluation of thermodynamic feasibility. This can be evaluated by band edge alignment with the redox potentials of the reaction. An analogous analysis could be applied to any pair of redox reactions, also with sacrificial agents, co-catalysts, and in a Z-scheme.[31]

Photocatalytic DFT descriptors

Firstly, we take a closer look at each photocatalytic DFT descriptor. To aid the evaluation of structure-property relationships for our descriptors, we determined a set of 26 substructures (see SI for the complete list) based on chemical intuition and on their occurrence as building blocks (linkages and linkers) and functional groups in COFs.[8,30,32] Searching for the presence of those substructures in the CURATED COFs allows for methodical pattern-seeking when it comes to the structure-property relationships. Such a pattern has been shown previously, indicating benzene-based functional groups with C_{3v} symmetry and nitrogen-based functional groups as potential catalytic active sites for OER and HER, respectively.[30] Among the defined substructures, we proceed to discuss the ones that presented statistically significant effects on our photocatalytic DFT descriptors, not considering the ones with only a few data points or that showed no effect.

Energy-based descriptors 75% of the evaluated CURATED COFs have their band gaps in the range of visible light ($1.6\,\mathrm{eV} < E_{bg} < 3.2\,\mathrm{eV}$). Structural analysis shows that most structures containing porphyrin, pyrene, triphenylamine, and 2,4,6-triphenyl-1,3,5-triazine (TPTA, with triazine linkage) have their PBE0-adjusted band gaps within the visible range (Figure 3). This observation agrees with the photophysical nature of these organic groups known for having visible-light absorbing properties.[13,23,33–38] The presence of these groups has been reported for applications involving visible light absorption in COFs.[33,39–41]

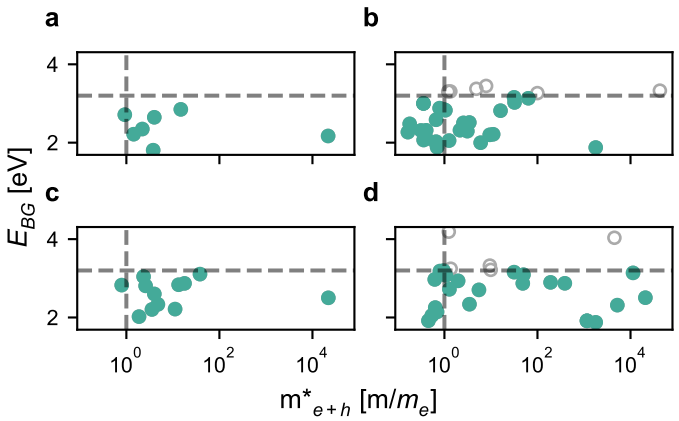


Figure 3: All COFs containing a) porphyrin, b) pyrene, c) triphenylamine, and d) 2,4,6-triphenyl-1,3,5-triazine. Jungle green dots represent structures among them with band gaps in the visible range (porphyrin and triphenylamine 100%, pyrene 83%, 2,4,6-triphenyl-1,3,5-triazine 85%). Grey circles represent structures with those functional groups but out of the visible range (porphyrin and triphenylamine 0%, pyrene 17%, and 2,4,6-triphenyl-1,3,5-triazine 15%). The horizontal line represents the upper limit of the visible light range, the vertical line represents $m^*_{e+h} = 1m_e$.

Moreover, 14% of the COFs' ionization potation (IP) and electron affinity (EA) values straddle the OER and HER potentials, all with band gaps within the visible range. Figure 4 shows the band edge alignment for the filtered structures in this case study. Indeed, we see in Figure 4 that all values for electron affinity are above the hydrogen evolution reaction potential (jungle green dashed line), and all values for ionization potential are below the oxygen evolution reaction potential (Persian indigo dashed line). Such straddling indicates thermodynamic feasibility for OWS with visible light absorption (the black dashed line is the maximum value in the visible range). Tighter alignment, *i.e.* lower band gaps ($E_{bg} < 2.36\,eV$) in this plot are desired to achieve the minimum requirement of 10% solar-to-hydrogen (STH) efficiency for profitable industrialization.[42] Furthermore, the filtered structures can also be investigated as photocatalysts for individual HER and OER.

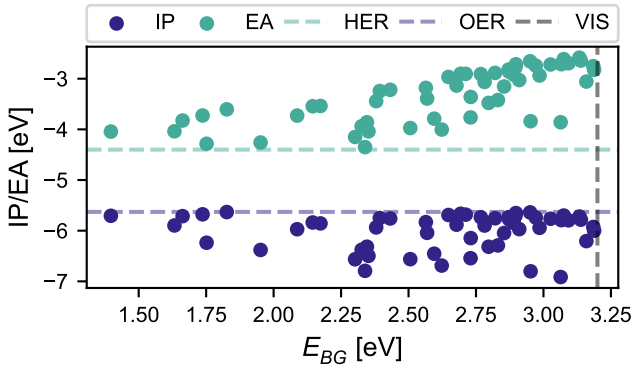


Figure 4: Simultaneous filtering of the energy-based DFT descriptors computed for 419 CURATED COFs. For the case study of overall water splitting, good candidates should have their IP and EA straddling the redox potentials of HER (jungle green dashed line) and OER (Persian indigo dashed line), all aligned to vacuum, and band gap in the visible range (grey dashed line as the maximum value). IP and EA in the same vertical line correspond to the same structure.

We highlight that all calculations are kept at the GGA level, with the PBE functional, and posteriorly adjusted empirically to PBE0 values. This choice is justified by our screening approach, considering the large number of COFs. However, DFT, even with hybrid functionals, can present shortcomings regarding the prediction of fundamental band gaps and IP/EA levels. For GGA functionals, one example is the difficulty in capturing excitonic effects in organic systems, which can be in the order of a few eV and severely impact visible light absorption.[18] Therefore, this approach might not be enough when compared to more accurate calculations, but it can at least describe the nature of the electronic properties of COFs. Moreover, the strategy adopted here allows for a cost-effective way of shortlisting photocatalyst candidates, which can later be redirected for more accurate calculations.

Charge separation descriptor The charge separation descriptors computed in this work range between 0.037 and 0.931 (it can go from 0 to 1). To test the validity of our charge separation descriptor, we compared our results with reported lifetimes in COFs. Our computed values agree qualitatively with experimental results of charge carrier lifetime for, e.g., COF-366 (low Λ=0.3, lifetime of $\approx$80 µs) and N_3-COF (high Λ=0.7, average lifetime of

≈4 ns).[43,44] Considering that lower values of Λ are an indication of lower chances of electron-hole recombination, we filtered the structures presenting Λ less than 0.5 as potentially favorable for charge separation. The structures obtained after this filter represent 28% of the database. With a lower threshold of 0.15, 13 structures are filtered, 4 of them composing a series of multiple-component donor-acceptor COFs (MC-COF-TPs), strategically designed to contain spatially separated electron donor and acceptor groups.[45] Those COFs contain the group 2,3,6,7,10,11-hexahydroxytriphenylene (TP) as knots, benzene-1,4-dialdehyde, nitro-*p*-phenylenediamine and *p*-phenylenediamine as electron donors and 2,1,3-benzothiadiazole-4,7-diboronic acid (BTDADA) as an electron acceptor, with N, B, and S heteroatoms present and spatially separated in each structure.[45] The authors state that charge transfer is triggered from TP to BTDADA and that the sequenced π-arrays may play an important role in the electronic correlations.[45] The presence of groups that can stabilize charge carriers is crucial for enhancing photocatalytic activity and is desired if one would like to design COFs with low detrimental charge recombination.[46] The remaining nine structures with the lowest computed charge recombination descriptors ($\Lambda < 0.15$) are BP-COF-1 and BP-COF-2 (both with B, S, P, O, and F as heteroatoms), PIA-AA and PIC-AA (both with dicarboximide groups), CCOF-1, NN-TAPH-COF, (R)-DTP-COF, DhaTab, and COF-119.[47–52] Among them, triphenylbenzene seems to be a common group (for DhaTab, PIC-AA, and (R)-DTP-COF).

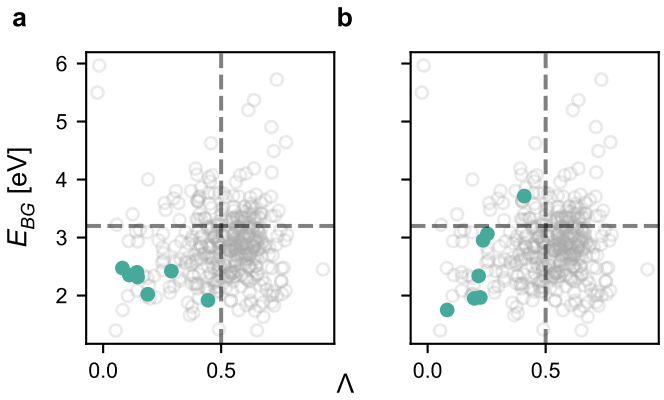


Figure 5: Structural analysis of all COFs distributed by their band gap and overlap values, with jungle green dots representing all structures that contain a) thiadiazole, or b) phthalimide. The horizontal line represents the upper limit of the visible light range, and the vertical line represents $\Lambda = 0.5$.

Moreover, our substructure-based structural analysis shows that the presence of thiadiazole and phthalimide is associated with low overlap values, as displayed in Figure 5. All COFs with those substructures have a charge separation descriptor lower than 0.5. Statistical analysis with bootstrapped effect sizes corroborates the lowering of charge recombination descriptor with the presence of those groups (see Figures S14 and S15).

Charge-carrier mobility descriptor Concerning the charge-carrier mobility descriptor, 67% of the COFs have effective masses of both electron and hole below ten times the electron rest mass (m_e). The threshold is chosen because most of the effective masses reported for COFs with high conductivity in the literature are below this value.[44,53–55] At a much tighter threshold of 1 m_e, still, 35% remain, among which only 6% are 3D COFs. Figure 6 shows a distribution of the effective masses of both electron and hole for the evaluated CURATED COFs. Out of the structures with effective masses of either electron or hole higher than 100 m_e, 18% are 3D COFs.

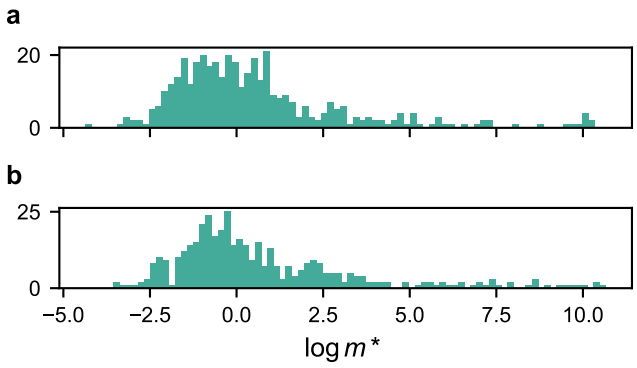


Figure 6: Distribution of effective masses of a) electron and b) hole computed for 419 CURATED COFs.

When comparing the COFs with very low (m*<1 m_e) and very high (m*>100 m_e) effective masses, it was observed that COFs with low effective masses uniquely contain β-ketoenamine linkages with the group 2,4,6-trimethanimidoylbenzene-1,3,5-triol (TIBT), and COFs with high effective masses uniquely contain glutarimide (with dicarboximide group). In Figure 7, a pattern can be seen for β-ketoenamine-linked COFs with TIBT, its presence occurring mostly on COFs with low effective masses and appropriate IP/EA values for aligning with HER. Statistical analysis with bootstrapped effect sizes shows a significant difference between the means of effective masses of COFs with and without β-ketoenamine linkage with TIBT (lowering of 2 units in the log scale for both electron and hole Figures S12,S13). Indeed, β-ketoenamine-linked COFs usually display good photocatalytic properties for HER, with increased stability with respect to their imine counterparts.[8,56] However, the same COFs, except for a few, present higher values of the computed charge recombination descriptor (Figure 7b). It has been suggested that introducing donor-acceptor units to COFs with this linkage can enhance their photocatalytic activity by stabilizing excited charge transfer processes.[56]

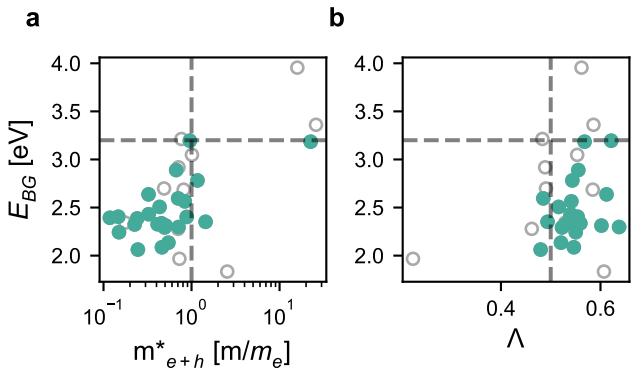


Figure 7: Structural analysis of all the β-ketoenamine-linked COFs with TIBT. Jungle green dots represent structures with band edges aligning with the HER redox potential, and grey circles represent those that do not align. HER alignment occurs mostly in structures with low effective masses and appropriate band gaps but relatively high overlap. The horizontal line represents the upper limit of the visible light range, and the vertical line represents in a) $m^*_{e+h} = 1m_e$, and in b) $\Lambda = 0.5$.

Concerning the before-mentioned MC-COF-TPs series with low charge recombination descriptors, their uniformly distributed N and S heteroatoms are also expected to enhance charge-carrier mobility.[57] Such observation is corroborated by the computed m* for both electron and hole for MC-COF-TPs, that range between 5e−10 to 13 m_e, with the exception of m^*_h for MC-COF-TP-$E_1{}^2E_2{}^1$. In summary, when designing COFs to enhance charge-carrier mobility, choosing 2D dimensionality and β-ketoenamine linkage with TIBT could be advantageous.

Moreover, the presence of dicarboximide functional groups was associated with lower charge recombination descriptors, but higher effective masses. A similar pattern is present for CCOF-1, with high effective masses but low charge recombination descriptor. Such observations could derive from lower band dispersion, which reduces the possibility of detrimental charge recombination, but can also reduce charge-carrier mobility. In fact, this behavior is observed in CCOF-1, with calculated band dispersion for VB and CB of 0.72 and 1.4 meV, respectively. Those values represent the absolute difference between the highest and lowest energy values for each band.

Shortlisted candidates for the case study of OWS

Combining the optimal values for all the descriptors, we end up with a set of 13 COFs that are promising for the case study of overall water splitting. The structures of interest are represented by the dark green and blue dots in the specified bottom-left region in Figure 8a, and detailed in Table S2. A thorough literature review shows that, to date, some of the filtered COFs have been explored for different photocatalytic processes, such as CO_2 reduction,[58] which corroborates their usage as light harvesters.

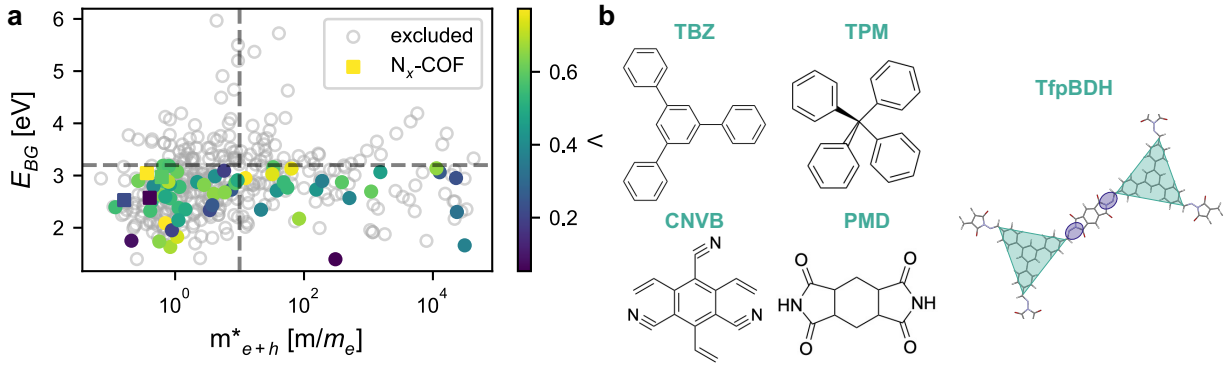


Figure 8: a) Simultaneous filtering of all the computed DFT descriptors for 419 CURATED COFs for the case study of overall water splitting. The region of interest is in the lower left region, which indicates more appropriate band gaps and low carrier effective masses. Grey dots are structures whose band edges do not straddle HER and OER redox potentials. Dark colors indicate a lower possibility of charge recombination. N_x-COFs filtered for HER and already investigated for this application are represented by colored squares. b) Common substructures in the set of 13 filtered COFs for OWS. To the right, the backbone is highlighted, which is similar to the donor-acceptor COFs designed for OWS, with HER and OER active sites highlighted in jungle green and Persian indigo, respectively.[30]

Most filtered COFs, however, have not yet been explored for photocatalytic HER/OER (79%, see Table S2), thus highlighting the usefulness of the current work and suggesting further theoretical and/or experimental investigation. Among the 13 COFs, common substructures are 1,3,5-(triphenyl)benzene, pyrometallic diimide, tetraphenylmethane, and 1,3,5-tricyano-2,4,6-tris(vinyl)benzene, see Figure 8b. Common linkages are imine, imide, azo, and β-ketoenamine (with TIBT). Moreover, some COFs in the final list, such as TfpBDH and PI-COF-SR, display a similar backbone compared to donor-acceptor COFs designed strategically for

photocatalytic OWS (Figure 8b).[30]

As a proof of concept, N_x-COFs (x=0-3) were evaluated by this work as prospective candidates for photocatalytic HER and are already reported for this application.[59] Indeed, it showed proper alignment of IP/EA, band gap values within the visible range that agree reasonably with experimental results and good results for effective masses (below 5 m_e), see Table S3 for more details. Figure 8a represents this series in colored squares. N_3-COF, in particular, has its IP/EA straddling both HER and OER, thus could also be a candidate for OWS. Although its charge separation descriptor is slightly higher than our threshold of 0.5, it was nevertheless reported among the N_x-COF series as the most effective HER photocatalyst when triethanolamine is chosen as the sacrificial donor.

Structure-property relationship

Figure 9 summarizes the effect of the defined building blocks and functional groups whose presence statistically affects at least one of our photocatalytic DFT descriptors. The presence of β-ketoenamine with TIBT, for example, promotes a statistically significant lowering of charge carrier effective masses, but increases the values of the charge recombination descriptor. More details on the statistical analysis are displayed in the SI, showing as an example the statistically-based effect of β-ketoenamine (with TIBT) on each descriptor (Figures S10-S13).

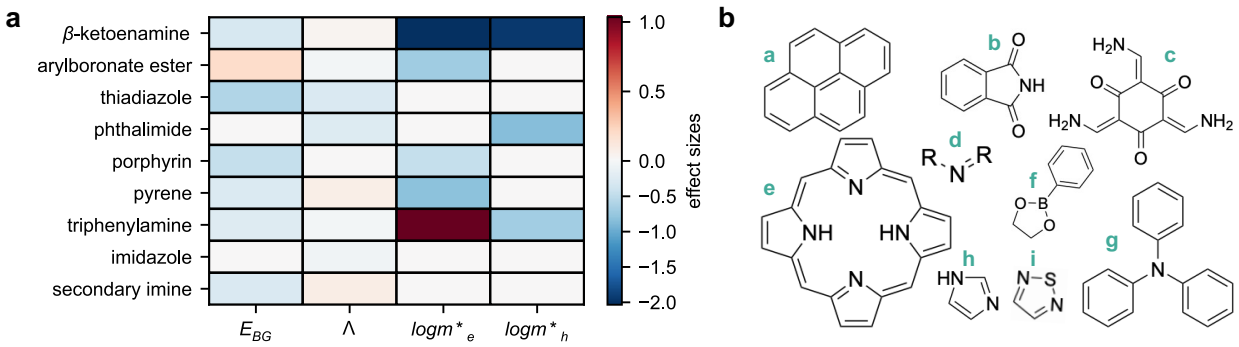


Figure 9: a) Effect of the presence of defined substructures on the band gap, charge recombination descriptor, and effective masses based on statistical analysis with bootstrapped effect sizes. A lowering trend is desired for Λ and m*, indicating lower chances of electron and hole recombination, and higher charge carrier mobility. b) Visualization of substructures, namely, a. pyrene, b. phthalimide, c. TIBT (β-ketoenamine linkage), d. secondary imine, e. porphyrin, f. arylboronate ester (2-phenyl-1,3,2-dioxaborolane), g. triphenylamine, h. imidazole, and i. 1,2,5-thiadiazole.

Furthermore, topology analysis with CrystalNets.jl[60] in the "all nodes" clustering was successfully performed for 300 structures. The distribution of topologies found is displayed in Figure S16. The results show that most structures aligned for OWS and individual HER and OER present honeycomb (**hcb**), augmented honeycomb (**hca**), **fes**, and **hnb** topologies. For the case of HER, although the occurrence of **hbn** topology is practically unchanged in the filtered COFs (2.3% of occurrence in all COFs, 2.7% in the filtered ones for HER), we noticed an increase in the occurrence of **hca** (64.5% of occurrence in all COFs, 75.7% in the filtered ones for HER), **hcb** (6.4% of occurrence in all COFs, 10.8% in the filtered ones for HER), and **fes** (5.0% of occurrence in all COFs, 10.8% in the filtered ones for HER). In the literature, COFs that have been reported for HER oftentimes display honeycomb (**hcb** and **hca**) topologies.[61,62] Our findings suggest that **fes** nets could be another possible direction of exploration for photocatalytically active COFs.

Other photo-redox reactions

To allow for an in-depth analysis of the evaluated COFs, we have chosen the study case of overall water splitting. However, as previously mentioned, the strategy adopted here can be applied to any pair of redox reactions without the need to perform further calculations. As a demonstration of the versatility of this work, we developed a bokeh application that allows for interactive filtering of the evaluated COFs. Such filtering can be done by choosing different redox reactions and different thresholds for the charge carrier mobility and charge recombination descriptors. The video in the supplementary materials exemplifies the usage of this application by filtering candidates for photocatalytic CO_2 reduction to methanol, with triethanolamine (TEOA) as a sacrificial agent. The process of reducing CO_2 to value-added chemicals such as methanol is a compelling alternative to recycle CO_2 from the environment.[63] After applying the desired filters, which can also be customized to feasible values, it is possible to obtain the dataset of the filtered COFs.

Conclusions

The screening approach employed by this work allowed the shortlisting of photocatalytically active COFs based on a cost-effective calculation of DFT-based descriptors. We selected 13 COFs as prospective photocatalysts for OWS after screening the database for alignment to redox potentials, visible light absorption, charge separation, and charge-carrier mobility descriptors. The list of selected COFs for HER contains the series of N_x-COF (x=0-3) already studied experimentally for this purpose. Interestingly, our approach can seamlessly be extended to filter candidates for any desired pair of photo-redox reactions.

Structural analysis for each DFT descriptor provided suggestions for the rational design of photocatalytically active COFs. Porphyrin, pyrene, triphenylamine, and TPTA-containing COFs were associated with visible light absorption, which can be considered when designing photocatalytic COFs to harvest sunlight. Most β-ketoenamine-linked COFs with TIBT in

the database have their IPs/EAs aligned to the HER potential. Concerning charge separation, COFs containing thiadiazole and phthalimide groups display a statistically significant lowering of the computed averaged spatial overlap that assesses the likelihood of charge recombination. Furthermore, the choice of 2D dimensionality and β-ketoenamine linkages with TIBT can benefit the design of COFs with enhanced charge-carrier mobility.

We highlight that other aspects need to be further considered for evaluating photocatalytic performance, such as photo-stability, stacking modes, and kinetics, which are open questions for this screening approach. Suggested next steps are the experimental evaluation of the shortlisted COFs in Table S2 for OWS and further theoretical investigations with higher accuracy. Future works could also address the influence of how stacking modes and statistical stacking order can affect our photocatalytic descriptors. Altogether, this work can pave the way for using in-silico methodologies to design COFs for photocatalysis.

Computational methods

The Automated Interactive Infrastructure and Database for Computational Science (AiiDA,[64,65] v1.6.5) was used as a workflow manager to perform the DFT calculations for 419 structures from the CURATED COFs database.[11] A new workchain was built as depicted by the flowchart in Figure 2 based on workchains for multistage optimization and base calculations from the plugins aiida-cp2k and aiida-lsmo. The logic of the developed workchain consists of evaluating the band gap to ensure a closed-shell system/semiconducting band gap, followed by optimization of the cell parameters and calculations of our DFT descriptors for photocatalysis.[15,21]

Ground-state unrestricted Kohn-Sham (UKS) DFT calculations were employed for electron and hole injection to evaluate the charge separation descriptor as previously defined.[21]

Electronic chemical potentials were aligned to vacuum as reported by Fumanal et al.[15], based on the procedure established by Butler et al.[16] The band extrema energies were

corrected with our empirical adjustment to mimic PBE0 values.

For the charge separation descriptor, an in-house code was developed to compute the averaged spatial overlap of the cube files for electron and hole injection.[66]

The parabolic approximation implemented in sumo library[67] was used to compute effective masses of electrons and holes based on the electronic bands' dispersion in the reciprocal space.[21,22]

More details about the DFT calculations can be found in the Supporting Information. Structural analysis was performed with the CSD substructure search Python API via SMARTS,[68,69] and we used CrystalNets.jl[60] for topology analysis. It was necessary to consider flexible SMARTS definition, *i.e.* not explicitly specifying aromaticity or unsaturations, for proper recognition with the CSD substructure search due to incompatibilities in bond lengths. Other computational tools utilized in this work are detailed in the Supporting Information.

Statistics

We performed statistical analysis to determine whether the presence of a building block or functional group could have positive, negative, or no effects on our photocatalytic DFT descriptors. We employed boostrapped effect sizes with the DABEST package[70] for quantitative estimation of effect sizes. Detailed information on the statistical tests is available as a supplementary material.

Acknowledgement

The authors acknowledge PRACE and MARVEL for awarding access to Piz Daint (project ID: pr128) and Eiger (project ID: mr30) at the Swiss National Supercomputing Centre (CSCS), Switzerland. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945363.

Supporting Information Available

Further details on the datasets, computational methods, statistical analysis and preliminary tests are detailed in the Supporting Information.

Authors contributions

A.O.-G., K.M.J. and B.M. designed the research strategy, K.M.J. and B.M. developed the workflow, B.M. performed data analysis. B.M. wrote the first draft. B.M., A.O.-G., K.M.J. and B.S. contributed to writing and revising the manuscript, directing the work and providing crucial inputs.

Data availability

The aiida workflow and post-processing codes utilized by this work are available on `https://github.com/bmourino/cof_photocatalysis`. The final dataset with the values for all the descriptors, and a separate dataset containing details on statistical analysis with bootstrapped effect sizes are available on `https://doi.org/10.5281/zenodo.7590815`. The interactive application developed is available on `https://go.epfl.ch/cof-photocatalysis-app`.

Ethics declarations

Competing interests

The authors declare no competing interests.

References

(1) Galushchinskiy, A.; González-Gómez, R.; McCarthy, K.; Farràs, P.; Savateev, A. Progress in Development of Photocatalytic Processes for Synthesis of Fuels and Organic Compounds under Outdoor Solar Light. *Energy & Fuels* **2022**, *36*, 4625–4639.

(2) Melchionna, M.; Fornasiero, P. Updates on the Roadmap for Photocatalysis. *ACS Catalysis* **2020**, *10*, 5493–5501.

(3) Cote, A. P.; Benin, A. I.; Ockwig, N. W.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. Porous, Crystalline, Covalent Organic Frameworks. *Science* **2005**, *310*, 1166–1170.

(4) Diercks, C. S.; Yaghi, O. M. The atom, the molecule, and the covalent organic framework. *Science* **2017**, *355*.

(5) Zhang, Y.; Položij, M.; Heine, T. Statistical Representation of Stacking Disorder in Layered Covalent Organic Frameworks. *Chemistry of Materials* **2022**, *34*, 2376–2381.

(6) Chen, R.; Wang, Y.; Ma, Y.; Mal, A.; Gao, X.-Y.; Gao, L.; Qiao, L.; Li, X.-B.; Wu, L.-Z.; Wang, C. Rational design of isostructural 2D porphyrin-based covalent organic frameworks for tunable photocatalytic hydrogen evolution. *Nature Communications* **2021**, *12*.

(7) Nguyen, H. L.; Alzamly, A. Covalent Organic Frameworks as Emerging Platforms for CO2 Photoreduction. *ACS Catalysis* **2021**, *11*, 9809–9824.

(8) Li, Y.; Song, X.; Zhang, G.; Wang, L.; Liu, Y.; Chen, W.; Chen, L. 2D Covalent Organic Frameworks Toward Efficient Photocatalytic Hydrogen Evolution. *ChemSusChem* **2022**, *15*, e202200901.

(9) Ongari, D.; Talirz, L.; Smit, B. Too Many Materials and Too Many Applications: An Experimental Problem Waiting for a Computational Solution. *ACS Central Science* **2020**, *6*, 1890–1900.

(10) Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R. Efficient Computational Screening of Organic Polymer Photovoltaics. *The Journal of Physical Chemistry Letters* **2013**, *4*, 1613–1623.

(11) Ongari, D.; Yakutovich, A. V.; Talirz, L.; Smit, B. Building a Consistent and Reproducible Database for Adsorption Evaluation in Covalent–Organic Frameworks. *ACS Central Science* **2019**, *5*, 1663–1675.

(12) Kronik, L.; Stein, T.; Refaely-Abramson, S.; Baer, R. Excitation Gaps of Finite-Sized Systems from Optimally Tuned Range-Separated Hybrid Functionals. *Journal of Chemical Theory and Computation* **2012**, *8*, 1515–1531, PMID: 26593646.

(13) Ortega-Guerrero, A.; Fumanal, M.; Capano, G.; Smit, B. From Isolated Porphyrin Ligands to Periodic Al-PMOF: A Comparative Study of the Optical Properties Using DFT/TDDFT. *The Journal of Physical Chemistry C* **2020**, *124*, 21751–21760.

(14) Hait, D.; Head-Gordon, M. Delocalization Errors in Density Functional Theory Are Essentially Quadratic in Fractional Occupation Number. *The Journal of Physical Chemistry Letters* **2018**, *9*, 6280–6288.

(15) Fumanal, M.; Capano, G.; Barthel, S.; Smit, B.; Tavernelli, I. Energy-based descriptors for photo-catalytically active metal–organic framework discovery. *J. Mater. Chem. A* **2020**, *8*, 4473–4482.

(16) Butler, K. T.; Hendon, C. H.; Walsh, A. Electronic Chemical Potentials of Porous Metal–Organic Frameworks. *Journal of the American Chemical Society* **2014**, *136*, 2703–2706, PMID: 24447027.

(17) Tamirat, A. G.; Rick, J.; Dubale, A. A.; Su, W.-N.; Hwang, B.-J. Using hematite for photoelectrochemical water splitting: a review of current progress and challenges. *Nanoscale Horizons* **2016**, *1*, 243–267.

(18) Wang, H.; Jin, S.; Zhang, X.; Xie, Y. Excitonic Effects in Polymeric Photocatalysts. *Angewandte Chemie International Edition* **2020**, *59*, 22828–22839.

(19) Botti, S.; Sottile, F.; Vast, N.; Olevano, V.; Reining, L.; Weissker, H.-C.; Rubio, A.; Onida, G.; Sole, R. D.; Godby, R. W. Long-range contribution to the exchange-correlation kernel of time-dependent density functional theory. *Physical Review B* **2004**, *69*.

(20) Janesko, B. G.; Henderson, T. M.; Scuseria, G. E. Screened hybrid density functionals for solid-state chemistry and physics. *Phys. Chem. Chem. Phys.* **2009**, *11*, 443–454.

(21) Fumanal, M.; Ortega-Guerrero, A.; Jablonka, K. M.; Smit, B.; Tavernelli, I. Charge Separation and Charge Carrier Mobility in Photocatalytic Metal-Organic Frameworks. *Advanced Functional Materials* **2020**, *30*, 2003792.

(22) Muschielok, C.; Oberhofer, H. Aspects of semiconductivity in soft, porous metal-organic framework crystals. *The Journal of Chemical Physics* **2019**, *151*, 015102.

(23) Kinik, F. P.; Ortega-Guerrero, A.; Ebrahim, F. M.; Ireland, C. P.; Kadioglu, O.; Mace, A.; Asgari, M.; Smit, B. Toward Optimal Photocatalytic Hydrogen Generation from Water Using Pyrene-Based Metal–Organic Frameworks. *ACS Applied Materials & Interfaces* **2021**, *13*, 57118–57131, PMID: 34817166.

(24) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(25) Baerends, E. J. From the Kohn–Sham band gap to the fundamental gap in solids. An integer electron approach. *Phys. Chem. Chem. Phys.* **2017**, *19*, 15639–15656.

(26) Bao, J. L.; Gagliardi, L.; Truhlar, D. G. Self-Interaction Error in Density Functional Theory: An Appraisal. *The Journal of Physical Chemistry Letters* **2018**, *9*, 2353–2358, PMID: 29624392.

(27) Zhou, P.; Navid, I. A.; Ma, Y.; Xiao, Y.; Wang, P.; Ye, Z.; Zhou, B.; Sun, K.; Mi, Z. Solar-to-hydrogen efficiency of more than 9% in photocatalytic water splitting. *Nature* **2023**, *613*, 66–70.

(28) Bie, C.; Wang, L.; Yu, J. Challenges for photocatalytic overall water splitting. *Chem* **2022**, *8*, 1567–1574.

(29) Christoforidis, K. C.; Fornasiero, P. Photocatalytic Hydrogen Production: A Rift into the Future Energy Supply. *ChemCatChem* **2017**, *9*, 1523–1544.

(30) Wan, Y.; Wang, L.; Xu, H.; Wu, X.; Yang, J. A Simple Molecular Design Strategy for Two-Dimensional Covalent Organic Framework Capable of Visible-Light-Driven Water Splitting. *Journal of the American Chemical Society* **2020**, *142*, 4508–4516, PMID: 32043354.

(31) Wang, Y.; Vogel, A.; Sachs, M.; Sprick, R. S.; Wilbraham, L.; Moniz, S. J. A.; Godin, R.; Zwijnenburg, M. A.; Durrant, J. R.; Cooper, A. I.; Tang, J. Current understanding and challenges of solar-driven hydrogen generation using polymeric photocatalysts. *Nature Energy* **2019**, *4*, 746–760.

(32) *Introduction to Reticular Chemistry*; John Wiley & Sons, Ltd, 2019; Chapter 8, pp 197–223.

(33) Stegbauer, L.; Zech, S.; Savasci, G.; Banerjee, T.; Podjaski, F.; Schwinghammer, K.; Ochsenfeld, C.; Lotsch, B. V. Tailor-Made Photoconductive Pyrene-Based Covalent Organic Frameworks for Visible-Light Driven Hydrogen Generation. *Advanced Energy Materials* **2018**, *8*, 1703278.

(34) Mohammed, O. F.; Dreyer, J.; Magnes, B.-Z.; Pines, E.; Nibbering, E. T. J. Solvent-Dependent Photoacidity State of Pyranine Monitored by Transient Mid-Infrared Spectroscopy. *ChemPhysChem* **2005**, *6*, 625–636.

(35) Kinik, F. P.; Ortega-Guerrero, A.; Ongari, D.; Ireland, C. P.; Smit, B. Pyrene-based metal organic frameworks: from synthesis to applications. *Chemical Society Reviews* **2021**, *50*, 3143–3177.

(36) Dumur, F. Recent advances on visible light Triphenylamine-based photoinitiators of polymerization. *European Polymer Journal* **2022**, *166*, 111036.

(37) Tiwari, A.; Pal, U. Effect of donor-donor-π-acceptor architecture of triphenylamine-based organic sensitizers over TiO_2 photocatalysts for visible-light-driven hydrogen production. *International Journal of Hydrogen Energy* **2015**, *40*, 9069–9079.

(38) Zhang, X.; Jin, R. Rational Design of Low-Band Gap Star-Shaped Molecules With 2,4,6-Triphenyl-1,3,5-triazine as Core and Diketopyrrolopyrrole Derivatives as Arms for Organic Solar Cells Applications. *Frontiers in Chemistry* **2019**, *7*.

(39) Fateeva, A.; Chater, P. A.; Ireland, C. P.; Tahir, A. A.; Khimyak, Y. Z.; Wiper, P. V.; Darwent, J. R.; Rosseinsky, M. J. A Water-Stable Porphyrin-Based Metal-Organic Framework Active for Visible-Light Photocatalysis. *Angewandte Chemie International Edition* **2012**, *51*, 7440–7444.

(40) Dai, C.; He, T.; Zhong, L.; Liu, X.; Zhen, W.; Xue, C.; Li, S.; Jiang, D.; Liu, B. 2,4,6-Triphenyl-1,3,5-Triazine Based Covalent Organic Frameworks for Photoelectrochemical H_2 Evolution. *Advanced Materials Interfaces* **2021**, *8*, 2002191.

(41) Luo, B.; Chen, Y.; Zhang, Y.; Huo, J. Benzotrithiophene and triphenylamine based covalent organic frameworks as heterogeneous photocatalysts for benzimidazole synthesis. *Journal of Catalysis* **2021**, *402*, 52–60.

(42) Ng, K. H.; Lai, S. Y.; Cheng, C. K.; Cheng, Y. W.; Chong, C. C. Photocatalytic water splitting for solving energy crisis: Myth, Fact or Busted? *Chemical Engineering Journal* **2021**, *417*, 128847.

(43) Wan, S.; Gandara, F.; Asano, A.; Furukawa, H.; Saeki, A.; Dey, S. K.; Liao, L.; Ambrogio, M. W.; Botros, Y. Y.; Duan, X.; Seki, S.; Stoddart, J. F.; Yaghi, O. M. Covalent Organic Frameworks with High Charge Carrier Mobility. *Chemistry of Materials* **2011**, *23*, 4094–4097.

(44) Ma, S.; Deng, T.; Li, Z.; Zhang, Z.; Jia, J.; Li, Q.; Wu, G.; Xia, H.; Yang, S.-W.; Liu, X. Photocatalytic Hydrogen Production on a sp^2-Carbon-Linked Covalent Organic Framework. *Angewandte Chemie* **2022**, *134*.

(45) Huang, N.; Zhai, L.; Coupry, D. E.; Addicoat, M. A.; Okushita, K.; Nishimura, K.; Heine, T.; Jiang, D. Multiple-component covalent organic frameworks. *Nat. Commun.* **2016**, *7*, 12325.

(46) Kailasam, K.; Mesch, M. B.; Möhlmann, L.; Baar, M.; Blechert, S.; Schwarze, M.; Schröder, M.; Schomäcker, R.; Senker, J.; Thomas, A. Donor-Acceptor-Type Heptazine-Based Polymer Networks for Photocatalytic Hydrogen Evolution. *Energy Technology* **2016**, *4*, 744–750.

(47) Zhao, S.; Dong, B.; Ge, R.; Wang, C.; Song, X.; Ma, W.; Wang, Y.; Hao, C.; Guo, X.; Gao, Y. Channel-wall functionalization in covalent organic frameworks for the enhancement of CO_2 uptake and CO_2/N_2 selectivity. *RSC Advances* **2016**, *6*, 38774–38781.

(48) van der Jagt, R.; Vasileiadis, A.; Veldhuizen, H.; Shao, P.; Feng, X.; Ganapathy, S.; Habisreutinger, N. C.; van der Veen, M. A.; Wang, C.; Wagemaker, M.; van der Zwaag, S.; Nagai, A. Synthesis and Structure–Property Relationships of Polyimide Covalent Organic Frameworks for Carbon Dioxide Capture and (Aqueous) Sodium-Ion Batteries. *Chemistry of Materials* **2021**, *33*, 818–833.

(49) Wang, J.-C.; Kan, X.; Shang, J.-Y.; Qiao, H.; Dong, Y.-B. Catalytic asymmetric synthesis of chiral covalent organic frameworks from prochiral monomers for heterogeneous asymmetric catalysis. *J. Am. Chem. Soc.* **2020**, *142*, 16915–16920.

(50) Gropp, C.; Ma, T.; Hanikel, N.; Yaghi, O. M. Design of higher valency in covalent organic frameworks. *Science* **2020**, *370*, eabd6406.

(51) Wang, X.; Han, X.; Zhang, J.; Wu, X.; Liu, Y.; Cui, Y. Homochiral 2D porous covalent organic frameworks for heterogeneous asymmetric catalysis. *J. Am. Chem. Soc.* **2016**, *138*, 12332–12335.

(52) Zhao, C.; Lyu, H.; Ji, Z.; Zhu, C.; Yaghi, O. M. Ester-linked crystalline covalent organic frameworks. *J. Am. Chem. Soc.* **2020**, *142*, 14450–14454.

(53) Thomas, S.; Li, H.; Zhong, C.; Matsumoto, M.; Dichtel, W. R.; Bredas, J.-L. Electronic Structure of Two-Dimensional π-Conjugated Covalent Organic Frameworks. *Chemistry of Materials* **2019**, *31*, 3051–3065.

(54) Er, D.; Dong, L.; Shenoy, V. B. Mechanisms for Engineering Highly Anisotropic Conductivity in a Layered Covalent-Organic Framework. *The Journal of Physical Chemistry C* **2015**, *120*, 174–178.

(55) Wang, M. et al. Unveiling Electronic Properties in Metal-Phthalocyanine-Based Pyrazine-Linked Conjugated Two-Dimensional Covalent Organic Frameworks. *Journal of the American Chemical Society* **2019**, *141*, 16810–16816.

(56) Li, Y.; Liu, M.; Wu, J.; Li, J.; Yu, X.; Zhang, Q. Highly stable β-ketoenamine-based covalent organic frameworks (COFs): synthesis and optoelectrical applications. *Frontiers of Optoelectronics* **2022**, *15*.

(57) Kim, Y. H.; Kim, N.; Seo, J.-M.; Jeon, J.-P.; Noh, H.-J.; Kweon, D. H.; Ryu, J.; Baek, J.-B. Benzothiazole-Based Covalent Organic Frameworks with Different Symmetrical Combinations for Photocatalytic CO_2 Conversion. *Chemistry of Materials* **2021**, *33*, 8705–8711.

(58) Peng, L.; Chang, S.; Liu, Z.; Fu, Y.; Ma, R.; Lu, X.; Zhang, F.; Zhu, W.; Kong, L.; Fan, M. Visible-light-driven photocatalytic CO_2 reduction over ketoenamine-based covalent organic frameworks: role of the host functional groups. *Catalysis Science & Technology* **2021**, *11*, 1717–1724.

(59) Vyas, V. S.; Haase, F.; Stegbauer, L.; Savasci, G.; Podjaski, F.; Ochsenfeld, C.; Lotsch, B. V. A tunable azine covalent organic framework platform for visible light-induced hydrogen generation. *Nat. Commun.* **2015**, *6*, 8508.

(60) Zoubritzky, L.; Coudert, F.-X. CrystalNets.jl: Identification of Crystal Topologies. *SciPost Chemistry* **2022**, *1*.

(61) Stegbauer, L.; Schwinghammer, K.; Lotsch, B. V. A hydrazone-based covalent organic framework for photocatalytic hydrogen production. *Chem. Sci.* **2014**, *5*, 2789–2793.

(62) Yang, J. et al. Protonated Imine-Linked Covalent Organic Frameworks for Photocatalytic Hydrogen Evolution. *Angewandte Chemie International Edition* **2021**, *60*, 19797–19803.

(63) Hamad, S.; Hernandez, N. C.; Aziz, A.; Ruiz-Salvador, A. R.; Calero, S.; Grau-Crespo, R. Electronic structure of porphyrin-based metal–organic frameworks and their suitability for solar fuel production photocatalysis. *Journal of Materials Chemistry A* **2015**, *3*, 23458–23465.

(64) Huber, S. P. et al. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Scientific Data* **2020**, *7*.

(65) Uhrin, M.; Huber, S. P.; Yu, J.; Marzari, N.; Pizzi, G. Workflows in AiiDA: Engineering a high-throughput, event-based engine for robust and modular computational workflows. *Computational Materials Science* **2021**, *187*, 110086.

(66) Cube analysis. `https://github.com/kjappelbaum/cubes`, Accessed: 2022-09-17.

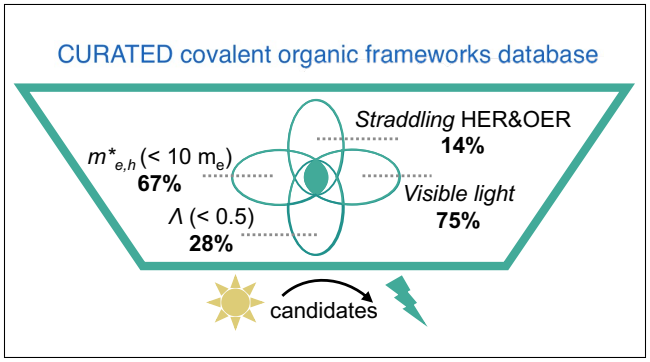
(67) Ganose, A. M.; Jackson, A. J.; Scanlon, D. O. sumo: Command-line tools for plotting and analysis of periodic *ab initio* calculations. *Journal of Open Source Software* **2018**, *3*, 717.

(68) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge structural database. *Acta Crystallogr. B: Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171–179.

(69) Ehrt, C.; Krause, B.; Schmidt, R.; Ehmki, E. S. R.; Rarey, M. SMARTS.plus – A Toolbox for Chemical Pattern Design. *Molecular Informatics* **2020**, *39*, 2000216.

(70) Ho, J.; Tumkaya, T.; Aryal, S.; Choi, H.; Claridge-Chang, A. Moving beyond P values: data analysis with estimation graphics. *Nature Methods* **2019**, *16*, 565–566.

TOC Graphic


CURATED covalent organic frameworks database
Straddling HER&OER
14%
m*e,h (< 10 me)
67%
Visible light
75%
Λ (< 0.5)
28%
candidates