Organic Photoredox Catalysts for CO$_2$ Reduction: Driving Discovery with Genetic Algorithms

Kareesa J. Kron,[1] Andres Rodriguez-Katakura,[1] Pranesh Regu,[1, 2] Maria N. Reed,[1] Rachelle Elhessen,[1] and Shaama Mallikarjun Sharada[1, 3]

[1)]*Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, California, 90089, United States*

[2)]*Alexion Pharmaceuticals, Los Angeles, California, United States*

[3)]*Department of Chemistry, University of Southern California, Los Angeles, California, 90089, United States*

(*Electronic mail: ssharada@usc.edu)

(Dated: 1 August 2022)

This work implements a genetic algorithm (GA) to discover organic catalysts for photoredox CO$_2$ reduction that are both highly active and resistant to degradation. The LUMO energy of the ground state catalyst is chosen as the activity descriptor and average Mulliken charge on all ring carbons as the descriptor for resistance to degradation via carboxylation (both obtained using density functional theory), to construct the fitness function of the GA. We combine the results of multiple GA runs, each based on different relative weighting of the two descriptors, and rigorously assess GA performance by calculating electron transfer barriers to CO$_2$ reduction. A large majority of GA predictions exhibit improved performance relative to experimentally studied *o*-, *m*-, and *p*-terphenyl catalysts. Based on stringent cut-offs imposed on the average charge, barrier to electron transfer to CO$_2$, and excitation energy, we recommend 25 catalysts for further experimental investigation of viability towards photoredox CO$_2$ reduction.

I. INTRODUCTION

The primary objective of this work is to establish an automated discovery protocol for viable molecular organic photoredox catalysts to perform single-electron reduction of CO$_2$. Green, photoredox routes for CO$_2$ reduction using phenylene oligomers were first developed in the 1990's.[1,2] Growing emphasis on sustainable chemical transformations in recent years has renewed both general interest in organic photoredox chemistry[3,4] and synthesis efforts to reduce CO$_2$ to make amino acids and hydrocarboxylation products.[5,6] Organic chromophores are very attractive as they can catalyze thermally challenging reactions via the generation of highly reactive radical species upon excitation and quenching.

Figure 1 describes the catalytic cycle for CO$_2$ reduction with phenylene oligomers, including proposed pathways that lead to catalyst degradation. For this reaction, Matsuoka and coworkers found that *p*-terphenyl produced the most formate among *o*-, *m*-, and *p*- isomers.[2] Higher activity of *p*-terphenyl is counter-intuitive as the free energy driving force for CO$_2$ reduction is more favorable for *o*-, and *m*- isomers relative to the *p*- isomer. Carboxylation products are observed for *o*- and *m*-isomers, indicating that this reaction is preferred to electron transfer (ET).[2] Furthermore, all isomers exhibit single-digit turnover numbers. This is believed to be caused by proton capture and subsequent Birch reduction of the catalysts.[7]

Based on higher activity of the *p*-terphenyl isomer observed in this study, our group carried out an analysis of substituent electrophilicity effects on the kinetics of ET from the *p*-terphenyl radical anion to CO$_2$.[9] The ET process is adiabatic and electron-donating groups (indicated by negative Hammett parameter,[10] σ_p), symmetrically substituted at both *p*- ends of *p*-terphenyl lead to lower ET barriers and consequently higher rate coefficients than electron-withdrawing groups. We
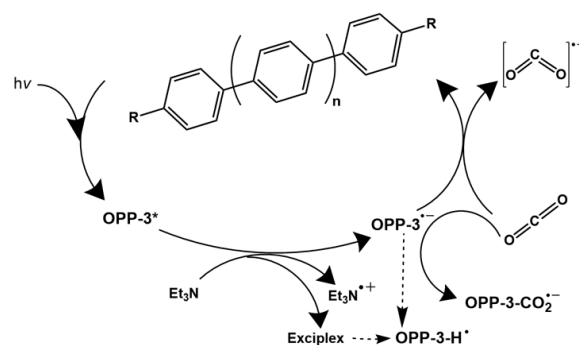

FIG. 1. Proposed catalytic cycle of single electron reduction of CO$_2$ with oligo(p-phenylene), or OPP.[8,9] OPP-3 is terphenyl with n = 1 and R = H. Reproduced with permission from J. Phys. Chem. A. 124, 5359–5368 (2020).[9] Copyright 2020 American Chemical Society.

also determined that the lowest unoccupied molecular orbital (LUMO) of the ground state catalyst is a reliable descriptor for the reaction free energy of ET. LUMO energies drop steeply with increasing electron-withdrawing character, mirrored in the sharp increase in free energy. With increasing electron-donating character, however, while LUMO energies exhibit a modest increase with a corresponding decrease in free energies, both quantities plateau beyond a certain value of σ_p. Our prior work therefore demonstrates that free energies cannot be lowered indefinitely by enhancing electron-donating character alone.

The experimental and computational studies together demonstrate that the core challenge facing catalyst design for CO$_2$ reduction is that it must not only be active but also exhibit high turnover numbers by being resistant to degradation via carboxylation and Birch reduction. While such a task can seem daunting given the sheer vastness of chemical space,

one can leverage emerging approaches that automate screening and search to drive materials discovery. For example, machine learning methods trained on experimental and computational hydrogen evolution activity data for a database of organic photoredox catalysts have been utilized to screen materials libraries and identify new candidates.[11,12] Methods that drive discovery of novel organic molecules by utilizing common structural or latent features of known molecules are also being developed. However with some exceptions,[13] both evolutionary approaches such as genetic algorithms and generative models such as variational autoencoders have been restricted to non-catalysis applications such as drug development and organic light-emitting diodes.[14–16]

This study aims to develop a genetic algorithm (GA), constructed using insights from prior experimental and computational studies, to discover novel catalysts for photoredox CO_2 reduction. GA is an evolutionary approach that creates new 'offsprings' by combining or modifying their 'parents.'[17,18] In any step of the evolution (generation), the 'fitness' of individuals determines whether they constitute the parent pool for the subsequent generation. GAs have been successfully deployed across several domains ranging from art education[19] to cloud security[20]. This work adapts the open-source GA program developed by the Jensen group, which was originally employed to discover drug candidates based on lipophilicity, synthetic accessibility, and ring size characteristics.[14]

In this work, the fitness function is modified to favor candidates that possess high LUMO energy and more positive average carbon charge, chosen as descriptors for catalyst activity towards CO_2 reduction and and resistance to degradation via carboxylation, respectively. The relative weights of the two descriptors are systematically varied. The three isomers of terphenyl, both with and without electron-donating substituents, constitute the parent pool. GA outcomes are assessed using more rigorous calculations of catalyst excited states and ET to CO_2. Most catalysts predicted by the GA show improved performance compared to the terphenyl benchmarks. The GA run with fitness biased slightly in favor of LUMO energy (60% weight) yields the largest number of viable candidates relative to their parent pool. We identify at least 25 catalyst candidates, out of the 103 unique catalysts predicted by the GA, that are promising starting points for experimental studies of photoredox CO_2 reduction.

II. METHODS

Our goal is to utilize genetic algorithms to identify substituted oligophenylenes that are both active towards CO_2 reduction and resistant to degradation via carboxylation and Birch reduction. To rigorously assess the performance of the GA approach that relies on inexpensive, approximate descriptors, ET barriers are calculated for all the catalysts predicted by the GA. The activity and degradation resistance of the new candidates are benchmarked to experimentally studied *o-, m-,* and *p-* terphenyl isomers.[2]

A. Genetic algorithm

We adapt the string GA developed by Jensen (`github.com/jensengroup/String-GA`) for catalyst discovery.[14,21] The string GA was employed to identify drug-like molecules based on a fitness function that favors high lipophilicity and ease of synthesis and penalizes large ring sizes. The 'string' in the name refers to the use of the SMILES (Simplified Molecular Input Line Entry System)[22,23] string representation of molecules to carry out mutation and crossover operations. Our adaptation (`github.com/andresmrk/Phenyl_GA`) also employs the SMILES representation owing to its widespread use and ability to represent complex phenyl ring substitutions,[24–26] despite some recent concerns regarding its robustness.[27] The RDKit software package[28] is employed to convert between SMILES representation and Cartesian coordinates. We modify the fitness function and GA procedure from the original implementation to search for catalytically active and degradation-resistant oligophenylene derivatives for one-electron CO_2 reduction.

The initial population for every GA run consists of 10 molecules chosen at random from a pool of 30 (Table S1 of Supporting Information, SI). The group of 30 molecules includes the reference unsubstituted *p-*, *m-*, and *o-* terphenyl, as well as substituted isomers where a single H in an outer phenyl ring is replaced with one of -CH$_3$, -NH$_2$, or -OH groups, illustrated in Figure 2. These functional groups are chosen based on our prior finding that electron donating groups on the terminal *p-* position of the outer phenyl ring enhance ET rates relative to unsubstituted *p-*terphenyl.[9] After the initial population is randomly selected, the GA is allowed to propagate 10 generations. Every generation is created using crossover and mutation operations. The crossover operation 'mates' two parents to produce an offspring by exchanging 'genes.' The multi-point crossover operation is utilized to create one fragment (or 'gene') from every parent. Choosing points on the SMILES string at random for crossover can result in the creation of unrealistic or erroneous SMILES representations. Therefore, we employ constraints that preserve phenyl rings. In other words, the crossover operator allows only for the swapping of one or more phenyl groups (substituted/unsubstituted) between parent molecules. A selected
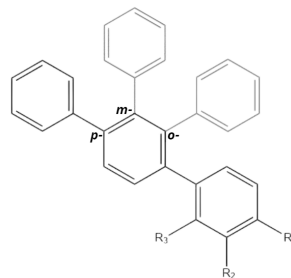

FIG. 2. Parent pool of terphenyl catalysts composed of substituted *p-*, *m-*, and *o-*terphenyl with a single substitution at one of three positions (R_1, R_2, R_3) with one of several electron-donating substituents or hydrogen (-H, -OH, -NH$_2$, CH$_3$)

gene from a parent molecule has an equal probability of replacing a gene in the receiving parent or being randomly appended to produce the offspring. Ten offspring are created at the end of every generation. After crossover, the mutation rate dictates the probability that a mutation will be applied to a randomly selected carbon atom in the offspring. A mutation rate of 0.1 is applied and the mutation is selected at random from a list of 15 substituents (Table S2, SI).

Once a new generation is created, all individuals compete according to a user-defined fitness function. We turn to prior experimental and computational studies of CO_2 reduction with phenylene catalysts to identify the most appropriate form of the fitness function. Our prior computational study on reduction of CO_2 by substituted p-terphenyls revealed that the catalyst's reduction potential, or free energy of the reaction has the largest impact on ET rates.[9] Further, free energy is more favorable for phenylenes with higher ground state LUMO energy. Therefore, the LUMO energy of the catalyst, obtained inexpensively (albeit approximately) from density functional theory (DFT) calculations, constitutes one component of the fitness function.

To assess degradation resistance via Birch reduction, we turned to work by Zimmerman and coworkers.[29,30] They show that for aromatic systems, the position that is most likely to be protonated is that with the highest charge density, or the most anionic carbon. However, as protonation can occur from many electronic states of the catalyst and the most favorable proton source is not yet known,[31] the choice of carbon charge as a descriptor for degradation via Birch reduction is unclear. To the best of our knowledge, studies similar to those by Zimmerman and coworkers for the likelihood or the position of attack of carboxylation have not yet been performed. As CO_2 is an electrophile, charge density may be the determining factor for carboxylation along similar lines to Birch reduction. For the experimentally studied o-, m-, and p-terphenyl isomers therefore, we calculate transition structures and barriers to carboxylation at various ring positions (Figure S1, SI). Described in Section 3 of the SI, these calculations show that the barriers at various positions are not directly correlated to the charge on the respective carbon for either the anionic or neutral ground state of the catalyst (Figure S2, SI). On the other hand, both average and minimum carboxylation barriers for isomers are strongly, linearly correlated to the respective average charge on the ring carbons, both in the anionic and ground state of the catalyst (Figure S3, SI). Furthermore, charges on ring carbons in the anionic state of the catalyst are directly correlated (slope $\approx$ 1) with the corresponding charges in the ground state (Figure S4, SI), consistent with our prior work that shows that trends in anion properties closely follow those of the ground states of phenylenes.[9] Therefore, the average carbon charge of the ground state catalyst is used as an approximate measure of likelihood of degradation via carboxylation, with more positive charge signifying higher resistance to degradation.

Lastly, to limit the selection of catalysts to those with optical gaps close to phenylene isomers,[32–34] we add a simple constraint to the fitness function that the HOMO-LUMO energy gap lies between 200 and 400 nm. Unlike prior GA studies that inspired this work,[14,35] there is no need to impose a penalty for large ring sizes due to the constraint imposed on the crossover operation. There are also no constraints necessary on overall system size as very large oligophenylenes are expected to have lower fitness scores than short-chain molecules on account of lower LUMO energies. We verify this by showing that system sizes of the candidates identified by the GA are not arbitrarily large. There is also no noticeable correlation between ET parameters and system size (Figure S5, SI).

Normalized fitness scores are calculated for LUMO energy ($s_L(i)$) and average carbon charge ($s_C(i)$) of candidate i as follows:

$$s_L(i) = \left(\frac{LUMO(i)}{\sum_n LUMO(n)} \right); \quad s_C(i) = \left(\frac{\overline{C(i)}}{\sum_n \overline{C(n)}} \right) \quad (1)$$

where $\overline{C(i)}$ is the average Mulliken ring carbon charge for candidate i. The gap contribution ($s_{G,i}$) is assigned as 0 or -1 depending on whether it lies within or beyond the 200-400 nm region, respectively.

$$s_G(i) = \begin{cases} 0 & \text{if 200 nm} \leq \text{(LUMO-HOMO)} \leq 400 \text{ nm} \\ -1 & \text{otherwise} \end{cases}$$

$$(2)$$

However, DFT typically underestimates HOMO-LUMO gaps.[36] To assess the impact of applying an approximate orbital gap constraint using ground-state DFT, we calculate vertical excitation energies for all catalysts predicted by the GA using time-dependent DFT (TDDFT)[37] with the same level of theory as ET calculations. Excitation energies confirm that the approximated gap contribution to the fitness function ensures all GA predictions possess excitation energies that lie in nearly the same range as the reference p-, m-, and o-terphenyl and the parent pool (Tables S3 and S4, SI).

The fitness function, $f(i)$ is given by:

$$f(i) = \omega s_L(i) + (1 - \omega)s_C(i) + s_G(i) \quad (3)$$

The weight, ω, specifies the influence of each descriptor score (LUMO or average charge) on the overall fitness score, with zero factoring only the charge, 0.5 assigning equal weighting to charge and LUMO, and unity factoring only LUMO. As the optimal value is not known a priori, we vary ω from 0 to 1 in steps of 0.1 and execute 11 GA runs.

During the GA runs, LUMO energies and charges are calculated in the gas phase using the *ab initio* quantum chemistry software, Q-Chem (version 5.4.2)[38] at the B3LYP/def2-SVP level of theory.[39–41] Ten fittest candidates in every generation are passed to the next generation. From a randomly chosen initial population, every GA run evolves the catalysts over 10 generations. At the end of 10 generations, the top 10 fittest structures are selected. ET rates and excitation energies are calculated for these structures (procedure described below) and contrasted with the phenylene oligomers examined in prior studies.[2] Although corresponding studies of degradation pathways are also necessary, they entail computationally intensive calculations of transition structures for carboxylation (and protonation) across multiple catalysts. This is a topic

of ongoing and future work in our group. For the purpose of assessing GA performance, we continue to use average carbon charge $\overline{C}$ as a descriptor of susceptibility to degradation.

B. Electron transfer barriers

GA predictions are rigorously analyzed by quantifying catalyst activity towards CO_2 reduction. The computational procedure is very similar to our previous work.[9] We use constrained density functional theory (CDFT)[42–44] to calculate initial and final states for ET where the negative charge is localized on the catalyst and substrate fragment, respectively. The guess reactant (initial state) structures for optimization are created by placing linear CO_2 slightly above and near the center of every predicted catalyst. Our prior work shows that upon optimization, CO_2 relaxes closer towards the center of the catalyst for most substituted p-terphenyls even if the guess structure is created with CO_2 off-center. Product (or final) states are calculated with CDFT using the optimized reactant geometries as guesses and by switching charges between fragments. In some cases, the CO_2 fragment is manually bent to accelerate convergence to the final state. While it is possible that there are multiple local minima corresponding any reactant or product state, our prior work predicts similar ET parameters across at least some of these local minima. Therefore, we employ a single representative geometry for each of the reactant and product states for a given catalyst. The state-specific model for non-equilibrium solvation is employed to calculate vertically excited charge transfer states.[45–49] We find that the donor-acceptor pairs are electronically strongly coupled for most predicted catalysts, based on values of the diabatic coupling constant calculated with CDFT configurations interaction (CDFT-CI) (Table S5, SI).[50] Therefore, the adiabatic expression[51] for rate coefficients can be employed:

$$k = \nu e^{-\frac{\Delta G^*}{RT}} \tag{4}$$

where ν is a weighted average of vibrational frequencies for modes that contribute to the reaction coordinate ($\approx 10^{13}$ s^{-1})[51] and R is the molar equivalent of Boltzmann's constant, k_B. ΔG^* is the effective barrier to ET obtained from Marcus theory:[52]

$$\Delta G^* = \frac{\lambda}{4}\left(1 + \frac{\Delta G}{\lambda}\right)^2 \tag{5}$$

where λ is the reorganization energy and ΔG is the free energy change for the ET step. All simulations are carried out using Q-Chem at the ωB97X-D/def2-TZVP level of theory.[53,54] The conductor-like polarizable continuum model (C-PCM) for implicit solvation[55–57] is used for all electronic structure calculations with dichloromethane (dielectric = 8.93, optical dielectric = 2.02834) as the solvent. Vertical excitation energies of catalysts are calculated using TDDFT at the same level of theory with non-equilibrium solvation.

III. RESULTS AND DISCUSSION

A. GA predictions

Eleven GA runs generate 103 unique catalysts (with 7 duplicates) that vary widely in size and structure in comparison to the parent population (Figures S6-S7, SI). The smallest are biphenyls while the largest contain seven phenyl rings, composed of various o-, p-, and m- terphenyl sub-structures. Quaterphenyls (4 phenyl rings) are observed for every LUMO weight, and terphenyls are observed in all but the case where $\omega = 1$. The smallest predicted catalyst contains 23 atoms while the largest contains 82 atoms. Two largest structures include o-, m-, and p-terphenyl as sub-structures in the arrangement of the seven phenyl rings. The p- sub-structure is found in only 7 of the 103 catalysts, while o- and m- are observed in approximately 80 and 65 catalysts respectively, with some catalysts containing multiple o- or m- terphenyl sub-structures. Nine substituents are found in the GA catalysts – -OH: 97 catalysts, -NH$_2$: 19, -OCH$_3$: 11, -NHCH$_3$: 11, and -CH$_3$, -SCH$_3$, -NO$_2$, -F, and -N(CH$_3$)$_2$ each occurring in < 10 catalysts. Hydroxyl groups are by far the most common. Other substituents are only represented in specific weights; for example, all catalysts for LUMO weight of 50% ($\omega = 0.5$) include OCH$_3$. Only one other OCH$_3$ is found in a catalyst predicted in the 20% LUMO weighted ($\omega = 0.2$) GA run.

Figure 3 demonstrates that the GA discovery procedure successfully identifies a diverse range of structures by seeking to increase both the average carbon charge ($\overline{C}$) and LUMO energy. As is desired, most of the predicted catalysts (green circles) are located to the left of and higher than the parent population (black squares). Vertical and horizontal lines represent the reference unsubstituted o-, m-, and p-terphenyl systems. p-Terphenyl has the most positive $\overline{C}$ of the three isomers, in agreement with its lower susceptibility to degradation via carboxylation in experiments.[2] As a result, despite its lower LUMO energy (higher ET barrier), it is the isomer that exhibits highest activity towards CO_2 reduction via ET. Most predicted catalysts are above and to the left of p-terphenyl, suggesting that they may exhibit improved performance relative to the experimentally studied system. To assess whether the LUMO energy is indeed a reliable descriptor of catalytic activity, we carry out more computationally intensive ET calculations, described below.

B. Impact of LUMO weight on catalyst performance

Reaction energy (ΔG), reorganization energy (λ), and the resulting free energy barrier (ΔG^*, Equation 5) are calculated for all catalysts predicted by the GA. Figure 4 depicts the variation of these quantities and $\overline{C}$ with varying LUMO weighting parameter (ω) in the fitness function. The corresponding data is presented in (Table S6, SI). Four predicted catalysts for $\omega = 0$ yield ΔG^* values higher than 120 kJ/mol and are excluded from this and all other figures to enhance readability. High ΔG^* values are the likely outcome of GA search based
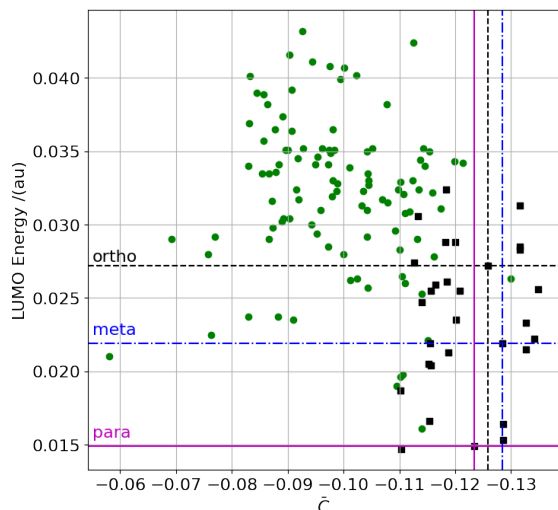
FIG. 3. GA Predictions – LUMO energies /(a.u.) versus average carbon charges ($\overline{C}$, Mulliken charges) for both the parent pool (black squares) and the predicted catalysts (green circles) with vertical and horizontal lines representing the LUMO energies and $\overline{C}$ of *p*- (magenta), *m*- (blue), and *o*-terphenyl (black). All values are calculated for optimized ground state catalysts at the ωB97X-D/def2-TZVP level of theory with solvation in dichloromethane.

exclusively on charge and no weight assigned to LUMO energies. In other words, even though the predicted catalysts are expected to be resistant to degradation, they exhibit no activity towards CO_2 reduction.

At the other end of the fitness spectrum where only LUMO is weighted ($\omega = 1$), $\overline{C}$ values are more negative than those for most intermediate values of ω. We note however that these observations are purely qualitative. Quantitative comparisons are precluded by random processes inherent to the GA, in addition to the fact that the initial population is distinct for every ω. Trends in barriers ΔG^* closely follow those for ΔG as reorganization energies lie within a narrow range for most systems (with only a few exceptions for ω of 0.4, 0.7, 0.8 and 0.9). This indicates that in most cases, ΔG calculations are sufficient to predict trends in ET rate coefficients for CO_2 reduction with these catalysts. In addition, Figure 5 demonstrates that LUMO energies serve as reliable descriptors for reaction free energies and therefore ET barriers. Although the degree of correlation between ΔG^* and LUMO energy is slightly lower (R^2=0.64) than what is normally desirable, the additional scatter compared to ΔG versus LUMO energy (R^2 = 0.73) likely originates in the incorporation of λ (Equation 5). Nevertheless, pursuing higher LUMO energies via the fitness function yields systems that possess lower ET barriers. Most predicted catalysts (purple circles) yield lower ET barriers to CO_2 reduction relative to both the parent population (black squares) as well as *p*-terphenyl.

Figure 6 depicts ΔG^* vs. $\overline{C}$ for each weight of LUMO, with labels that represent the ranking of that catalyst within its weight pool. The horizontal blue line on each figure reflects the average barriers of the initial pool fed into that GA run,

while the vertical blue line reflects the average carbon charge of the initial pool. Overall, around 90 candidates generated by the GA outperform their respective initial populations. In all cases including the one in which charge is not taken into account in the fitness ($\omega = 1$), the GA predicts catalysts with more positive $\overline{C}$. Most catalysts also yield lower ET barriers as well, although weights of less than 50 percent LUMO ($\omega < 0.5$) often yield higher ΔG^*'s relative to the initial pool. All the candidates generated in GA runs with ω values of 0.6 and 0.7, where the fitness is slightly more biased in favor of LUMO energy, possess more positive $\overline{C}$ and and lower barriers than their respective initial pools. LUMO weights of 0.2 and 0.3 yield many catalysts that exhibit poor activity towards CO_2 reduction and are possibly more susceptible to degradation compared to other GA runs. For these two runs in particular, the initial pools have higher average barriers and lower $\overline{C}$, which may be partially responsible for the poor outcomes. As structural diversity of the initial pool also affects the quality of GA predictions, we analyze counts of isomers, functional groups, and positions of attachment in initial pools (Figures S9-S11 and Table S8, SI). For instance, 6 out of 10 parents have a single substituent (-OH) in the parent pool of $\omega = 0.3$, unlike the remaining pools where the distributions are less skewed towards a single substituent. Further exploration of the dependence of predicted structural diversity on parent population is a topic for future work as it will require multiple GA runs for a given value of ω.

In the next step, we contrast predicted catalysts with those for which experimental studies are available, specifically *o*-, *m*-, and *p*-terphenyl. Figure 7 depicts ΔG^* vs. $\overline{C}$ for the initial pool as well as the GA-generated catalysts. Vertical and horizontal lines are used to represent the barriers and average carbon charge of *p*-, *m*-, and *o*- terphenyl, which are labeled accordingly. The overarching objective of this work is to discover catalysts that yield lower barriers as well as higher (more positive) average carbon charges than *p*-terphenyl, which is known to to be active for CO_2 reduction and is more resistant to degradation than its *o*- or *m*- isomers. Therefore, all points below and to the left of the magenta lines representing *p*-terphenyl in Figure 7 constitute improvements upon known experimental successes. The majority of the predicted catalysts are therefore viable based on the two criteria employed in this study.

C. Selection of viable catalysts

We select a subset of GA candidates based on stricter thresholds. Represented by the green, shaded region in Figure 7, we identify 38 candidates that have both $\overline{C} \geq -0.10$ and $\Delta G^* \leq 90$ kJ/mol. Of the 38, 25 catalysts possess vertical excitation energies (calculated using TDDFT) that lie between the calculated excitation energies of unsubstituted *p*- and *o*-terphenyl (3.56-3.86 eV). Out of the 103 catalysts predicted by the GA, 47 meet the excitation energy criterion, with scope for future improvement by replacing the current limits for determining the gap contribution to the fitness function ($s_g(i)$) with those determined using DFT. Constraining the excitation
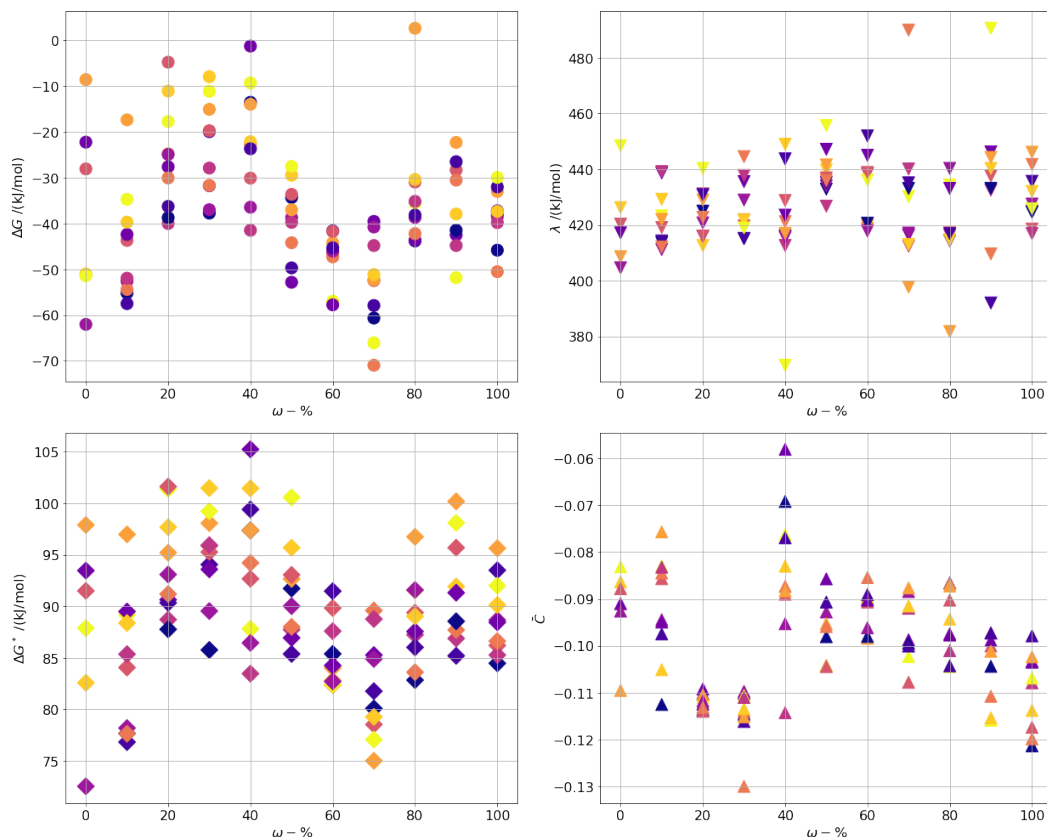
FIG. 4. Impact of choice of LUMO weight, ω – Reaction free energy for ET (ΔG /(kJ/mol)), reorganization energy (λ /(kJ/mol)), and ET barrier (ΔG^* /(kJ/mol)) as well as average carbon charge ($\overline{C}$). The color gradient within weight groups represents the rank of that catalyst (yellow: 1 to purple: 10).
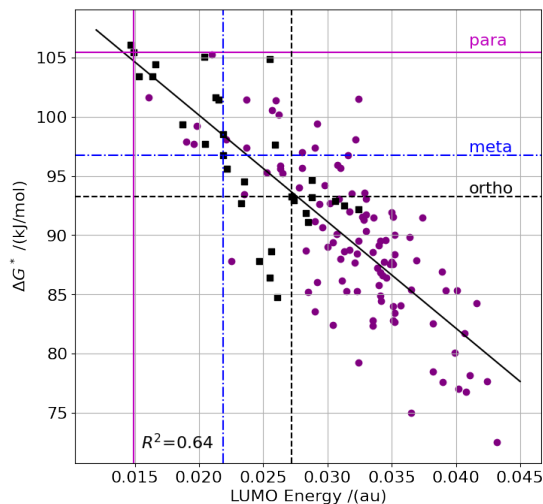

FIG. 5. ET barriers (ΔG^*) vs LUMO energy – for parent pool (black squares) and predicted catalysts (purple circles), with vertical and horizontal lines representing the LUMO energies and $\overline{C}$ of p- (magenta), m- (blue), and o-terphenyl (black). The black solid line represents a linear fit to the data.

energy to this region helps us ensure that excitations of the predicted catalysts are achievable with similar experimental set-ups, and the candidates possess similar optical properties to known active catalysts.

Viable catalysts identified based on our prescribed cut-off values of ET barriers, $\overline{C}$, and excitation energies are shown in Figure 8. Figure S12 of the SI groups subsets of these catalysts by similarity in their phenylene backbone to illustrate structural features common across viable catalysts. The 26 structures in Figure 8 represent 25 unique catalysts spanning several LUMO weights. The repeated structure is o-terphenyl that is doubly substituted with -OH on the p-terminal positions of the phenyl rings and emerges as a viable candidate in GA runs with 60% and 80% LUMO weighting. The GA with 60% weighting of LUMO energy contributes the largest pool of catalysts to the best 25 catalysts with 9 of its 10 structures fulfilling all criteria. GA runs with 70% and 80% weights introduce electron-donating NH_2 groups and also contribute significantly to the final set of viable catalysts. GA runs with LUMO weights higher than 80% do not yield viable candidates. GA runs with weights lower than 50% yield a handful of viable catalysts that are structurally similar to those generated with 60% weighting. Our recommended weight for future discovery studies with this GA framework is therefore 60% in favor of LUMO energy.
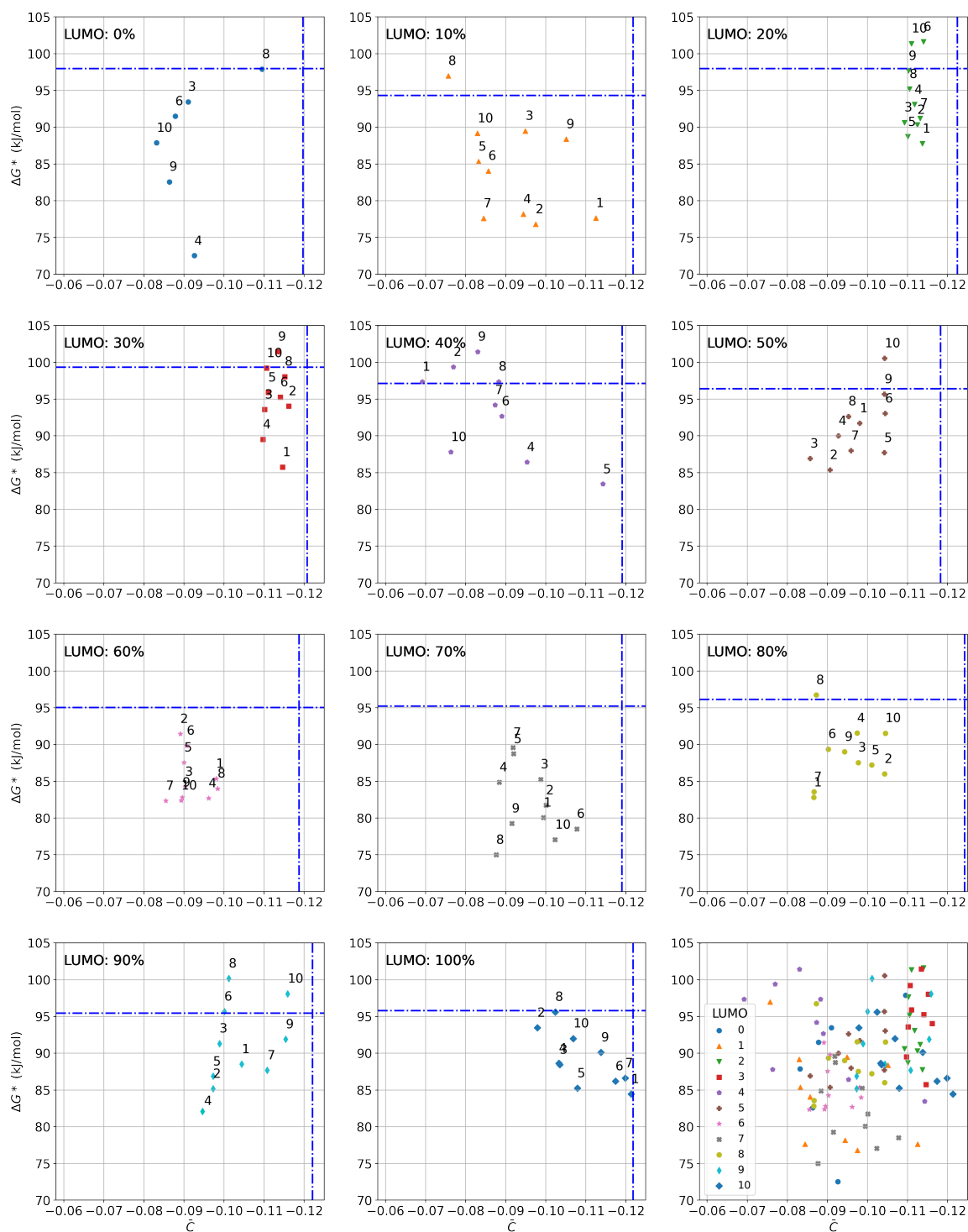
FIG. 6. Impact of ω and initial population – ET barrier (ΔG^*) vs. average carbon charge ($\overline{C}$), with labels representing catalyst rank in its respective GA run. Blue horizontal and vertical lines represent the average ET barrier and average carbon charge for the parent population that was randomly selected for that GA run.
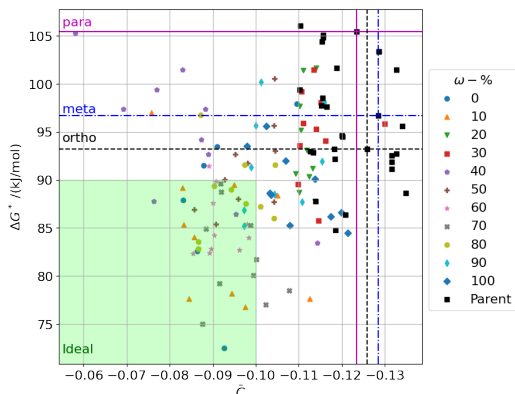
FIG. 7. Identifying candidates for experiment – ET barrier vs average carbon charge for GA predictions and parent catalysts. Vertical and horizontal lines represent the barrier and average Mulliken charge of *p*- (magenta), *m*- (blue), and *o*-terphenyl (black).

The catalysts vary in length from terphenyl to oligophenylenes with 7 phenyl rings. Our largest predicted catalyst is shown in the middle right of the 'LUMO: 70%' box in Figure 8. This catalyst contains every reference isomer sub-structure as well as multiple NH_2 and OH substitutions, and has the second-lowest barrier in this set of 25 catalysts. In the same way that *o*- sub-structures are the most common in the set of 103 predicted catalysts, *o*-sub-structures are found in 23 of the 25 catalyst structures. The *o*-terphenyl sub-structure substituted with at least one -OH yields three unique catalysts found with 3 different LUMO weights ranging from 10% to 80%.

Quaterphenyls represent the highest fraction of the best 25 catalysts and are found across the span of LUMO weights. Almost every structural isomer of quaterphenyl is represented in these 25 catalysts and most are found at least twice, with the sole exception of the *p*-terphenyl substructure. As quaterphenyls with *p*-sub-structures yield ET barriers higher than 90 kJ/mol, they do not meet our viability requirements. The quaterphenyls consisting of *p*-substituted OH-groups illustrated in Figure 8 are expected to be easier to synthesize as it often more straightforward to target the same position on several phenyl groups than it is target different positions of attack on neighboring phenyl groups, lowering the barrier to synthesis and testing of catalyst performance with experiments. As selective synthesis of specific structural isomers can be challenging, the fact that GA predicts simple structural isomers with comparable performance may ease synthesis efforts. An excellent instance of this is the set of viable candidates obtained with 60% LUMO weight, which are diverse (range of chain lengths) yet consist of similar features (substructures, substituents) to facilitate synthesis and testing. Although 70% LUMO weighting also contributes significantly to the pool of ideal catalysts, the candidates are larger and more complex. In addition to difficulties with synthesis, these candi-

dates may also be less soluble as is observed in the case of longer oligo(*p*-phenylenes) in experiments.[2]

Overall, our GA protocol enables the discovery of several novel catalyst candidates for photoredox CO_2 reduction by using simple DFT-based descriptors for activity (LUMO energy), degradation resistance (Mulliken charge), and excitation energy (HOMO-LUMO gap). Going forward, one way to enhance the GA's ability to predict synthetically viable candidates is to include a synthetic accessibility score in the fitness function similar to prior GA studies.[14] Other possibilities include the use of emerging machine learning approaches that predict the synthetic complexity (number of steps, costs of reactants, and so on) of organic molecules.[58–60] Explicitly considering ease of synthesis will accelerate the selection of viable catalysts. To identify truly sustainable pathways for CO_2 utilization, it is also important to incorporate parameters which reflect the environmental costs of catalyst synthesis routes as well as downstream processes to recycle or dispose of the proposed catalysts. Recent studies assessing the greenness of various chemical reactions have identified parameters such as greenhouse gas emissions, environmental/human toxicity, and atom efficiency, that must be considered while designing catalytic reaction pathways.[61–63]

One of the driving hypotheses in this study that is yet to be rigorously tested is the appropriateness of $\bar{C}$ obtained from Mulliken analysis as a descriptor for degradation resistance. In addition to examining substituent effects on carboxylation barriers, our group aims to examine pathways for proton transfer and Birch reduction both from the anion radical state of the catalyst as well as the excited state when complexed with the sacrificial electron donor. As our mechanistic and physical understanding of these catalysts improve, we can develop more robust discovery frameworks for viable photoredox catalysts based on the descriptors that emerge from such studies.

IV. CONCLUSIONS

We present a genetic algorithm for discovery of catalysts for photoredox CO_2 reduction that are both active and resistant to degradation via carboxylation. The fitness function favors candidates with higher LUMO energies and more positive average carbon charges, which represent descriptors for activity and degradation resistance, respectively. By varying relative weights on each descriptor in the fitness function, we find that biasing the function slightly in favor of LUMO energy (60%) yields the maximum number of candidates that satisfy our viability criteria. The GA successfully predicts several new structures, a majority of which promise substantial performance enhancements over previously studied *o*-, *m*-, and *p*-terphenyl isomers. By imposing more restrictive requirements on activity, degradation resistance, and excitation energy, we recommend 25 catalysts out of 103 GA predictions as potential first candidates for experimental studies of photoredox CO_2 reduction. Going forward, we aim to enhance GA robustness by both testing the reliability of average charge as descriptor for degradation resistance as well as predicting catalysts that are synthetically viable and sustainable by identifying appro-
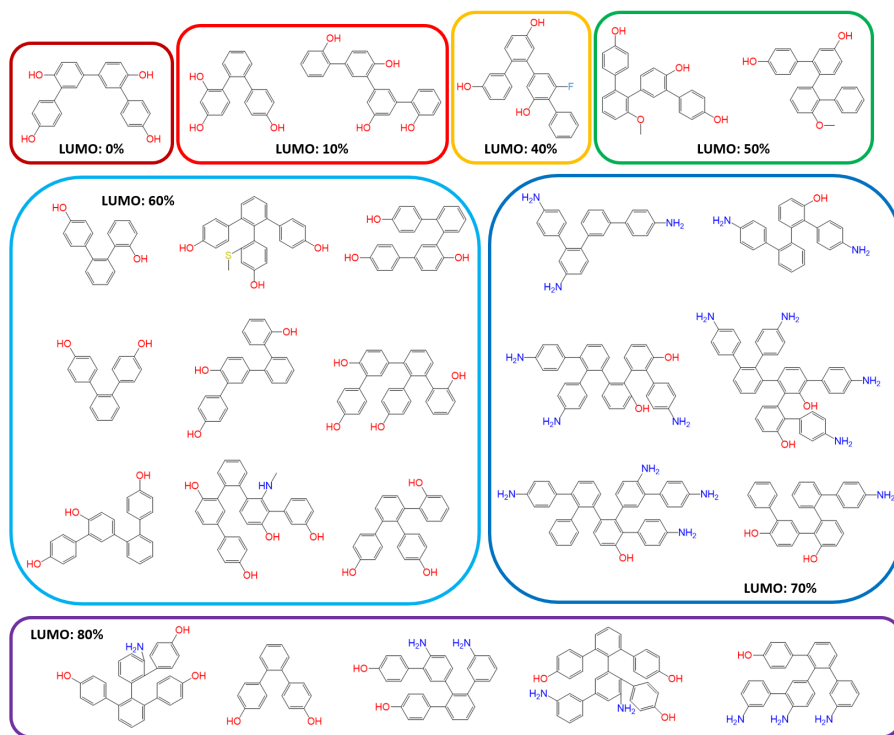
FIG. 8. Catalysts predicted by the GA with $\Delta G^* \leq 90$ kJ/mol, $\overline{C} \geq$ -0.10, and excitation energy between p- and o-terphenyl (3.56-3.86 eV), grouped by LUMO weight.

priate additional parameters for the fitness function.

V. SUPPLEMENTARY MATERIAL

The supplementary material consists of SMILES notations of parent and mutation candidates for the GA and examination of choice of descriptor for degradation using barriers to carboxylation. The material also summarizes key properties of candidates predicted by the GA – geometries, diabatic coupling constants, TDDFT excitation energies, and ET rate coefficients. Cartesian coordinates of all optimized structures are provided.

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available within the article and its supplementary material.

[1] S. Matsuoka, T. Kohzuki, C. Pac, and S. Yanagida, "Photochemical reduction of carbon dioxide to formate catalyzed by p-terphenyl in aprotic polar solvent," Chemistry Letters **19**, 2047–2048 (1990).

[2] S. Matsuoka, T. Kohzuki, C. Pac, A. Ishida, S. Takamuku, M. Kusaba, N. Nakashima, and S. Yanagida, "Photocatalysis of oligo (p-phenylenes): photochemical reduction of carbon dioxide with triethylamine," The Journal of Physical Chemistry **96**, 4437–4442 (1992).

[3] N. A. Romero and D. A. Nicewicz, "Organic photoredox catalysis," Chemical Reviews **116**, 10075–10166 (2016).

[4] Y. Du, R. M. Pearson, C.-H. Lim, S. M. Sartor, M. D. Ryan, H. Yang, N. H. Damrauer, and G. M. Miyake, "Strongly reducing, visible-light organic photoredox catalysts as sustainable alternatives to precious metals," Chemistry–A European Journal **23**, 10962–10968 (2017).

[5] H. Seo, A. Liu, and T. F. Jamison, "Direct β-selective hydrocarboxylation of styrenes with CO_2 enabled by continuous flow photoredox catalysis," Journal of the American Chemical Society **139**, 13969–13972 (2017).

[6] H. Seo, M. H. Katcher, and T. F. Jamison, "Photoredox activation of carbon dioxide for amino acid synthesis in continuous flow," Nature Chemistry **9**, 453 (2017).

[7] J. A. Barltrop, "The photoreduction of aromatic systems," Pure and Applied Chemistry **33**, 179–196 (1973).

[8] Y. Wada, T. Kitamura, and S. Yanagida, "CO_2-fixation into organic carbonyl compounds in visible-light-induced photocatalysis of linear aromatic compounds," Research on Chemical Intermediates **26**, 153–159 (2000).

[9] K. J. Kron, S. J. Gomez, Y. Mao, R. J. Cave, and S. Mallikarjun Sharada, "Computational Analysis of Electron Transfer Kinetics for CO_2 Reduction with Organic Photoredox Catalysts," Journal of Physical Chemistry A **124**, 5359–5368 (2020).

[10]L. P. Hammett, "The effect of structure upon the reactions of organic compounds. benzene derivatives," Journal of the American Chemical Society **59**, 96–103 (1937).

[11]Y. Bai, L. Wilbraham, B. J. Slater, M. A. Zwijnenburg, R. S. Sprick, and A. I. Cooper, "Accelerated discovery of organic polymer photocatalysts for hydrogen evolution from water through the integration of experiment and theory," Journal of the American Chemical Society **141**, 9063–9071 (2019).

[12]X. Li, P. M. Maffettone, Y. Che, T. Liu, L. Chen, and A. I. Cooper, "Combining machine learning and high-throughput experimentation to discover photocatalytically active organic molecules," Chemical Science **12**, 10742–10754 (2021).

[13]M. Foscato and V. R. Jensen, "Automated in silico design of homogeneous catalysts," ACS Catalysis **10**, 2354–2377 (2020).

[14]J. H. Jensen, "A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space," Chemical Science **10**, 3567–3572 (2019).

[15]S.-Y. Lu, S. Mukhopadhyay, R. Froese, and P. M. Zimmerman, "Virtual screening of hole transport, electron transport, and host layers for effective oled design," Journal of Chemical Information and Modeling **58**, 2440–2449 (2018).

[16]R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules," ACS Central Science **4**, 268–276 (2018).

[17]S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," Multimedia Tools and Applications **80**, 8091–8126 (2021).

[18]D. Whitley, "A genetic algorithm tutorial," Statistics and Computing **4**, 65–85 (1994).

[19]J. Liu, Q. Chen, and X. Tian, "Illustration design model with clustering optimization genetic algorithm," Complexity **2021**, 6668929 (2021).

[20]M. Tahir, M. Sardaraz, Z. Mehmood, and S. Muhammad, "CryptoGA: a cryptosystem based on genetic algorithm for cloud data security," Cluster Computing **24**, 739–752 (2021).

[21]E. S. Henault, M. H. Rasmussen, and J. H. Jensen, "Chemical space exploration: how genetic algorithms find the needle in the haystack," PeerJ Physical Chemistry **2**, e11 (2020).

[22]D. Weininger, "SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules," Journal of Chemical Information and Computer Sciences **28**, 31–36 (1988).

[23]D. Weininger, A. Weininger, and J. L. Weininger, "SMILES. 2. Algorithm for Generation of Unique SMILES Notation," Journal of Chemical Information and Computer Sciences **29**, 97–101 (1989).

[24]J. J. Irwin and B. K. Shoichet, "ZINC - A free database of commercially available compounds for virtual screening," Journal of Chemical Information and Modeling **45**, 177–182 (2005).

[25]J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," Journal of Molecular Graphics and Modelling **25**, 247–260 (2006).

[26]D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "DrugBank: A comprehensive resource for in silico drug discovery and exploration," Nucleic Acids Research **34**, D668–D672 (2006).

[27]M. Krenn, F. Häse, A. Nigam, P. Friederich, and A. Aspuru-Guzik, "Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation," Machine Learning: Science and Technology **1**, 045024 (2020), 1905.13741.

[28]G. Landrum, P. Tosco, B. Kelley, Ric, Sriniker, Gedeck, R. Vianello, NadineSchneider, E. Kawashima, A. Dalke, D. N, D. Cosgrove, B. Cole, M. Swain, S. Turk, AlexanderSavelyev, G. Jones, A. Vaucher, M. Wójcikowski, I. Take, D. Probst, K. Ujihara, V. F. Scalfani, G. Godin, A. Pahl, F. Berenger, JLVarjo, Strets123, JP, and DoliathGavid, "rdkit/rdkit: 2022_03_1 (q1 2022) release," (2022).

[29]H. E. Zimmerman and P. A. Wang, "The regioselectivity of the birch reduction," Journal of the American Chemical Society **115**, 2205–2216 (1993).

[30]H. E. Zimmerman, "A Mechanistic Analysis of the Birch Reduction," Accounts of Chemical Research **45**, 164–170 (2011).

[31]K. J. Kron, J. R. Hunt, J. M. Dawlaty, and S. Mallikarjun Sharada, "Modeling and characterization of exciplexes constituting photoredox cycles for CO_2 reduction: Insights from quantum chemistry and fluorescence spectroscopy," Accepted by Journal of Physical Chemistry A (2021), 10.26434/CHEMRXIV-2021-4JL66.

[32]V. Lukeš, R. Šolc, M. Barbatti, M. Elstner, H. Lischka, and H.-F. Kauffmann, "Torsional potentials and full-dimensional simulation of electronic absorption and fluorescence spectra of para-phenylene oligomers using the semiempirical self-consistent charge density-functional tight binding approach," The Journal of Chemical Physics **129**, 164905 (2008).

[33]C. S. Hartley, "Excited-state behavior of ortho-phenylenes," The Journal of Organic Chemistry **76**, 9188–9191 (2011).

[34]V. Lukeš, A. J. A. Aquino, H. Lischka, and H.-F. Kauffmann, "Dependence of optical properties of oligo-para-phenylenes on torsional modes and chain length," The Journal of Physical Chemistry B **111**, 7954–7962 (2007).

[35]X. Yang, J. Zhang, K. Yoshizoe, K. Terayama, and K. Tsuda, "ChemTS: an efficient python library for de novo molecular generation," Science and Technology of Advanced Materials **18**, 972–976 (2017).

[36]S. Refaely-Abramson, R. Baer, and L. Kronik, "Fundamental and excitation gaps in molecules of relevance for organic photovoltaics from an optimally tuned range-separated hybrid functional," Physical Review B **84**, 075144 (2011).

[37]A. Dreuw and M. Head-Gordon, "Single-reference ab initio methods for the calculation of excited states of large molecules," Chemical Reviews **105**, 4009–4037 (2005).

[38]E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. Gonzalez-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernandez Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, A. Jasz, H. Ji, H. Jiang, B. Kaduk, S. Kahler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjanszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavosevic, Z. Pei, S. Prager, E. I. Proynov, A. Rak, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. Mallikarjun Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, and A. I. Krylov, "Software for the frontiers of quantum chemistry: An overview of developments in the q-chem 5 package," The Journal of Chemical Physics **155**, 084801 (2021).

[39]A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," The Journal of Chemical Physics **98**, 5648–5652 (1993).

[40]F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy," Physical Chemistry Chemical Physics **7**, 3297–

3305 (2005).

[41] D. Rappoport and F. Furche, "Property-optimized gaussian basis sets for molecular response calculations," The Journal of Chemical Physics **133**, 134105 (2010).

[42] Q. Wu and T. Van Voorhis, "Direct optimization method to study constrained systems within density-functional theory," Physical Review A - Atomic, Molecular, and Optical Physics **72**, 024502 (2005).

[43] Q. Wu and T. Van Voorhis, "Direct calculation of electron transfer parameters through constrained density functional theory," The Journal of Physical Chemistry A **110**, 9212–9218 (2006).

[44] B. Kaduk, T. Kowalczyk, and T. Van Voorhis, "Constrained density functional theory," Chemical Reviews **112**, 321–370 (2012).

[45] R. Cammi and J. Tomasi, "Nonequilibrium solvation theory for the polarizable continuum model: A new formulation at the scf level with application to the case of the frequency-dependent linear electric response function," International Journal of Quantum Chemistry **56**, 465–474 (1995).

[46] M. Cossi and V. Barone, "Separation between fast and slow polarizations in continuum solvation models," The Journal of Physical Chemistry A **104**, 10614–10622 (2000).

[47] R. Improta, V. Barone, G. Scalmani, and M. J. Frisch, "A state-specific polarizable continuum model time dependent density functional theory method for excited state calculations in solution," The Journal of Chemical Physics **125**, 054103 (2006).

[48] Z.-Q. You, J.-M. Mewes, A. Dreuw, and J. M. Herbert, "Comparison of the Marcus and Pekar partitions in the context of non-equilibrium, polarizable-continuum solvation models," The Journal of Chemical Physics **143**, 204104 (2015).

[49] J. M. Mewes, Z. Q. You, M. Wormit, T. Kriesche, J. M. Herbert, and A. Dreuw, "Experimental benchmark data and systematic evaluation of two a posteriori, polarizable-continuum corrections for vertical excitation energies in solution," Journal of Physical Chemistry A **119**, 5446–5464 (2015).

[50] Q. Wu and T. Van Voorhis, "Extracting electron transfer coupling elements from constrained density functional theory," The Journal of Chemical Physics **125**, 164105 (2006).

[51] B. S. Brunschwig, J. Logan, M. D. Newton, and N. Sutin, "A semiclassical treatment of electron-exchange reactions. application to the hexaaquoiron (ii)-hexaaquoiron (iii) system," Journal of the American Chemical Society **102**, 5798–5809 (1980).

[52] R. A. Marcus, "On the theory of oxidation-reduction reactions involving electron transfer. i," The Journal of Chemical Physics **24**, 966–978 (1956).

[53] J.-D. Chai and M. Head-Gordon, "Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections," Physical Chemistry Chemical Physics **10**, 6615–6620 (2008).

[54] J.-D. Chai and M. Head-Gordon, "Systematic optimization of long-range corrected hybrid density functionals," The Journal of Chemical Physics **128**, 084106 (2008).

[55] T. N. Truong and E. V. Stefanovich, "A new method for incorporating solvent effect into the classical, ab initio molecular orbital and density functional theory frameworks for arbitrary shape cavity," Chemical Physics Letters **240**, 253 – 260 (1995).

[56] V. Barone and M. Cossi, "Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model," The Journal of Physical Chemistry A **102**, 1995–2001 (1998).

[57] M. Cossi, N. Rega, G. Scalmani, and V. Barone, "Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model," Journal of Computational Chemistry **24**, 669–681 (2003).

[58] C. W. Coley, W. H. Green, and K. F. Jensen, "Machine Learning in Computer-Aided Synthesis Planning," Accounts of Chemical Research **51**, 1281–1289 (2018).

[59] K. Molga, S. Szymkuć, and B. A. Grzybowski, "Chemist Ex Machina: Advanced Synthesis Planning by Computers," Accounts of Chemical Research **54**, 1094–1106 (2021).

[60] C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, and K. F. Jensen, "Prediction of Organic Reaction Outcomes Using Machine Learning," ACS Central Science **3**, 434–443 (2017).

[61] X. Zhu, C. H. Ho, and X. Wang, "Application of Life Cycle Assessment and Machine Learning for High-Throughput Screening of Green Chemical Substitutes," ACS Sustainable Chemistry and Engineering **8**, 11141–11151 (2020).

[62] M. J. Eckelman, "Life cycle inherent toxicity: A novel LCA-based algorithm for evaluating chemical synthesis pathways," Green Chemistry **18**, 3257–3264 (2016).

[63] L. M. Tufvesson, P. Tufvesson, J. M. Woodley, and P. Börjesson, "Life cycle assessment in green chemistry: Overview of key parameters and methodological concerns," International Journal of Life Cycle Assessment **18**, 431–444 (2013).