# Closed-Loop Transfer Enables AI to Yield Chemical Knowledge

Nicholas H. Angello[1,2]†, David M. Friday[1,2]†, Changhyun Hwang[2,4]†, Seungjoo Yi[2,3]†, Austin H. Cheng[5,7]^, Tiara C. Torres-Flores[2,4]^, Edward R. Jira[2,4], Wesley Wang[1,2], Alán Aspuru-Guzik[5,6,7,8,9,10,11]*, Martin D. Burke[1,2,12,13,14]*, Charles M. Schroeder[1,2,3,4]*, Ying Diao[1,2,4]*, Nicholas E. Jackson[1,2]*

**Author affiliations**

1. Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
2. Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
3. Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
4. Department of Chemical & Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, United States
5. Department of Chemistry, University of Toronto, Toronto, ON, Canada
6. Department of Computer Science, University of Toronto, Toronto, ON, Canada
7. Vector Institute for Artificial Intelligence, Toronto, ON, Canada
8. Canadian Institute for Advanced Research, Toronto, ON, Canada
9. Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON, Canada
10. Department of Materials Science and Engineering, University of Toronto, Toronto, ON, Canada
11. Acceleration Consortium, Toronto, ON, Canada
12. Carle R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
13. Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
14. Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

†Contributed equally, co-first author.

^Contributed equally, co-second author.

*Corresponding authors. Emails: jacksonn@illinois.edu; yingdiao@illinois.edu; cms@illinois.edu; mdburke@illinois.edu, alan@aspuru.com

AI-guided closed-loop experimentation has recently emerged as a promising method to optimize objective functions,[1,2] but the substantial potential of this traditionally black-box approach to reveal new scientific knowledge has remained largely untapped. Here, we report a new AI-guided approach, dubbed Closed-Loop Transfer (CLT), that integrates closed-loop experiments with physics-based feature selection and supervised learning to yield new scientific knowledge in parallel with optimization of objective functions. CLT surprisingly revealed that high-energy regions of the triplet state manifold are paramount in dictating molecular photostability in solution across a diverse chemical library of light-harvesting donor-bridge-acceptor oligomers. Remarkably, this insight emerged after automated modular synthesis and experimental characterization of only ~1.5% of the theoretical chemical space. Supervised learning models considering millions of combinations of 100+ physics-based descriptors further showed that high energy triplet states most strongly correlate with photostability, while excluding more commonly considered predictors such as the lowest energy triplet state. The physics-informed model for photostability was even further confirmed and then strengthened using an explicit experimental test set, validating the substantial power of the CLT method. Broadly, these findings show that interfacing physics-based modeling with closed-loop discovery campaigns unimpeded by synthesis bottlenecks can rapidly illuminate fundamental chemical insights and guide more rational pursuit of frontier molecular functions.

## INTRODUCTION

Achieving systematic and efficient chemical design, discovery, and understanding across large chemical search spaces is a grand challenge for chemistry. With recent developments in automated modular synthesis,[3–6] synthetically accessible chemical space is vast,[7] and

advanced physics-based approaches for targeted functional design are critically needed. But the paradigm of rational design and discovery has increasingly been supplanted by theory-driven exploration and combinatorial, high-throughput experimentation.[8–10] Recently, artificial intelligence (AI)-guided closed-loop platforms, where predictions, experiments, and analysis are automated and connected in a positive feedback loop, have shown great potential to accelerate scientific discovery in intractably large search spaces.[11–24] Despite these recent successes, it is not yet possible to broadly leverage such a paradigm to drive knowledge-based discovery of molecular functions. A major challenge lies in the need to choose *either* objective optimization or knowledge generation as the driver for data-scarce AI-guided campaigns. New closed-loop design paradigms are required to reveal fundamental scientific understanding[22] in parallel with data-driven problem solving.

The inability to rationally control photostability in organic light-harvesting molecules is emblematic of the limitations of existing discovery paradigms. Photostability is both a complex phenomenon and an essential property for light-active materials in technologies including organic photovoltaics,[25,26] electrochromic materials,[27] organic light-emitting diodes,[28] organic photocatalysts,[29] photoactive coatings,[30] and fluorescent dyes.[31,32] Due to the large number of complex physical and chemical descriptors potentially governing photostability in organic compounds, we lack a general set of design principles for controlling photodegradation across a broad chemical space.[33–37] Prior work has often considered the lowest lying excited triplet state ($T_1$) energy as a primary determinant of photostability and photodegradation across scattered chemical classes.[38–43] Although some recent studies have suggested that high energy triplet states ($T_n$, $n > 1$) may play a role in the performance and photostability of organic light-harvesting compounds,[44–47] experimental observation and computational modeling of such high-energy states is extremely challenging.[48] New experimental and computational approaches examining a wide range of chemically diverse molecules with unbiased evaluation of many possible physics-based features are critically required to understand the fundamental determinants of molecular photostability in the context of functional molecular discovery.[39]

Here we report a new AI-guided closed-loop method that simultaneously reveals fundamental knowledge while efficiently balancing exploration and exploitation across a broad chemical space. We hypothesized that AI-guided closed-loop experimentation could nucleate a physics-informed supervised learning model capable of identifying highly photostable compounds while concomitantly elucidating the fundamental determinants of molecular photostability. Our results show that a single campaign employing Bayesian optimization (BO) guided closed-loop experiments leveraging automated modular small molecule synthesis, performed in concert with physics-based modeling, resulted in both high-performing photoactive molecules and the unexpected revelation that molecular photostability is most strongly correlated with the high energy portion of the triplet excited state manifold. These results were achieved using BO to intentionally balance exploration and exploitation of the diverse chemical space with no pre-existing knowledge of the determinants of photostability in organic molecules. Our work shows that interfacing physics-based modeling with the data emerging from BO-guided closed-loop discovery can deliver physical insights into frontier molecular functionality.

## RESULTS & DISCUSSION

We introduce a new closed-loop approach called closed-loop transfer (CLT) (**Figure 1**) that consists of three phases: Phase I integrates BO that intentionally balances exploration and exploitation (GRYFFIN[49]) in a closed-loop process with automated modular small molecule synthesis[3,50,51] and multidimensional characterization (here, a solution-based solar irradiation cell[25]), until saturation of the functional objective target is achieved; Phase II integrates comprehensive whole molecule DFT calculations and feature selection with physics-informed supervised learning models to extract physical insights and extrapolate broadly across chemical space; Phase III uses an explicit experimental test set for statistical validation of the physical insights and understanding across the entire chemical search space. To establish the CLT approach, we chose to first focus on molecular photostability of conjugated oligomers in the solution state which intentionally avoids additional complicating factors due to processing, film morphology, and interfacial effects.

To initialize the CLT campaign, we first defined the chemical space (**Figure 2**) for the closed-loop (**Figure 3a**) procedure. We chose the donor-bridge-acceptor (D-B-A) motif ubiquitous in state-of-the-art light absorbing conjugated materials[52–54] as a molecular design scaffold that can be readily modularized into function-infused building blocks amenable to automated chemical synthesis. The donor and pi-bridge building blocks were inspired by successful motifs found in molecular electronics, whereas the acceptor building blocks were algorithmically chosen in a down-selection process (SI Section 1) maximizing molecular diversity within all purchasable (hetero)aryl-halide building blocks (Extended Data Figure 1). The resulting modularized chemical space was thus infused with light-harvesting functionality while also sampling a diverse chemical space rich with discovery potential. In total, the chemical space includes three donors, seven pi-bridges, and 100 acceptor blocks, yielding a total of 2,200 potential oligomers when accounting for symmetry and molecules lacking a pi-bridge. To facilitate AI-driven BO, the chemical space was featurized using concatenations of rapidly calculable structural and electronic descriptors of the building blocks computed with density functional theory (DFT) and RDKit[55] (Tables S1-S2).

Critical to the initialization of the CLT campaign was an emphasis on diversity early in the closed-loop campaign, followed by balanced exploration and exploitation in subsequent rounds (**Figure 3b**). To initiate the first round, molecular diversity sampling was utilized to decrease initiation bias in the closed loop (SI Section 1). The 2nd round contained additional diversity selected molecules and BO recommendations. For each subsequent closed-loop

iteration, a batch of six molecules was drawn to uniformly sample along the domain between fully exploitative and fully explorative within the BO algorithm.[49] This strategy ensured that while BO was optimizing for photostability, it was concurrently diversifying its knowledge of the chemical space to maximally inform general scientific understanding. At the end of each round of the closed-loop cycle, experimental photostability data were measured and passed to the BO model, which then suggested synthetic candidates for the next round. These molecular targets and a list of their nearest neighbors in feature space (known as 'backups' in cases where the target is not readily synthesizable or testable – **Figure 3b**) were then automatically populated on a custom-built digital project manager, a web-based database and dashboard, for subsequent automated synthesis (Extended Data Figure 2).

During the closed-loop iterations, automated modular small molecule synthesis was conducted using a version of our iterative C-C bond-forming robot optimized for reaction reproducibility (**Figure 3c** and Extended Data Figure 3).[14] Initial tests demonstrated that a fully automated two-step synthesis was feasible. The first step proved to be generally efficient, whereas the second step was initially more challenging and variable due to the chemical diversity in the 100 aryl-halide acceptor blocks. We thus opted to independently synthesize and scale-up the first coupling products (donor-bridges, SI Section 2), and separately optimize the second reaction step using slow-release cross-coupling.[56] We found that using both our recently reported general reaction conditions for heteroaryl cross-coupling discovered via an AI-guided closed-loop campaign[14] (GC1, **Figure 3c**) and newly discovered anhydrous slow-release coupling conditions (GC2, **Figure 3c**, and SI Section 2[50]) maximized the synthetic hit-rate (~60%).

Following synthesis, purification, and structural verification (SI Section 2) in each round of the closed-loop, the photophysical properties of donor-bridge-acceptor molecules were characterized via solution-based photodegradation in a solar irradiation cell (**Figure 3d**). The photophysical properties for all molecules were measured under standardized concentrations, in the same solvent (chlorobenzene), and under a controlled atmosphere using a glovebox with oxygen and humidity control. We measured two properties: spectral overlap (SO), defined as the integral of the normalized overlap of a molecule's absorbance spectrum and the solar irradiance spectrum, and spectral decay time ($T_{80}$), defined as the time required for the observed absorbance spectrum to decay to 80% of its initial value under constant irradiation. Given that photodegradation is sensitive to the local environment, an internal standard was used to ensure consistency (SI Section 3). Based on a first-order kinetic model, we chose to optimize photostability, defined as the product of SO and $T_{80}$ (Extended Data Figure 4).

BO-driven closed-loop experimentation proceeded across five rounds, automatically synthesizing 30 novel donor-bridge acceptor light-harvesting oligomers (**Figure 4a** and Extended Data Table 1), until saturation of the experimental photostability was observed. The first round of suggestions, relying on diversity-driven selection, resulted in a set of 10 molecules with low to moderate photostabilities. Subsequent rounds 2-4 relying on the BO strategy probed existing and new regions of functional chemical space, discovering molecules at the extremes of $T_{80}$ and SO, and some that maximized their product. By the conclusion of the 5$^{th}$ round of the closed-loop, the average photostability of the top-five molecules showed an effective plateau, signaling the end of the BO-guided closed-loop campaign (**Figure 4b**). Importantly, a nearly 400% increase in the average photostability of top-5 performers was achieved by only sampling < 1.5% of the total space of 2,200 possible molecules, a result consistent with prior theoretical predictions but not previously verified experimentally.[49]

Following saturation of the experimental photostabilities within the BO campaign, we used the new CLT method to enable a physics-based understanding of functional properties. We performed whole molecule DFT calculations[57] on all 2,200 D-B-A and D-A oligomers, and from these results a comprehensive set of 114 physical and chemical molecular descriptors were extracted (Tables S1-S4). These descriptors were then integrated in backward stepwise feature selection and support vector regression (SVR) with the experimental $T_{80}$ and SO values. SVR models were trained to predict SO and $T_{80}$ separately due to an observed inverse relationship between $T_{80}$ and SO, with the best model achieving leave-one-out-validation (LOOV) predictive accuracy of $R^2 > 0.80$ on the experimental photostability using 6 features from the TDOS (TDOS at 2.6, 2.8, 3.8, 3.9, 4.0, and 4.6 eV). We next explored all possible SVR models (12,996) using two-descriptor combinations of the 114-feature set and LOOV performance metric (**Figure 4c**). Unexpectedly our results showed that, as opposed to the conventional $T_1$ energy descriptor of photostability, high-energy TDOS emerged as a primary determinant of molecular photostability across the entire chemical space (high $T_{80}$ LOOV $R^2$) (Figure S4-S5). We then examined 234,136 four-descriptor SVR models and plotted the distribution of performance ($R^2$) for all models containing the most predictive region of the TDOS at 4.0 eV compared to the $T_1$ energy (**Figure 4d**). Our results clearly show a stark difference in the predictive capabilities of supervised learning models employing the high-energy region of the TDOS as opposed to the low energy region. Importantly, this knowledge emerged *a priori*, across a broad chemical space, with equal weight given to all possible hypotheses described by 114 physical and chemical descriptors. This data-driven physical insight suggests a more nuanced understanding of the triplet state manifold's role in molecular photostability is required that extends beyond simple $T_1$ energetics.

As learning in the low-data limit can induce spurious correlations in supervised learning models, the final phase of the CLT campaign involved construction of an independent experimental test set to validate the observation that high-lying triplet states determine molecular photostability. The best $T_{80}$ SVR model trained on six values of the TDOS (Figures S6) was used to predict $T_{80}$ across the entire set of 2,200 oligomers (**Figure 5a**). The predicted $T_{80}$ value of each oligomer was then multiplied by its predicted SO value (a separate linear model

incorporating the TDDFT singlet absorption spectrum – Figure S1) to obtain the predicted photostabilities across the entire set of 2,200 oligomers (Figure S15). Using these predictions, we formed two batches of 7 molecules, one high-performing ("top 7") and one low-performing ("bottom 7"), to serve as experimental validation sets (Figure S8). These two batches possessed the following statistical features (SI Section 1): i) identical average SO within the 5.5%-9% SO region (which emphasizes $T_{80}$ effects in the photostability), and ii) similar standard deviations for the predicted $T_{80}$ (which samples a broad and chemically diverse space). We then synthesized and characterized the photophysical properties of the top 7 and bottom 7 batches (**Figure 5b**). We observed a statistically significant photostability difference between these groups in the predicted direction (**Figure 5b**; average $T_{80}$*SO = 165 for top 7 versus average $T_{80}$*SO = 97 for bottom 7, p=0.026, Spearman $R^2$=0.54), validating our CLT-derived physical insight that the triplet manifold is a key descriptor of molecular photostability. Interestingly, one outlier in the bottom 7 showed surprisingly high observed photostability; subsequent analysis demonstrated that this quinone-based molecule exhibited a high TDOS approximately 1 eV above its $T_1$ energy while maintaining a low structural complexity – a feature unseen relative to the 30 molecules synthesized in the five closed-loop rounds (Extended Data Figure 5).

Using the full experimental photostability data set generated by the CLT campaign (44 molecules, Extended Data Figure 6a), we performed a final retraining of supervised learning models to maximally extract physics-informed insights into photostability. Strikingly, upon re-training all 12,996 two-feature SVR models when considering the full experimental dataset, the high-energy TDOS even more strongly emerged as a critical determinant of molecular photostability across the entire chemical space (high $T_{80}$ LOOV $R^2$) (**Figures 5c**) and a superior descriptor relative to the conventional $T_1$ energy (**Figure 5d**). 2.5 million four-feature SVR models were trained to identify the most common feature present in the most predictive $T_{80}$ models ($R^2$ > 0.70) across all descriptors (Figure S9). The two most common features, present in ~30% of the most predictive models were the TDOS at 4.0 eV and the number of heteroatoms (Extended Data Figure 6b). Whereas the number of heteroatoms may be attributed to the reliably poor $T_{80}$ of the benzodithiophene donor, the observed correlation between a high TDOS at 4.0 eV and a low $T_{80}$ solidifies the closed-loop-derived knowledge that the high-lying TDOS is a critical determinant of molecular photostability.

Finally, we note that within our CLT campaign, the highest-performing four-feature SVR model trained on all 44 molecules of the campaign utilizes two simple structural features (the number of rotatable bonds and the number of heteroatoms) in addition to the TDOS at 3.9 eV and 2.5 eV (Extended Data Figure 6b, Figures S10-S11). That the data-scarce regime can still leverage simple structural descriptors is likely indicative of additional chemical or physical complexities not fully captured by the TDOS. However, due to the challenges of understanding the properties of the triplet manifold, substantial theoretical and experimental advances will be required to fully characterize the energetics and dynamics of photodegradation processes mediated by high energy triplet states beyond $T_1$.

## CONCLUSION

The closed-loop experiment reported here, augmented by physics-informed feature selection and supervised learning, elucidated fundamental knowledge regarding molecular photostability while simultaneously optimizing towards high function molecular targets. The remarkable efficiency with which this AI-guided campaign yielded physical insights (only 30 molecules) resulted from combining an AI-guided balance of exploration and exploitation with physics-based modeling and feature selection. This CLT paradigm may prove to be broadly applicable to other frontier applications and fields of inquiry, in particular low data regimes and multidimensional molecular properties which are challenging to predict *a priori*. We believe that CLT will serve as a valuable playbook for harnessing the strengths of BO and transitioning to hypothesis-driven discovery campaigns reinforced by physics-based insights. The CLT method may also provide a roadmap for AI-guided closed-loop algorithms to achieve physics-based, hypothesis-driven research independent of human input.

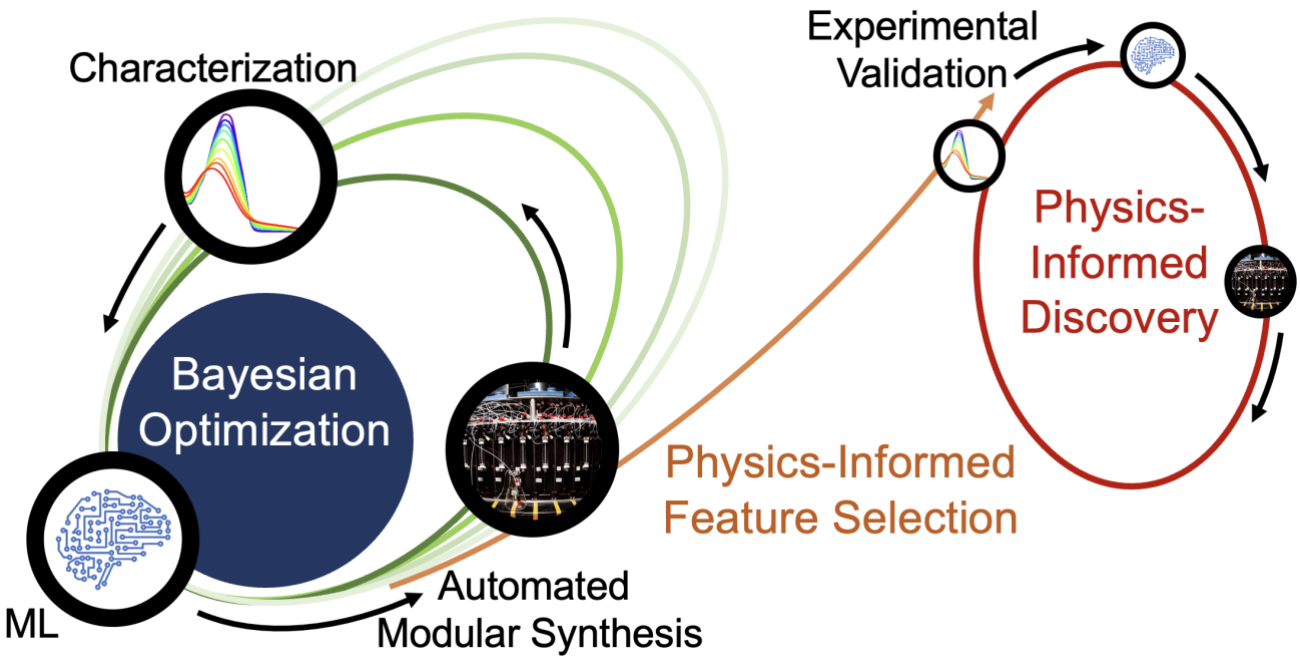



Figure 1. Closed-loop transfer paradigm for physics-informed functional molecular discovery.

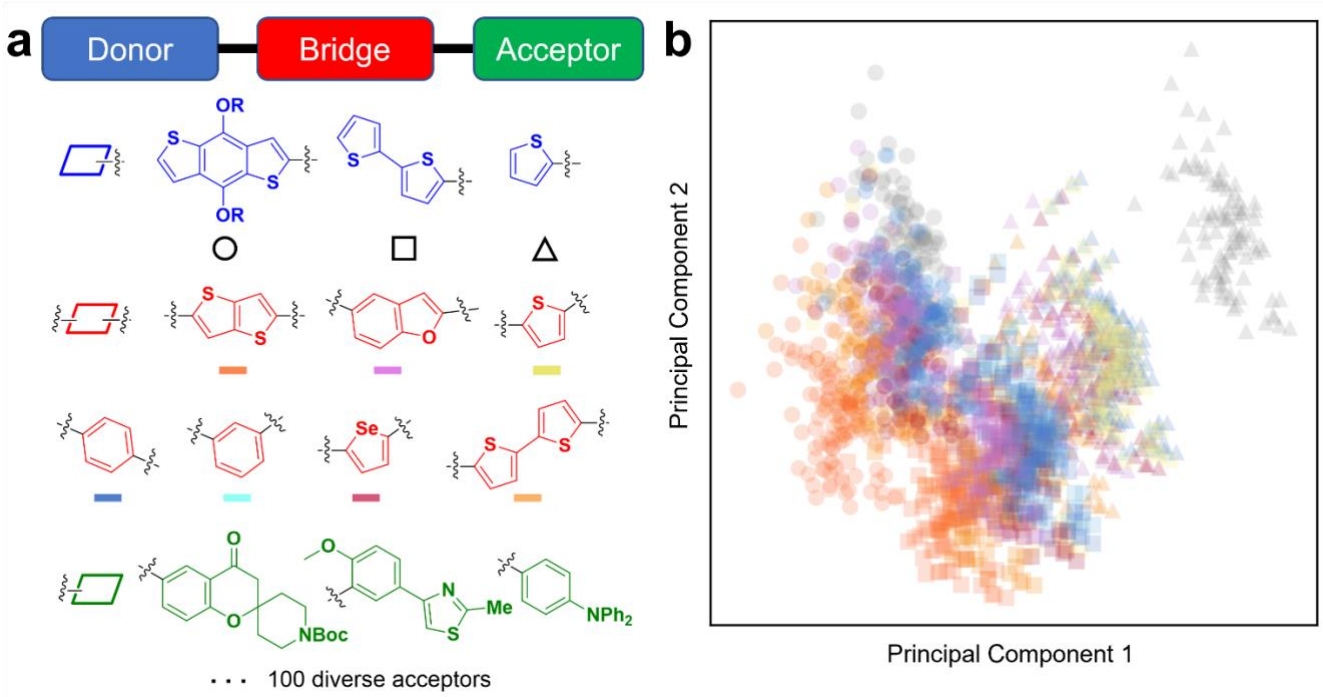
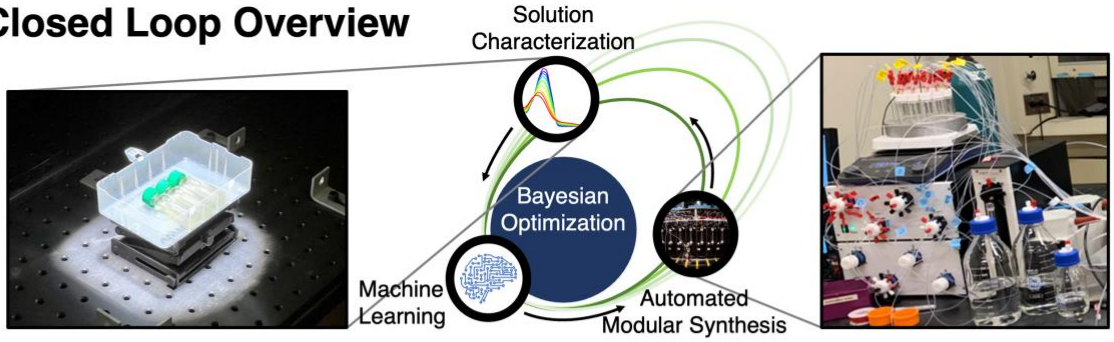
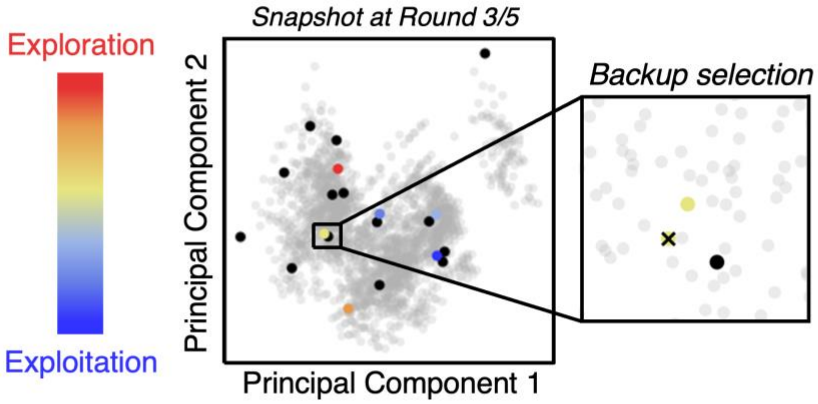



Figure 2. **A molecular building block set for light-harvesting small molecules. a,** The modular building blocks considered in this work for designing new light-harvesting donor-bridge-acceptor oligomers. The full list of 100 acceptor building blocks is shown in Extended Data Figure 1. **b,** Principal component analysis (PCA) projection of the full chemical space (2,200 molecules), shape- and color-coded according to the structure of the donor and bridge building blocks, respectively. Gray coloring indicates the absence of a bridge block, i.e., molecules that are donor-acceptor.

## a Closed Loop Overview

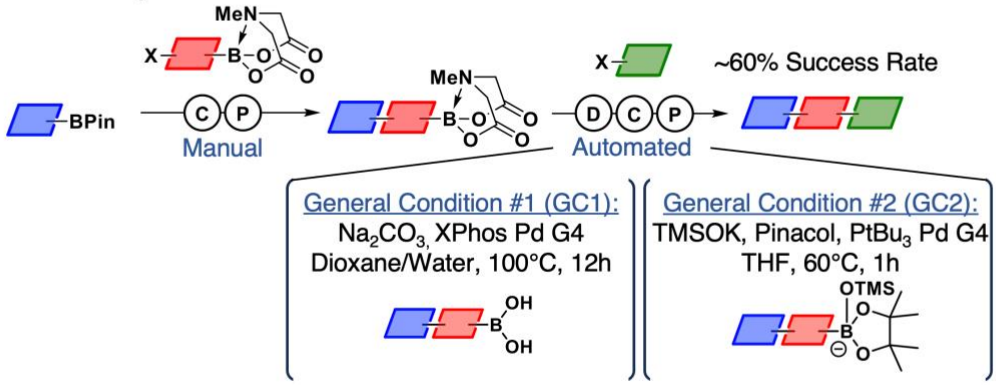



## b Machine Learning




| Recommended Molecules | Backup Molecules 1-10 |
|---|---|
| $A_0$ | $A_1, A_2, \ldots A_{10}$ |
| $B_0$ | $B_1, B_2, \ldots B_{10}$ |
| $C_0$ | $C_1, C_2, \ldots C_{10}$ |
| $D_0$ | $D_1, D_2, \ldots D_{10}$ |
| $E_0$ | $E_1, E_2, \ldots E_{10}$ |
| $F_0$ | $F_1, F_2, \ldots F_{10}$ |

## c Modular Synthesis




## d Solution Characterization

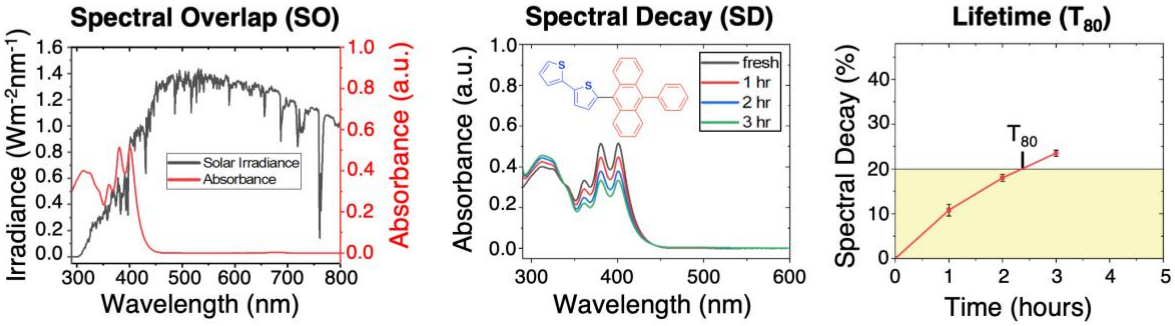


Figure 3. **Closed-loop experimental strategy.** An overview of the closed-loop process is shown in **a,** demonstrating the major steps under the guidance of AI-driven Bayesian optimization and images of the solution characterization process and automated synthesis equipment. **b,** A visualization of round 3 of the closed-loop campaign in which six molecules are recommended, balancing exploration and exploitation, along with several backup molecules for each. In round 3, the 3rd most explorative molecule (yellow) was unsynthesizable, requiring the use of a backup, as shown in the inset. **c,** The reaction conditions employed in the automated synthesis of ~60% of the molecules recommended by Bayesian optimization. **d,** Example data generated by the solution characterization, showing the fresh absorbance and solar simulator irradiance spectrum used to calculate spectral overlap, the decay of absorbance over time for a representative molecule (inset), and a plot of the decay of the absorption spectrum over time, which is used to calculate $T_{80}$ of each molecule with measurable absorbance.

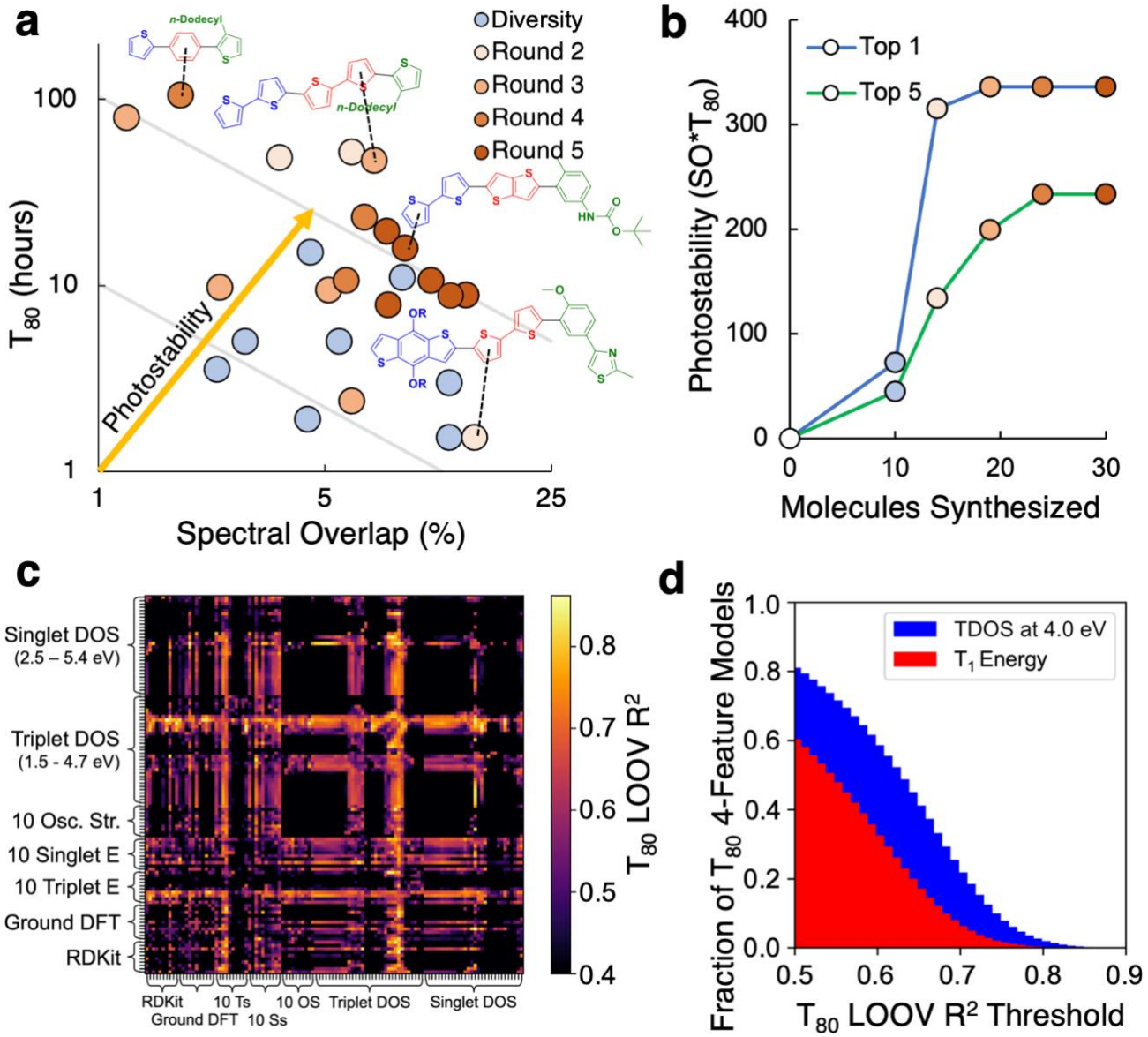



Figure 4. **Results from closed-loop Bayesian optimization. a**, Data generated per round of closed-loop experimentation, with representative highly photostable oligomers shown. **b,** Cumulative photostability performance per round of closed-loop experimentation. Top 5 corresponds to the average of the top 5 molecules. **c**, All support vector regression leave-one-out validation (LOOV) results for predicting $T_{80}$ from 2-feature combinations. **d,** Comparison of the prediction strength of all possible 4-feature models containing either the TDOS at 4.0 eV or the $T_1$ energy.

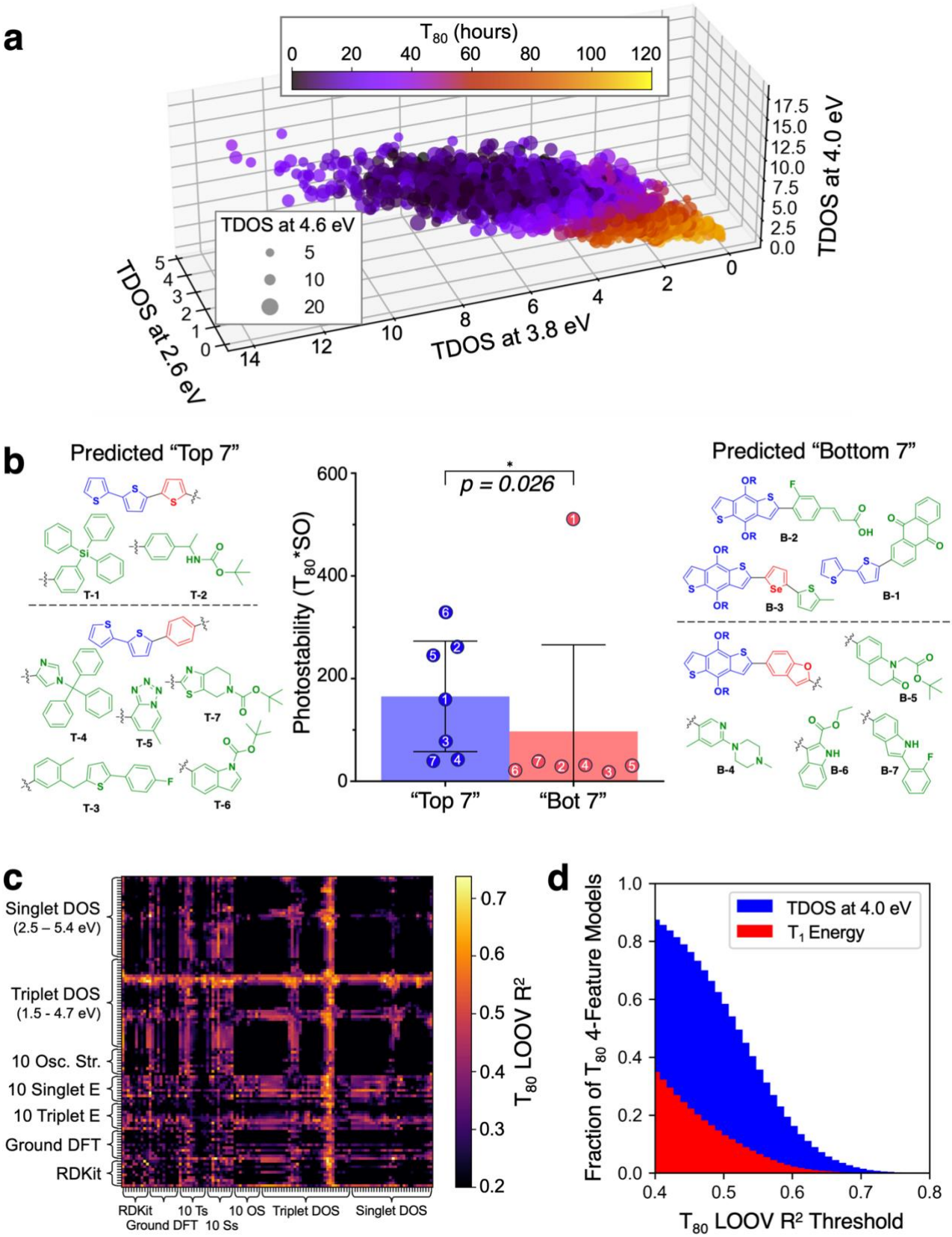


Figure 5. **Hypothesis-driven discovery model and experimental validation. a,** The physics-based model for $T_{80}$ using the triplet density of states of each molecule is plotted against four triplet density of states values relevant to the prediction task. These predicted $T_{80}$ values, along with the SO criteria, were used to select the Top 7 and Bottom 7 molecules, shown on the left and right side of **b**, respectively. **b,** Photostabilities of the molecules in the experimental validation set, their average, standard deviation, and the associated p-value from a Mann-Whitney test, are plotted in the center. These values demonstrate a statistically significant difference in performance between the two sets of molecules, as well as an outlier high performer (Bottom #2) as described in the text. **c,** All support vector regression leave-one-out validation (LOOV) results for predicting $T_{80}$ on the entire 44 molecule dataset from 2-feature combinations. **d,** Comparison of the prediction strength of all possible 4-feature models containing either the TDOS at 4.0 eV or the $T_1$ energy in predicting $T_{80}$ for the entire 44 molecule dataset. Comparing the plots in **c** and **d** to **Figure 4 c** and **d** shows the improved predictive strength of TDOS at 4.0 eV over all other features, especially $T_1$, after the validation dataset.

Online content
The data and methods that support the findings of this study are available in the Supplementary Information: machine learning and Bayesian optimization methods, details of automated synthesis and purification, and details of solution photodegradation testing.

AUTHOR CONTRIBUTIONS
N.H.A., D.M.F., C.H., and S.Y. contributed equally to this project (co-first author). A.H.C. and T.T.F. contributed equally to this project (co-second author). The project was designed by N.H.A., D.M.F., C.H., S.Y., E.R.J., A.A.-G., M.D.B., C.M.S., Y.D., and N.E.J. Molecule synthesis was conducted by N.H.A., S.Y., T.T.F., E.R.J., and W.W. Solution testing was conducted by C.H. Bayesian optimization and regression model training was conducted by D.M.F. and A.H.C. N.H.A, D.M.F., C.H., S.Y., A.A.-G., M.D.B., C.M.S., Y.D., and N.E.J. wrote the manuscript with contributions from all authors.

## AUTHOR INFORMATION


### Corresponding Authors

**Nicholas E. Jackson** – Department of Chemistry, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; orcid.org/0000-0002-1470-1903; Email: jacksonn@illinois.edu.

**Ying Diao** – Department of Chemical and Biomolecular Engineering, Department of Chemistry, Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, Urbana, Illinois, 61801, United States; orcid.org/0000-0002-8984-0051; Email: yingdiao@illinois.edu.

**Charles M. Schroeder** – Department of Materials Science and Engineering, Department of Chemical and Biomolecular Engineering, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; orcid.org/0000-0001-6023-2274; Email: cms@illinois.edu

**Martin D. Burke** – Department of Chemistry, Department of Biochemistry, Institute for Genomic Biology, Carle Illinois College of Medicine, and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; orcid.org/0000-0001-7963-7140; Email: mdburke@illinois.edu

**Alán Aspuru-Guzik –** Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3H6, Canada; CIFAR Artificial Intelligence Research Chair, Vector Institute, Toronto, Ontario M5S 1m1, Canada; Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario M5S 1M1, Canada; orcid.org/0000-0002-8277-4434. alan@aspuru.com.

### Authors

**Nicholas H. Angello** – Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; orcid.org/0000-0001-6436-3669.

**David M. Friday** – Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; orcid.org/0009-0004-3189-5115.

**Changhyun Hwang** – Department of Chemical and Biomolecular Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, United States orcid.org/0000-0003-1720-7852.

**Seungjoo Yi** – Department of Materials Science and Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, United States orcid.org/0009-0009-3881-9226.

**Austin H. Cheng** – Department of Chemistry, University of Toronto, Ontario M5S 3H6, Canada; Vector Institute, Toronto, Ontario M5S 1M1, Canada; orcid.org/0000-0002-9497-3932

**Tiara C. Torres-Flores** - Department of Chemical and Biomolecular Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, United States.

**Edward R. Jira** – Department of Chemical and Biomolecular Engineering, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States.

**Wesley Wang** - Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; orcid.org/0000-0002-1420-3031.


Notes
**The authors declare no competing financial interest.**

## ACKNOWLEDGMENT

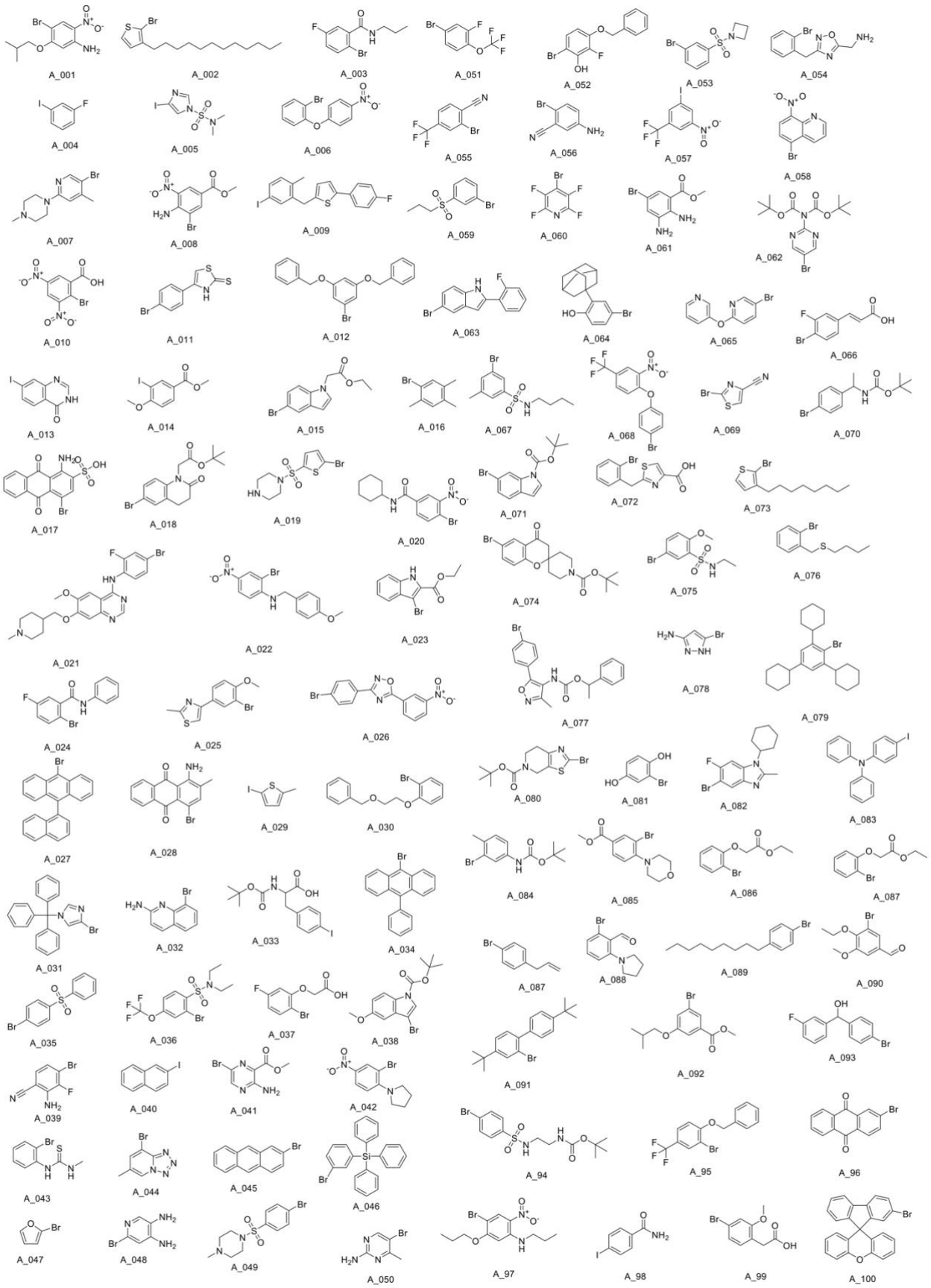

This work was supported by the Molecule Maker Lab Institute, an AI Research Institutes program supported by the US National Science Foundation under grant no. 2019897 (N.E.J., Y.D., C.M.S, M.D.B.). A. A.-G. and A. C. acknowledge support from the Canada 150 Research Chairs Program and the Acceleration Consortium at University of Toronto, as well as the generous support of Anders G. Frøseth. T.T.F., Y.D. and C.M.S. acknowledge support by the IBM-Illinois Discovery Accelerator Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the NSF.


## REFERENCES

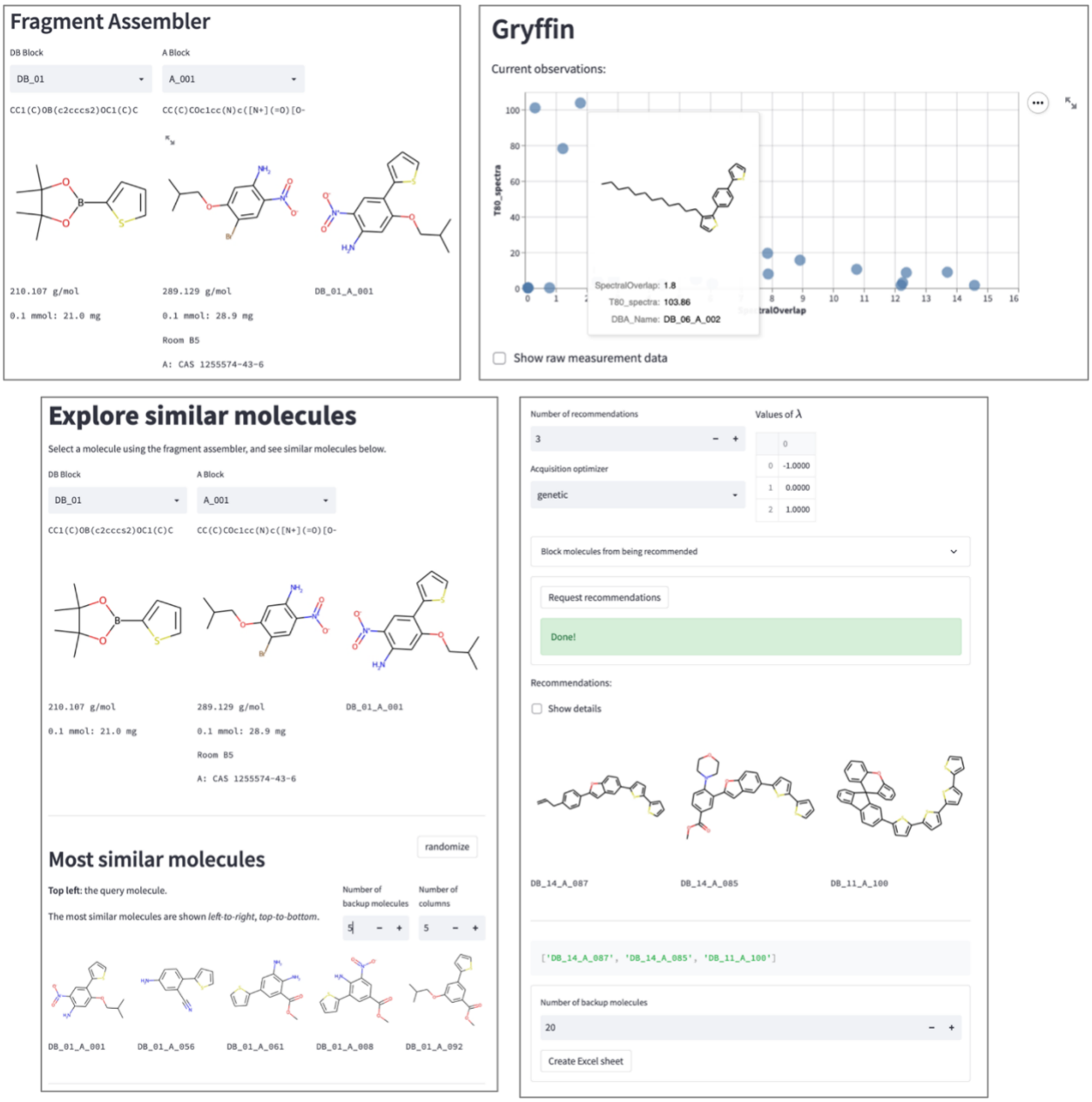

1. Flores-Leonar, M. M. *et al.* Materials Acceleration Platforms: On the way to autonomous experimentation. *Curr. Opin. Green Sustain. Chem.* **25**, 100370 (2020).
2. Peng, X. & Wang, X. Next-generation intelligent laboratories for materials design and manufacturing. *MRS Bull.* **48**, 179–185 (2023).
3. Li, J. *et al.* Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).
4. Plante, O. J., Palmacci, E. R. & Seeberger, P. H. Automated Solid-Phase Synthesis of Oligosaccharides. *Science* **291**, 1523–1527 (2001).

5. Caruthers, M. H. Gene Synthesis Machines: DNA Chemistry and Its Uses. *Science* **230**, 281–285 (1985).
6. Merrifield, R. B. Automated Synthesis of Peptides. *Science* **150**, 178–185 (1965).
7. Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **48**, 722–730 (2015).
8. Amis, E. J., Xiang, X. D. & Zhao, J. C. Combinatorial Materials Science: What's New Since Edison? *NIST* **27**, (2002).
9. Liu, A. L. *et al.* Composition Gradient High-Throughput Polymer Libraries Enabled by Passive Mixing and Elevated Temperature Operability. *Chem. Mater.* **34**, 6659–6670 (2022).
10. Venkatesh, R. *et al.* Data Science Guided Experiments Identify Conjugated Polymer Solution Concentration as a Key Parameter in Device Performance. *ACS Mater. Lett.* **3**, 1321–1327 (2021).
11. Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).
12. Burger, B. *et al.* A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
13. Rao, R. *et al.* Advanced machine learning decision policies for diameter control of carbon nanotubes. *Npj Comput. Mater.* **7**, 1–9 (2021).
14. Angello, N. H. *et al.* Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. *Science* **378**, 399–405 (2022).
15. Burés, J. & Larrosa, I. Organic reaction mechanism classification using machine learning. *Nature* **613**, 689–695 (2023).
16. Chang, J. *et al.* Efficient Closed-loop Maximization of Carbon Nanotube Growth Rate using Bayesian Optimization. *Sci. Rep.* **10**, 9040 (2020).
17. Nikolaev, P. *et al.* Autonomy in materials research: a case study in carbon nanotube growth. *Npj Comput. Mater.* **2**, 1–6 (2016).
18. Jablonka, K. M., Jothiappan, G. M., Wang, S., Smit, B. & Yoo, B. Bias free multiobjective active learning for materials design and discovery. *Nat. Commun.* **12**, 2312 (2021).
19. Shmilovich, K. *et al.* Discovery of Self-Assembling π-Conjugated Peptides by Active Learning-Directed Coarse-Grained Molecular Simulation. *J. Phys. Chem. B* **124**, 3873–3891 (2020).
20. Langner, S. *et al.* Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv. Mater.* **32**, 1907801 (2020).
21. Kusne, A. G. *et al.* On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
22. Krenn, M. *et al.* On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**, 761–769 (2022).
23. Christensen, M. *et al.* Data-science driven autonomous process optimization. *Commun. Chem.* **4**, 1–12 (2021).
24. Tamasi, M. J. *et al.* Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids. *Adv. Mater.* **34**, 2201809 (2022).
25. Burlingame, Q. *et al.* Intrinsically stable organic solar cells under high-intensity illumination. *Nature* **573**, 394–397 (2019).
26. Burlingame, Q., Ball, M. & Loo, Y.-L. It's time to focus on organic solar cell stability. *Nat. Energy* **5**, 947–949 (2020).
27. Tällberg, R., Jelle, B. P., Loonen, R., Gao, T. & Hamdy, M. Comparison of the energy saving potential of adaptive and controllable smart windows: A state-of-the-art review and simulation studies of thermochromic, photochromic and electrochromic technologies. *Sol. Energy Mater. Sol. Cells* **200**, 109828 (2019).
28. Chan, C.-Y. *et al.* Stable pure-blue hyperfluorescence organic light-emitting diodes with high-efficiency and narrow emission. *Nat. Photonics* **15**, 203–207 (2021).
29. Kosco, J. *et al.* Enhanced photocatalytic hydrogen evolution from organic semiconductor heterojunction nanoparticles. *Nat. Mater.* **19**, 559–565 (2020).
30. Kim, J.-H. *et al.* Hydrophilic/Hydrophobic Silane Grafting on TiO2 Nanoparticles: Photocatalytic Paint for Atmospheric Cleaning. *Catalysts* **11**, 193 (2021).
31. Eggeling, C., Widengren, J., Rigler, R. & Seidel, C. A. M. Photobleaching of Fluorescent Dyes under Conditions Used for Single-Molecule Detection: Evidence of Two-Step Photolysis. *Anal. Chem.* **70**, 2651–2659 (1998).
32. O'Brien, J., Wilson, I., Orton, T. & Pognan, F. Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur. J. Biochem.* **267**, 5421–5426 (2000).
33. Alem, S. *et al.* Degradation Mechanism of Benzodithiophene-Based Conjugated Polymers when Exposed to Light in Air. *ACS Appl. Mater. Interfaces* **4**, 2993–2998 (2012).
34. Mateker, W. R. & McGehee, M. D. Progress in Understanding Degradation Mechanisms and Improving Stability in Organic Photovoltaics. *Adv. Mater.* **29**, 1603940 (2017).
35. Liu, Z.-X. *et al.* Molecular insights of exceptionally photostable electron acceptors for organic photovoltaics. *Nat. Commun.* **12**, 3049 (2021).
36. Bekri, N., Asmare, E., Mammo, W. & Tegegne, N. A. Photostability of benzodithiophene based polymer: effect of PC60BM and intermolecular interactions. *Mater. Res. Express* **9**, 055502 (2022).
37. Anderson, M. A., Hamstra, A., Larson, B. W. & Ratcliff, E. L. Distinguishing photo-induced oxygen attack on alkyl chain versus conjugated backbone for alkylthienyl-benzodithiophene (BDTT)-based push–pull polymers. *J. Mater. Chem. A* **11**, 17858–17871 (2023).
38. Korycka-Dahl, M. & Richardson, T. Photodegradation of DNA with fluorescent light in the presence of riboflavin, and photoprotection by flavin triplet-state quenchers. *Biochim. Biophys. Acta* **610**, 229–234 (1980).
39. Groeneveld, I., Kanelli, M., Ariese, F. & van Bommel, M. R. Parameters that affect the photodegradation of dyes and pigments in solution and on substrate – An overview. *Dyes Pigments* **210**, 110999 (2023).
40. Kosumi, D., Horibe, T., Sugisaki, M., Cogdell, R. J. & Hashimoto, H. Photoprotection Mechanism of Light-Harvesting Antenna Complex from Purple Bacteria. *J. Phys. Chem. B* **120**, 951–956 (2016).
41. McNeill, K. & Canonica, S. Triplet state dissolved organic matter in aquatic photochemistry: reaction mechanisms, substrate scope, and photophysical properties. *Environ. Sci. Process. Impacts* **18**, 1381–1399 (2016).
42. Distler, A. *et al.* Effect of PCBM on the Photodegradation Kinetics of Polymers for Organic Photovoltaics. *Chem. Mater.* **24**, 4397–4405 (2012).
43. Zheng, Q. & Lavis, L. D. Development of photostable fluorophores for molecular imaging. *Curr. Opin. Chem. Biol.* **39**, 32–38 (2017).
44. Zhang, K. *et al.* New Insights about the Photostability of DNA/RNA Bases: Triplet nπ* State Leads to Effective Intersystem Crossing in Pyrimidinones. *J. Phys. Chem. B* **125**, 2042–2049 (2021).

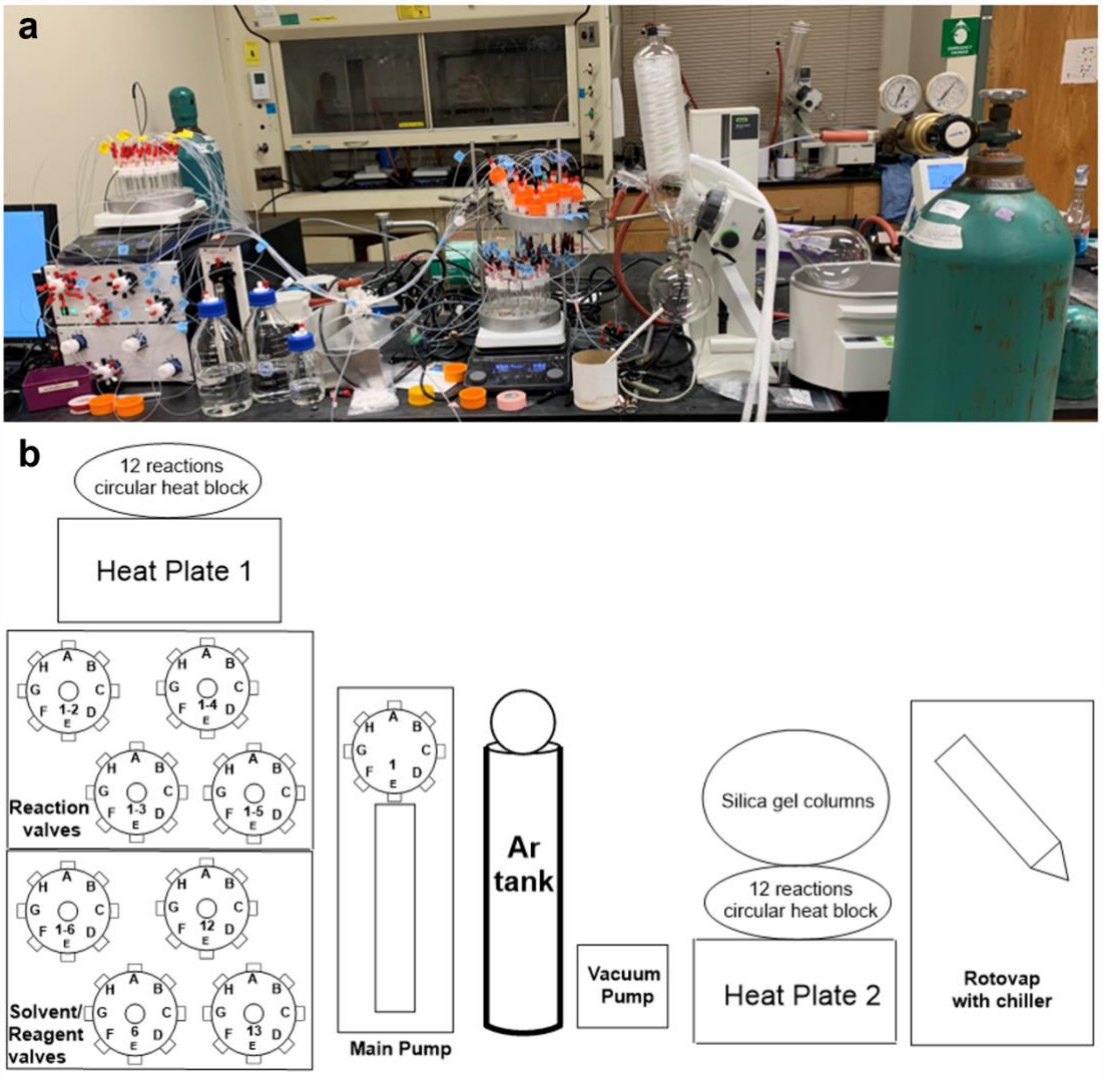
45. Guo, X. *et al.* Unraveling the Important Role of High-Lying Triplet–Lowest Excited Singlet Transitions in Achieving Highly Efficient Deep-Blue AIE-Based OLEDs. *Adv. Mater.* **33**, 2006953 (2021).
46. M. Kaminski, J. *et al.* Vibronic effects accelerate the intersystem crossing processes of the through-space charge transfer states in the triptycene bridged acridine–triazine donor–acceptor molecule TpAT-tFFO. *Chem. Sci.* **13**, 7057–7066 (2022).
47. Perun, S., Tatchen, J. & Marian, C. M. Singlet and Triplet Excited States and Intersystem Crossing in Free-Base Porphyrin: TDDFT and DFT/MRCI Study. *ChemPhysChem* **9**, 282–292 (2008).
48. Mukherjee, S., Fedorov, D. A. & Varganov, S. A. Modeling Spin-Crossover Dynamics. *Annu. Rev. Phys. Chem.* **72**, 515–540 (2021).
49. Häse, F., Aldeghi, M., Hickman, R. J., Roch, L. M. & Aspuru-Guzik, A. Gryffin: An algorithm for Bayesian optimization of categorical variables informed by expert knowledge. *Appl. Phys. Rev.* **8**, 031406 (2021).
50. Wang, W. *et al.* Rapid Automated Iterative Small Molecule Synthesis. Preprint at https://doi.org/10.26434/chemrxiv-2023-qpf2x (2023).
51. Blair, D. J. *et al.* Automated iterative Csp3–C bond formation. *Nature* **604**, 92–97 (2022).
52. Zhang, G. *et al.* Renewed Prospects for Organic Photovoltaics. *Chem. Rev.* **122**, 14180–14274 (2022).
53. Heitzer, H. M., Marks, T. J. & Ratner, M. A. Molecular Donor–Bridge–Acceptor Strategies for High-Capacitance Organic Dielectric Materials. *J. Am. Chem. Soc.* **137**, 7189–7196 (2015).
54. Bronstein, H., Nielsen, C. B., Schroeder, B. C. & McCulloch, I. The role of chemical design in the performance of organic semiconductors. *Nat. Rev. Chem.* **4**, 66–77 (2020).
55. RDKit: Open-source cheminformatics. https://www.rdkit.org.
56. Knapp, D. M., Gillis, E. P. & Burke, M. D. A General Solution for Unstable Boronic Acids: Slow-Release Cross-Coupling from Air-Stable MIDA Boronates. *J. Am. Chem. Soc.* **131**, 6961–6963 (2009).
57. Refaely-Abramson, S., Baer, R. & Kronik, L. Fundamental and excitation gaps in molecules of relevance for organic photovoltaics from an optimally tuned range-separated hybrid functional. *Phys. Rev. B* **84**, 075144 (2011).



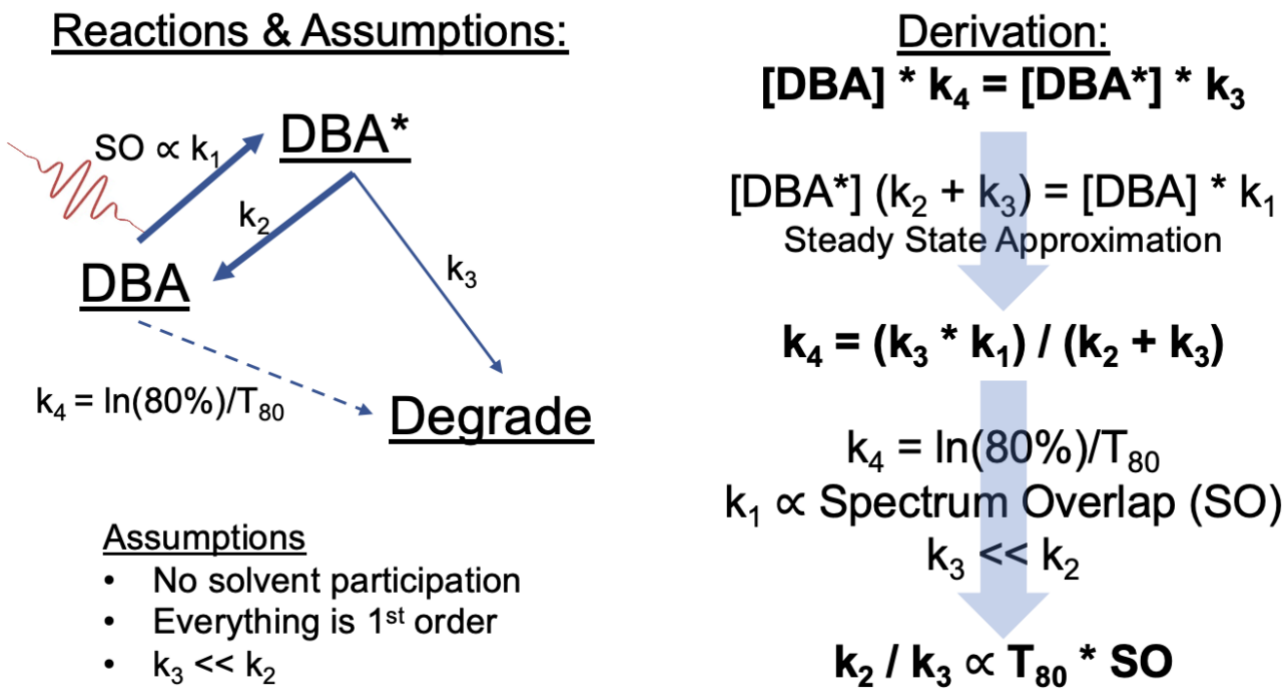

Extended Data Figure 1. Chemical diversity down-selected set of acceptors used in this work.




Extended Data Figure 2. The Streamlit web app digital project manager used in this work.




Extended Data Figure 3. The small molecule synthesizer used in this work. **a,** Picture of the hardware. **b,** Design schematic.

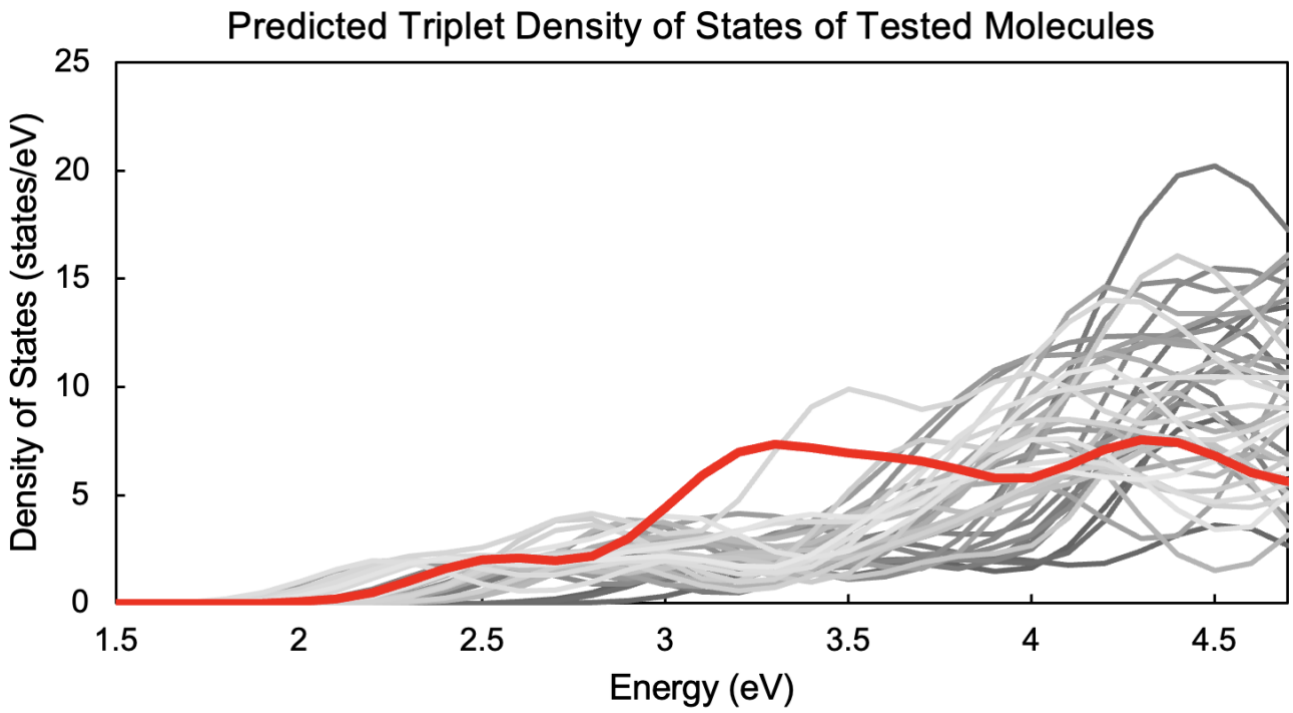



Extended Data Figure 4. Simplified kinetic model to describe oligomer degradation. See further discussion in SI section 1.

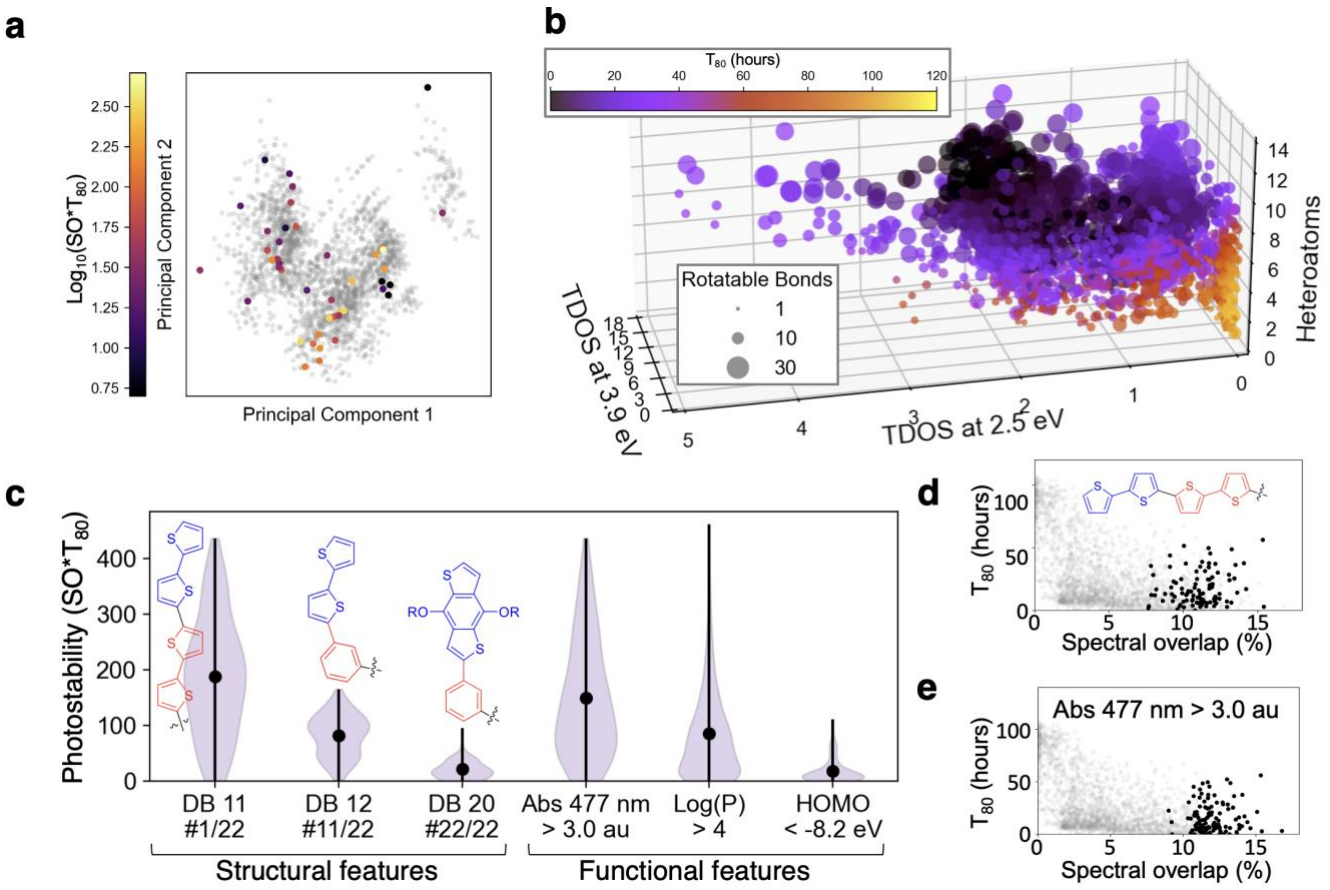
| Round-ID | DBA_Name | SO (%) | $T_{80}$ (hours) | SO*$T_{80}$ |
|---|---|---|---|---|
| 1-d1 | DB_01_A_010 | 0.8 | n/a | n/a |
| 1-d2 | DB_19_A_021 | 12.25 | 3 | 36.75 |
| 1-d3 | DB_15_A_053 | 2.88 | 5 | 14.4 |
| 1-d8 | DB_18_A_094 | 12.2 | 1.5 | 18.3 |
| 1-d9 | DB_05_A_073 | 0.1 | n/a | n/a |
| 1-d10 | DB_20_A_017 | 2.35 | 3.5 | 8.23 |
| 2-1 | DB_05_A_089 | 0.1 | n/a | n/a |
| 2-2s11 | DB_13_A_100 | 6.1 | 51.7 | 315.37 |
| 2-3s3 | DB_08_A_018 | 3.65 | 48 | 175.2 |
| 2-4 | DB_19_A_025 | 14.58 | 1.5 | 21.87 |
| 2-d5s1 | DB_20_A_027 | 4.55 | 15 | 68.25 |
| 2-d8s13 | DB_17_A_069 | 6.84 | 10.6 | 72.5 |
| 2-d9 | DB_12_A_022 | 5.55 | 5 | 27.75 |
| 2-d10 | DB_16_A_061 | 4.46 | 1.9 | 8.47 |
| 3-1 | DB_15_A_088 | 5.19 | 9.3 | 48.27 |
| 3-2 | DB_11_A_002 | 7.2 | 46.7 | 336.24 |
| 3-3s1 | DB_20_A_012 | 2.4 | 9.7 | 23.28 |
| 3-4 | DB_06_A_049 | 1.23 | 78.3 | 96.31 |
| 3-6 | DB_08_A_034 | 6.08 | 2.36 | 14.35 |
| 4-1 | DB_22_A_046 | 5.88 | 10.49 | 61.68 |
| 4-3 | DB_22_A_083 | 6.66 | 23.09 | 153.78 |
| 4-4 | DB_05_A_002 | 0.1 | n/a | n/a |
| 4-5 | DB_01_A_002 | 0.33 | 101.13 | 33.37 |
| 4-6 | DB_06_A_002 | 1.8 | 103.86 | 186.95 |
| 5-1 | DB_10_A_007 | 10.76 | 10.48 | 112.76 |
| 5-2 | DB_11_A_007 | 13.7 | 8.87 | 121.52 |
| 5-3s4 | DB_10_A_091 | 7.87 | 19.48 | 153.31 |
| 5-4s7 | DB_10_A_084 | 8.92 | 15.64 | 139.51 |
| 5-5 | DB_10_A_002 | 12.37 | 8.64 | 106.88 |
| 5-6s2 | DB_09_A_002 | 7.89 | 7.86 | 62.02 |
| Top7-1 | DB_04_A_046 | 6.36 | 24.91 | 158.43 |
| Top7-2 | DB_04_A_070 | 8.23 | 31.94 | 262.87 |
| Top7-3 | DB_13_A_009 | 2.42 | 32.17 | 77.85 |
| Top7-4 | DB_13_A_031 | 3.95 | 10.8 | 42.66 |
| Top7-5 | DB_13_A_044 | 5.45 | 44.96 | 245.03 |
| Top7-6 | DB_13_A_071 | 4.82 | 68.37 | 329.54 |
| Top7-7 | DB_13_A_080 | 6.1 | 6.67 | 40.69 |
| Bot7-1 | DB_08_A_096 | 7.1 | 71.88 | 510.35 |
| Bot7-2 | DB_15_A_066 | 5.46 | 5.34 | 29.16 |
| Bot7-3 | DB_17_A_029 | 7.48 | 2.42 | 18.10 |
| Bot7-4 | DB_22_A_007 | 3.4 | 9.15 | 31.11 |
| Bot7-5 | DB_22_A_018 | 3.14 | 9.77 | 30.68 |
| Bot7-6 | DB_22_A_023 | 4.3 | 4.8 | 20.64 |

| | | | | |
|---|---|---|---|---|
| Bot7-7 | DB_22_A_063 | 3.67 | 10.88 | 39.93 |

Extended Data Table 1. Characterized SO, $T_{80}$, and Photostability ($SO*T_{80}$) of synthesized molecules from Rounds 1-5 and the validation set (Top7 and Bot7). Molecules are numbered by their round number and intra-round ID, where 's' indicates substitute, and 'd' indicates selected via diversity. For molecules recommended by BO (e.g. Rounds 2 through 5 without a 'd' label) a lower intra-round ID corresponds to a more explorative recommendation, a higher number corresponds to more exploitative recommendation.




Extended Data Figure 5. Triplet Density of States (TDOS) for all experimentally measured molecules. DB08_A096 (the high performer in the predicted Bottom 15) is shown in red. All others are in gray, with the highest $T_{80}$ in darker colors, and the lowest $T_{80}$ in lighter colors.




Extended Data Figure 6. **a,** The distribution of photostabilities of the 44 molecules synthesized over the CLT campaign. **b,** The best 4 feature model for predicting $T_{80}$. Note the similarities of the TDOS features to those in the original physics based $T_{80}$ model (**Figure 5a** and S4. **c,** Examples of filtering predicted photostability using structural features (donor-bridge [DB] combinations) and functional features (calculated properties, independent of model training). DBs are ranked according to average predicted photostability values across all 100 acceptors. Functional features (Abs: absorbance, Log(P): octanol water partition coefficient), are filtered to contain ~100 molecules each, demonstrating that functional features can be equally predictive as structural features. **d-e**, individual predicted $T_{80}$ and SO plots for a structural search (DB 11) and a functional search (high absorbance at 477 nm). Black points correspond to donor-bridge-acceptor molecules that fulfill these criteria. Note that the molecule population shown in **e** contains six DB motifs.