Article

Closed-loop transfer enables artificial intelligence to yield chemical knowledge

https://doi.org/10.1038/s41586-024-07892-1

Received: 24 September 2023

Accepted: 31 July 2024

Published online: 28 August 2024

Check for updates

Nicholas H. Angello[1,2,3,16], David M. Friday[1,2,3,16], Changhyun Hwang[2,3,4,16], Seungjoo Yi[2,3,5,16], Austin H. Cheng[6,7], Tiara C. Torres-Flores[2,3,4], Edward R. Jira[2,3,4], Wesley Wang[1,2,3], Alán Aspuru-Guzik[6,7,8,9,10,11,12] ✉, Martin D. Burke[1,2,3,13,14,15] ✉, Charles M. Schroeder[1,2,3,4,5] ✉, Ying Diao[1,2,3,4] ✉ & Nicholas E. Jackson[1,2,3] ✉

Artificial intelligence-guided closed-loop experimentation has emerged as a promising method for optimization of objective functions[1,2], but the substantial potential of this traditionally black-box approach to uncovering new chemical knowledge has remained largely untapped. Here we report the integration of closed-loop experiments with physics-based feature selection and supervised learning, denoted as closed-loop transfer (CLT), to yield chemical insights in parallel with optimization of objective functions. CLT was used to examine the factors dictating the photostability in solution of light-harvesting donor–acceptor molecules used in a variety of organic electronics applications, and showed fundamental insights including the importance of high-energy regions of the triplet state manifold. This was possible following automated modular synthesis and experimental characterization of only around 1.5% of the theoretical chemical space. This physics-informed model for photostability was strengthened using multiple experimental test sets and validated by tuning the triplet excited-state energy of the solvent to break out of the observed plateau in the closed-loop photostability optimization process. Further applications of CLT to additional materials systems support the generalizability of this strategy for augmenting closed-loop strategies. Broadly, these findings show that combining interpretable supervised learning models and physics-based features with closed-loop discovery processes can rapidly provide fundamental chemical insights.

Artificial intelligence-guided closed-loop platforms in which predictions, experiments and analyses are automated and connected in a positive feedback loop have shown great potential in the acceleration of scientific discovery in intractably large search spaces[3–6]. Despite recent advances, it is not yet possible to leverage closed-loop optimization strategies to elicit fundamentally new chemical knowledge. At the frontiers of molecular function, such fundamental understanding is of equal importance to the practical results of artificial intelligence-guided optimization strategies and is critical for aligning artificial intelligence-guided discovery with the human scientific process.

Photostability represents a ubiquitous chemical function for which general chemical design principles are lacking[7–11]. This dearth of chemical knowledge limits progress in the areas of organic photovoltaics[12,13], dyed polymers[14], solar fuels[3,15], photosynthetic systems[16], electrochromic materials[17], organic light-emitting diodes[18], photoactive coatings[19] and fluorescent dyes[20,21]. Previous efforts to understand molecular photostability have focused on the energetics of the lowest-lying excited triplet state (T_1) and its relation to bond dissociation energy, with limited success across scattered chemical classes[22–29]. Recent studies have implicated higher energy triplet states (T_n, $n > 1$) for narrow chemical classes[30,31], but general design principles remain lacking[32]. Closed-loop paradigms are poised to enable breakthroughs in photostability in which traditional approaches have failed[23]. To realize this vision, new methods of knowledge extraction from closed-loop strategies are required before artificial intelligence can yield interpretable hypotheses and enhance our fundamental understanding of photostability and molecular function in general.

Here we report a three-phase approach that demonstrates chemical knowledge while optimizing molecular function across a broad

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]Molecule Maker Lab, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Molecule Maker Lab Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [4]Department of Chemical & Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [5]Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [6]Department of Chemistry, University of Toronto, Toronto, Ontario, Canada. [7]Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada. [8]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. [9]Canadian Institute for Advanced Research, Toronto, Ontario, Canada. [10]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada. [11]Department of Materials Science and Engineering, University of Toronto, Toronto, Ontario, Canada. [12]Acceleration Consortium, Toronto, Ontario, Canada. [13]Carle R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [14]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [15]Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [16]These authors contributed equally: Nicholas H. Angello, David M. Friday, Changhyun Hwang, Seungjoo Yi. ✉e-mail: alan@aspuru.com; mdburke@illinois.edu; cms@illinois.edu; yingdiao@illinois.edu; jacksonn@illinois.edu

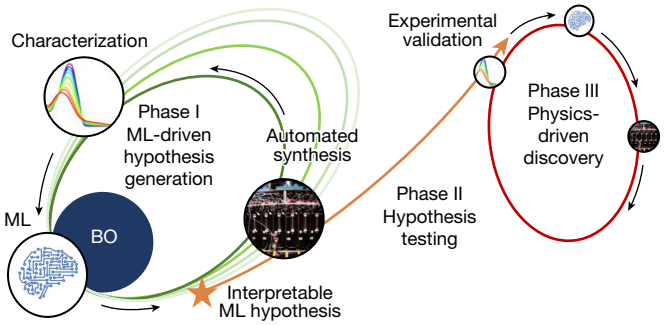


Fig. 1 | The CLT paradigm. Phase I (ML-driven hypothesis generation, green orbits) depicts BO-driven rounds of synthesis and characterization with concurrent use of interpretable ML using physics-based features. Following the emergence of a physics-based hypothesis via ML, phase II (hypothesis testing, orange orbit) experimentally tests the hypothesis, which, if validated, yields new knowledge that is exploited in phase III (physics-driven discovery, red orbit) for molecular optimization.

chemical space (Fig. 1). Phase I is machine learning (ML)-driven hypothesis generation: apply Bayesian optimization (BO) to improve photostability until the performance metric plateaus and an ML-derived hypothesis using physics-based molecular features emerges. Phase II is hypothesis testing: experimentally validate the ML-derived hypothesis to establish new-found chemical knowledge. Phase III is physics-driven discovery: apply the new physics-based knowledge to the chemical design space in a human-driven campaign to break out of the optimization plateau.

We show that artificial intelligence-guided closed-loop experimentation leads to the generation of a supervised learning model that not only identifies highly photostable compounds but also elucidates design rules for molecular photostability owing to interpretable physics-informed feature selection. Central to our approach is a BO-driven closed-loop experimental procedure leveraging automated modular small molecule synthesis that enables rapid search through a diverse chemical space to optimize photostability (phase I). The closed-loop strategy is performed in concert with physics-based modelling and feature selection, resulting in the unexpected finding that molecular photostability is strongly correlated with the high-energy portion of the triplet excited-state manifold—a hypothesis we further validate using an experimental test set (phase II). These results were achieved using BO-recommended molecules with no pre-existing knowledge of the determinants of photostability in organic molecules. Furthermore, this result launches us into a physics-driven discovery regime wherein we rationally improve molecular photostability, leveraging the newly discovered degradation pathway via Dexter triplet energy transfer from the solvent (phase III). Our work shows that interfacing physics-based modelling with the data emerging from BO-guided closed-loop discovery can deliver physical insights into frontier molecular functions. We further apply our approach beyond photostability to generate hypotheses underpinning the design of organic laser emitters[33] and stereoselective aluminium complexes for ring-opening polymerization[34] (Supplementary Information 1).

We introduce an approach called closed-loop transfer (CLT; Fig. 1) to extract and test physical insights from a closed-loop optimization campaign. CLT is a 'human-in-the-loop' approach that leverages domain expertise at the beginning of a closed-loop process, monitoring each optimization round with stopping criteria based on the simultaneous plateauing of an objective function and identification of a physics-based hypothesis by on-the-fly generation of interpretable ML models. When these criteria are met, an experimental test set is constructed across the entire chemical design space to assess the truth of the proposed hypothesis. If validated, this new-found chemical knowledge then seeds a human-driven campaign to break out of the plateau in the phase I closed-loop optimization stage. A CLT playbook is provided in Supplementary Information 1.

In this campaign we focus specifically on uncovering the fundamental determinants of photostability broadly across chemical space. Phase I integrates BO (GRYFFIN[35]) in a closed-loop process with automated modular small molecule synthesis[36–38] and multidimensional characterization (in this case solution-based, ultraviolet-visible absorbance and photodegradation in a solar irradiation cell[12]) run in parallel with on-the-fly ML modelling and molecular feature generation. This phase continues until the photostability plateaus and interpretable ML models using physics-based features emerge via a predictive model, referred to as the 'hypothesis', which is validated in phase II. In this work we focus on the molecular photostability of conjugated molecules in the solution state, intentionally avoiding additional complicating factors associated with film testing (that is, processing, film morphology and interfacial effects).

To initialize the CLT process, we first defined the chemical space (Fig. 2) for the closed-loop procedure (Fig. 3a). We chose the donor–bridge–acceptor motif, ubiquitous in state-of-the-art light-absorbing conjugated materials[39–41], 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 (Supplementary Information 2), maximizing molecular diversity within all purchasable (hetero)aryl-halide building blocks (Extended Data Fig. 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 molecules 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[42] (Supplementary Tables 2 and 3).

Critical to initialization of the CLT process was an early emphasis on diversity, followed by balanced exploration and exploitation in subsequent rounds (Fig. 3b). To initiate the first round, molecular diversity sampling was utilized to decrease initiation bias in the closed loop (Supplementary Information 2). The second round contained additional diversity-selected molecules and BO recommendations. For each subsequent closed-loop iteration a batch of six molecules was drawn for uniform sampling, along the domain between fully exploitative and fully explorative within the BO algorithm[35]. This strategy ensured that, although 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 neighbours in feature space (known as 'back-ups' in cases in which the target is not readily synthesizable or testable; Fig. 3b) were then automatically populated on a custom-built digital project manager with a web-based database and dashboard for subsequent automated synthesis (Extended Data Fig. 2).

During 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 (Fig. 3c and Extended Data Fig. 3)[4]. Initial tests demonstrated that a fully automated two-step synthesis was feasible. The first step proved to be generally efficient, whereas the second was initially more challenging and variable due to chemical diversity in the 100 aryl-halide acceptor blocks. We thus opted for independent synthesis and scale-up of the first coupling products (donor–bridges; Supplementary Information 3), and

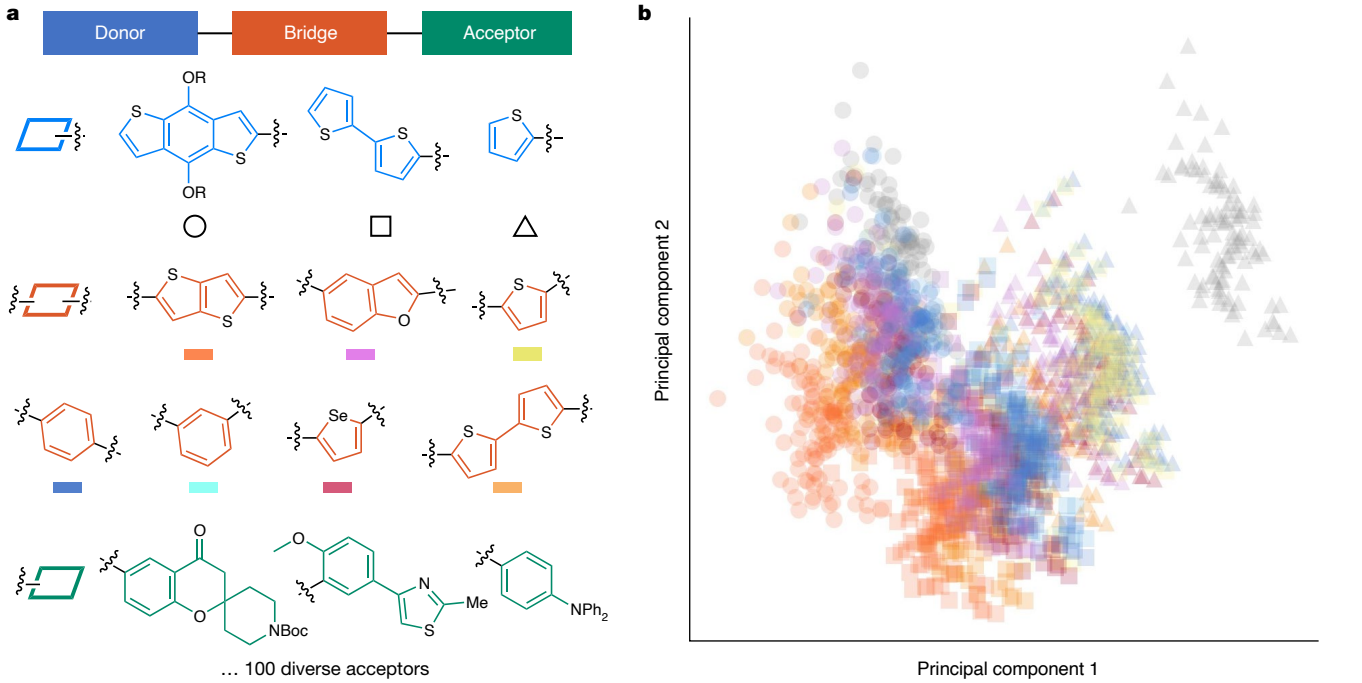


Fig. 2 | A molecular building-block set for light-harvesting small molecules. **a**, The modular building blocks considered in this work for the design of new light-harvesting donor–bridge–acceptor molecules. The full list of 100 acceptor building blocks is shown in Extended Data Fig. 1. **b**, Principal component analysis projection of the feature space (2,200 molecules featurized by concatenated donor–bridge and donor–acceptor features) used for BO. Alternative uniform manifold approximation and projections are provided in Supplementary Fig. 22. Shape and colour coding correspond to the structures of the donor and bridge building blocks, respectively. Grey colouring indicates the absence of a bridge block—that is, molecules that are donor–acceptor. The principal component analysis projection reveals the primacy of the donor block in organizing the chemical space, with benzodithophene (BDT) donor blocks (circles) in the top left corner and thiophene-based donor blocks (triangles and squares) in the bottom right regions. Similarly, a greater extent of conjugation in the bridge blocks correlates with lower values of principal component 1 and principal component 2 (lower left), further organizing the featurized chemical space.

separate optimization of the second reaction step using slow-release cross-coupling[43]. We found that, using our recently reported general reaction conditions for heteroaryl cross-coupling discovered via an artificial intelligence-guided closed-loop process[4] (general condition 1 (GC1); Fig. 3c), and newly discovered anhydrous slow-release coupling conditions (GC2) (Fig. 3c and Supplementary Information 3)[37] maximized the synthetic hit rate (approximately 60%).

Following synthesis, purification, and structural verification 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 (Fig. 3d). The photophysical properties of 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: (1) spectral overlap, defined as the integral of the normalized overlap of a molecule's absorbance spectrum and solar irradiance spectrum, and (2) 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 (Supplementary Information 4). Based on a first-order kinetic model we chose to optimize photostability, defined as the product of spectral overlap and T_{80} (Supplementary Fig. 5).

BO-driven closed-loop experimentation proceeded across five rounds, automatically synthesizing 30 new donor–bridge–acceptor light-harvesting molecules (Fig. 4a and Extended Data Table 1), until saturation of experimental photostability was observed. The first round of suggestions, relying on diversity-driven selection, resulted in a set of ten molecules with low to moderate photostability. 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 spectral overlap and some that maximized their product. By the conclusion of the fifth round of the closed loop the average photostability of the top five molecules had plateaued, signalling the end of the BO-guided closed-loop process (Fig. 4b). Importantly, an increase of over 500% in the average photostability of the top five performers was achieved by sampling less than 1.5% of the total space of 2,200 potential molecules, a result consistent with previous theoretical predictions but hitherto not verified experimentally[35].

In conjunction with the progress of the BO experiment, we trained interpretable ML models drawing on physics-based features following each BO round to generate hypotheses relating molecular features to photostability. To generate physics-based features we used whole-molecule, time-dependent DFT calculations[44] on donor–bridge–acceptor and donor–acceptor molecules synthesized following each closed-loop round, and from these results a comprehensive set of 114 physical and chemical molecular features were extracted (Supplementary Tables 2–5). These features were then integrated in support vector regression (SVR; Supplementary Table 1) following every BO round to predict experimental T_{80} values. SVR models with stepwise feature selection were selected due to their model interpretability and trained to predict T_{80} separately from spectral overlap due to an observed inverse relationship between T_{80} and spectral overlap (Supplementary Information 2). Spectral overlap was predicted via linear regression using the time-dependent, DFT-predicted absorption spectrum, solar simulator emission spectrum and a fitted constant energy shift due to solvent interactions (Supplementary Fig. 4). The best T_{80} model achieved leave-one-out-validation (LOOV) predictive accuracy of $R^2 = 0.86$ for experimental T_{80} following the fifth round using the triplet density of states (TDOS: density of triplet excited states at a particular

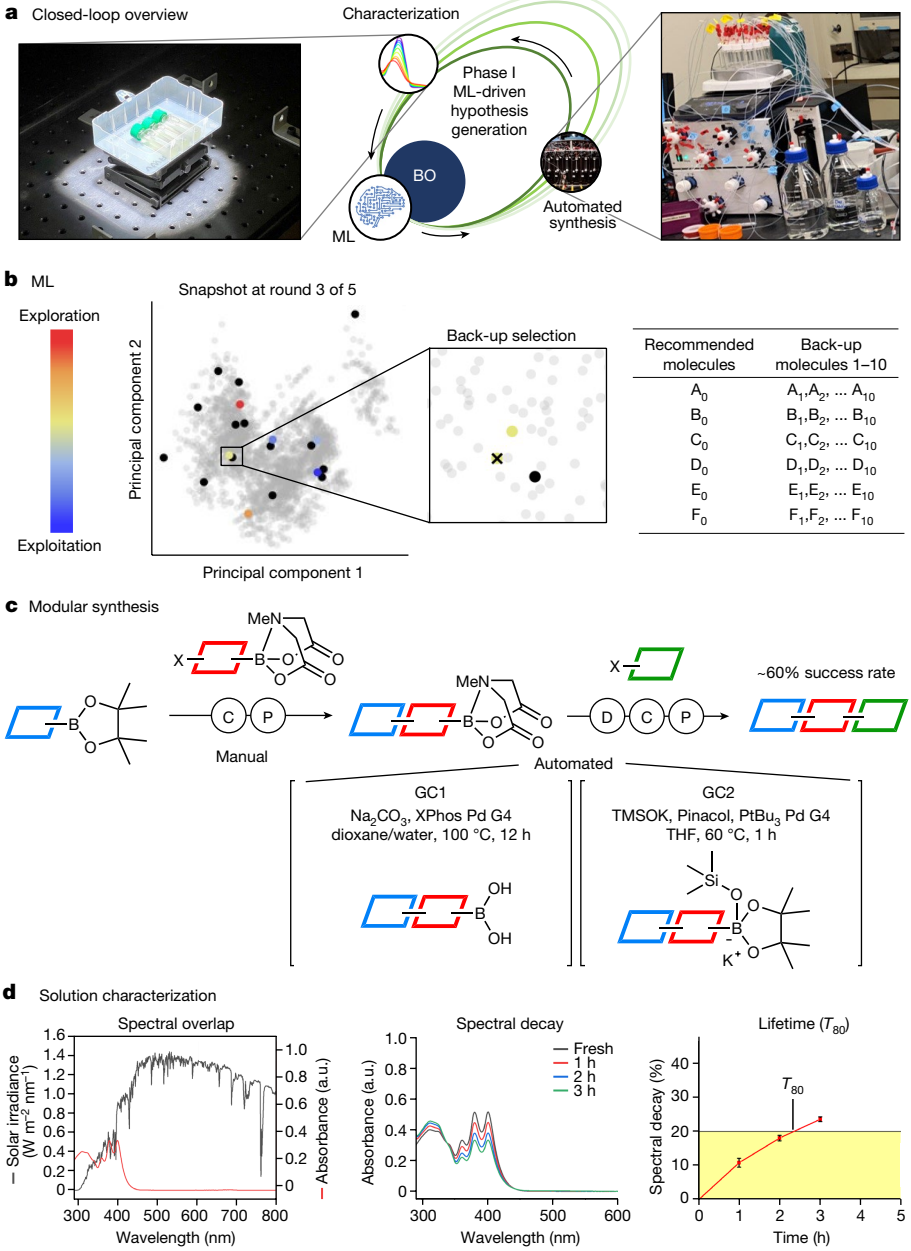


Fig. 3 | Closed-loop optimization in phase I. a, The major steps in the optimization portion of CLT phase I under the guidance of artificial intelligence-driven BO, and images of the solution characterization process and automated synthesis equipment. **b**, Visualization of round 3 of closed-loop optimization in which six molecules are recommended, balancing exploration and exploitation, along with back-up molecules. In round 3 the third-most explorative molecule (yellow) was unsynthesizable, requiring the use of a back-up, as shown in the inset. **c**, Reaction conditions used in the automated synthesis of roughly 60% of molecules recommended by BO. **d**, Example data generated by 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} for each molecule with measurable absorbance. a.u., arbitrary units; C, coupling; D, deprotection; P, purification; THF, tetrahydrofuran; TMSOK, potassium trimethylsilanolate.

energy above ground state) at 2.6, 2.8, 3.8, 3.9, 4.0 and 4.6 eV. We also explored all possible SVR models (12,996) using two-feature combinations of the 114-feature set and LOOV performance metric (Extended Data Fig. 5a). 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; Supplementary Fig. 9). This insight emerged following the fourth round of BO and was confirmed following the fifth (Supplementary Table 1). Because the ML models' convergence coincided with the plateau of molecular photostability, this signalled the end of our BO experiment. It is important to note that the positive aspect of CLT in this work is inherently linked to BO recommending molecules with sufficient representativeness for identification of a general physics-based hypothesis, which may occur either quickly or slowly (Supplementary Information 1).

We then examined all 234,136 four-feature SVR models containing the most predictive region of the TDOS at 4.0 eV or T_1 energy, and plotted the distribution of performance (R^2; Extended Data Fig. 5b). Our results show a marked difference in the predictive capabilities of supervised ML models using the high-energy region of TDOS as opposed to T_1. Importantly, this knowledge emerged a priori across a broad chemical space, with equal weight given to all possible hypotheses described

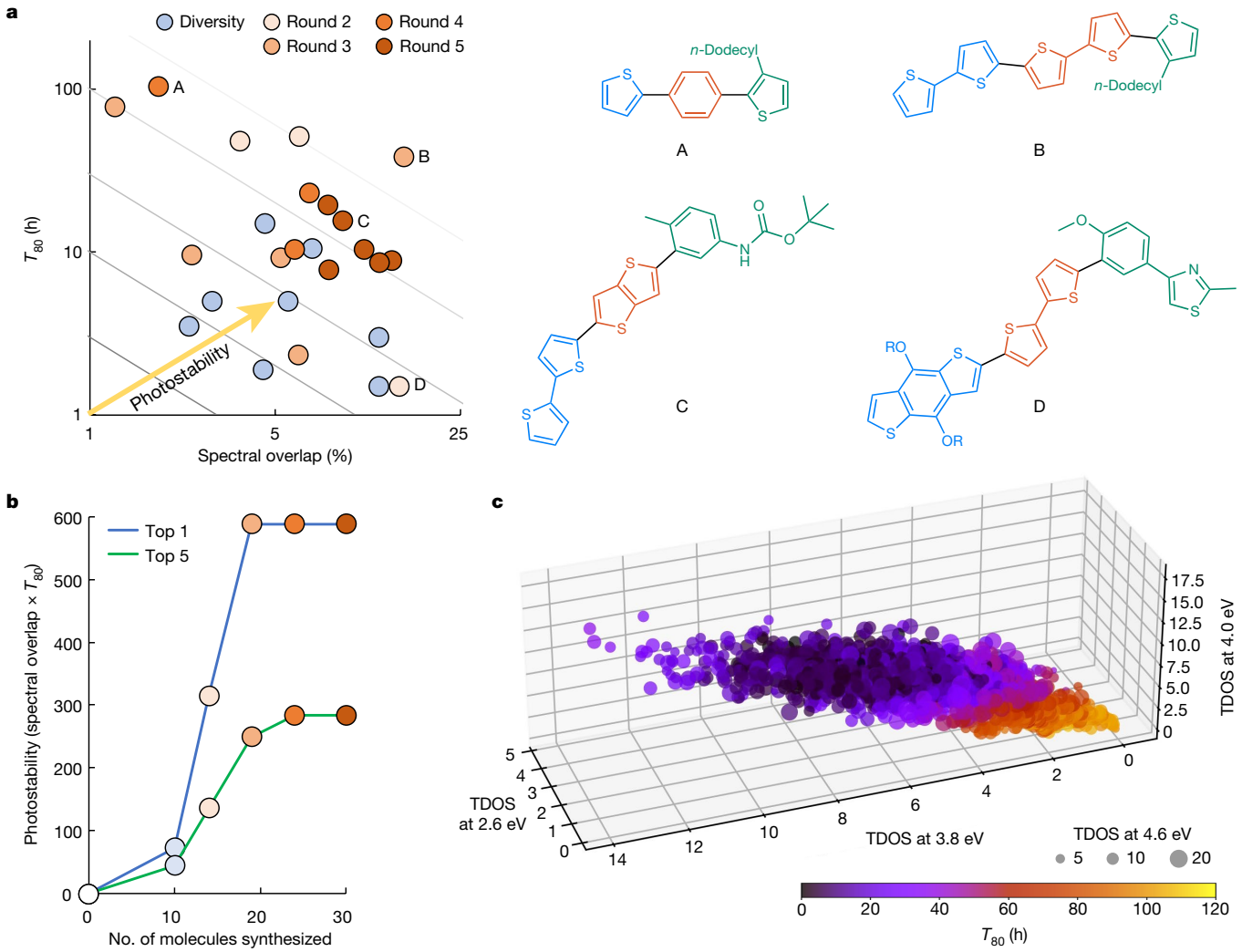


Fig. 4 | ML-driven hypothesis generation in phase I. a, Data generated per round of closed-loop experimental BO, with representative highly photostable molecules shown. **b**, Cumulative photostability performance per round of closed-loop experimentation. Top 5, average of the top five molecules across all rounds. **c**, Physics-based ML hypothesis for prediction of T_{80} using TDOS of each molecule is plotted against four triplet TDOS energies relevant to the prediction task. This model, along with spectral overlap criteria, was used to select the top seven and bottom seven molecules for hypothesis testing in phase II.

by 114 physiochemical descriptors. This data-driven physical insight suggests that a more nuanced understanding of the role of the triplet state manifold in molecular photostability is required that extends beyond simple T_1 energetics.

Because learning in the low-data limit can induce spurious correlations in supervised ML models, phase II of the CLT experiment involved construction of an independent experimental test set to validate the hypothesis that high-lying triplet states influence molecular photostability. The best T_{80} SVR model trained on six values of TDOS (Supplementary Fig. 9) was used to predict T_{80} across the entire set of 2,200 molecules (Fig. 4c). The predicted T_{80} value of each molecule was then multiplied by its predicted spectral overlap value to obtain predicted photostabilities across the entire set of 2,200 molecules. Using these predictions we formed two batches of seven molecules, one high performing (top seven) and one low performing (bottom seven), to serve as experimental validation sets (Supplementary Fig. 11). These two batches possessed the following statistical features (Supplementary Information 2): (1) identical average spectral overlap within the 5.5–9.0% SO region (which emphasizes the effects of T_{80} on photostability) and (2) similar standard deviation for the predicted T_{80} (which permits broader chemical diversity). We then synthesized and characterized the photophysical properties of the top seven and bottom seven batches. We observed a statistically significant photostability difference between these groups in the predicted direction (Fig. 5a; average T_{80} × spectral overlap = 165 for the top seven versus average T_{80} × spectral overlap = 97 for the bottom seven, $P = 0.026$, Spearman $R^2 = 0.54$), validating our CLT-derived hypothesis that the triplet manifold is a key descriptor of molecular photostability. Interestingly, one outlier in the bottom seven showed surprisingly high photostability; subsequent analysis demonstrated that this quinone-based molecule (a common motif in triplet quenchers) exhibited high TDOS while maintaining a low structural complexity—a feature unseen relative to the 30 molecules synthesized in the five closed-loop rounds (Extended Data Fig. 4).

Using the full experimental photostability dataset generated by the CLT campaign (44 molecules; Extended Data Fig. 6a), we performed a final retraining of supervised ML models to confirm and refine our photostability hypothesis. Strikingly, following retraining of all 12,996 two-feature SVR models when considering the full experimental dataset, high-energy TDOS emerged even more strongly as a critical determinant of molecular photostability across the synthesized chemical space (high T_{80} LOOV R^2; Fig. 5b) and a superior descriptor relative

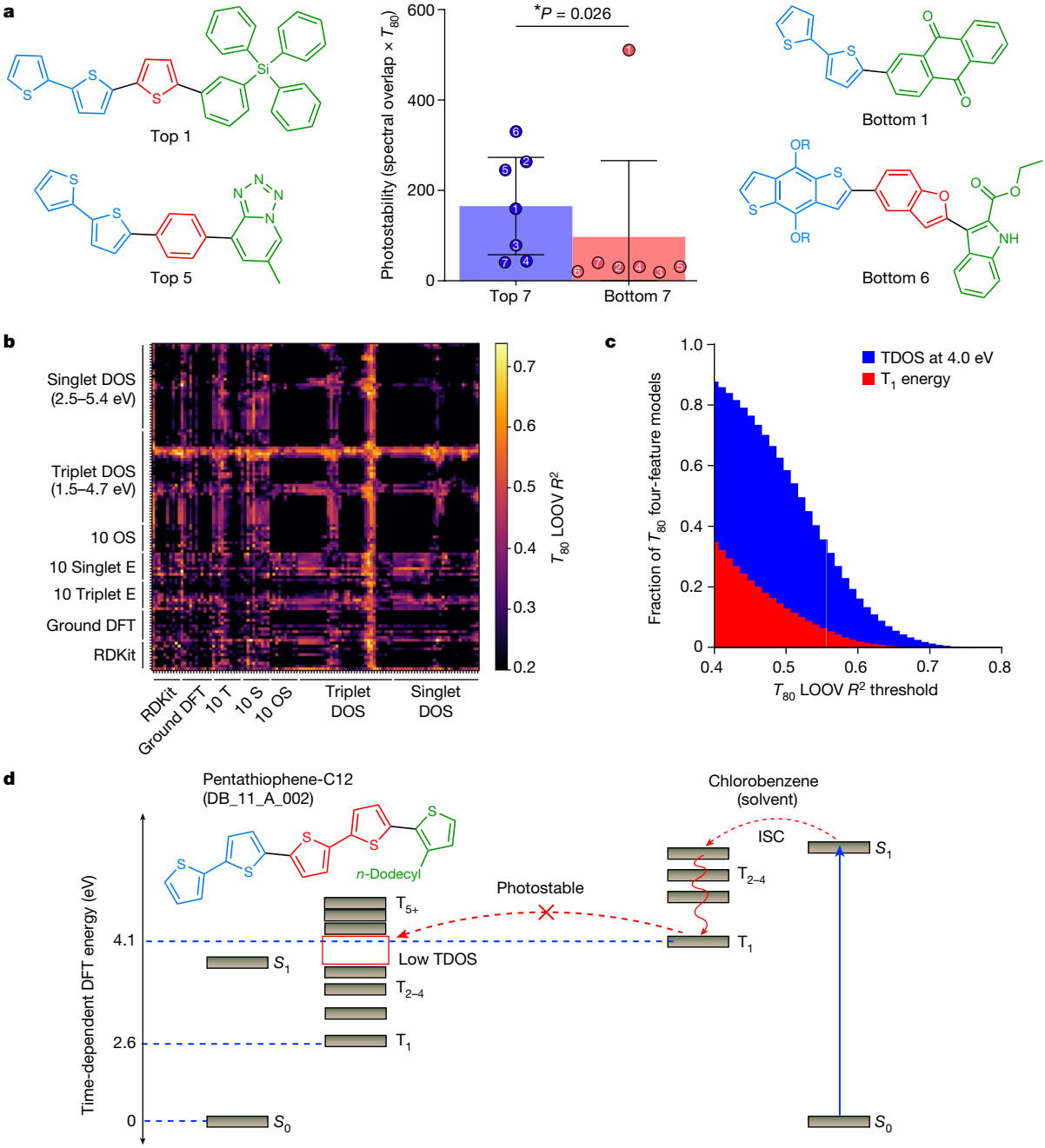


Fig. 5 | Hypothesis testing in phase II. a, Photostability of molecules in the experimental validation test set, their average, standard deviation (depicted as error bars) and associated *P* value from a Mann–Whitney test plotted in the centre. These values demonstrate a statistically significant difference in performance between the two sets of molecules, with the outlier high performer (bottom two) as described in the text. **b**, All two-feature SVR LOOV results for prediction of T_{80} on the entire 44-molecule dataset from two-feature combinations. E, energies; OS, oscillator strengths; S, singlet energies; T, triplet energies. **c**, Comparison of the prediction strength of all possible four-feature models containing either TDOS at 4.0 eV or T_1 energy in predicting T_{80} for the entire 44-molecule dataset. Comparison of the plots in **b**,**c** with Extended Data Fig. 5 shows the improved predictive strength of TDOS at 4.0 eV over all other features, in particular T_1, following the validation dataset. **d**, Physical mechanism derived from the CLT phase II hypothesis, demonstrating how a low TDOS at 4.0 eV reduces the frequency of Dexter energy transfer following intersystem crossing (ISC), imparting enhanced photostability to pentathiophene-C12 (DB_11_A_002). S_0, singlet ground state; S_1, the lowest energy singlet excited state.

to conventional T_1 energy (Fig. 5c). A total of 2.5 million four-feature SVR models were trained to identify the most common features present in the most predictive T_{80} models ($R^2 > 0.70$) across all descriptors (Supplementary Fig. 12). The two most common features, present in around 30% of the most predictive models, were TDOS at 4.0 eV and the number of heteroatoms. 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} consolidates closed-loop-derived chemical knowledge that the high-lying TDOS is a critical determinant of molecular photostability.

Because high-lying TDOS will be only transiently populated by intersystem crossing via direct excitation in the singlet manifold of the molecules, we posited that the photodegradation mechanism is mediated by chlorobenzene solvent (Fig. 5d) via resonant Dexter energy transfer. The resonance of the T_1 state of chlorobenzene (4.1 eV) is near that of the CLT-identified TDOS region at 4.0 eV, implying that the solvent could

act as an efficient sensitizer for populating the triplet states of DBA molecules, similar to work showing that Dexter transfer between the T_1 state of one molecule the excited triplet manifold of another leads to efficient triplet state transfer[45]. These high-lying triplet states are then anticipated to relax to the T_1 state in which any number of potential degradation mechanisms could be active. We anticipate the role of high-lying triplet states to be important in many multicomponent organic systems under solar irradiation. Beyond organic photovoltaics[46–48], these may also include dyed polymers[14], solar fuels[3,15] and photosynthetic systems[16].

To apply this new-found chemical knowledge towards improvement of photostability, in phase III we tested three molecules with varying TDOS at 4.0 eV in toluene (a more chemically stable solvent than chlorobenzene with comparable T_1 energy) and decane (with an inaccessible T_1 greater than 7 eV) to intentionally remove the potential for Dexter triplet energy transfer from the solvent. The photostability of all molecules improved by 90–150% in toluene, with an additional 10–100% improvement in decane and greater improvements for molecules with higher TDOS at 4.0 eV (Extended Data Fig. 6c). Further improvements and understanding of photostability were achieved by the addition of a triplet-quenching molecule to the solvent, yielding (1) 20% improvement in chlorobenzene, as well as direct chemical bonding of a triplet quencher to the highest-performing molecule via a hexyl linker, and (2) 75% poorer photostability in chlorobenzene via a hypothesized double-Dexter transfer mechanism (Extended Data Fig. 6d). These results are fully consistent with the proposed solvent sensitization mechanism for photostability (Fig. 5d and Supplementary Figs. 71–74), and demonstrate multiple phase III hypothesis-driven strategies to improve the photostability of light-absorbing molecules beyond the initial chemical space.

Conclusion

The closed-loop experiment reported here, augmented by interpretable ML models using physics-based features, elucidated fundamental chemical knowledge regarding molecular photostability while simultaneously optimizing towards high-function molecular targets. The marked efficiency with which our human-in-the-loop, artificial intelligence-guided campaign yielded our hypothesis (only 30 molecules) resulted from the combination of an artificial intelligence-guided balance of exploration and exploitation with interpretable physics-based modelling. CLT is shown to be broadly applicable to other frontier applications and fields of inquiry, in particular to low-data regimes and multidimensional molecular properties that are challenging to predict a priori. We believe that CLT will serve as a playbook for harnessing the strengths of BO in hypothesis-driven discovery investigations, reinforced by physics-based insights.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07892-1.

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. Burger, B. et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
4. Angello, N. H. et al. Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. *Science* **378**, 399–405 (2022).
5. Koscher, B. A. et al. Autonomous, multiproperty-driven molecular discovery: from predictions to measurements and back. *Science* **382**, eadi1407 (2023).
6. Chang, J. et al. Efficient closed-loop maximization of carbon nanotube growth rate using Bayesian optimization. *Sci. Rep.* **10**, 9040 (2020).
7. 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).
8. Mateker, W. R. & McGehee, M. D. Progress in understanding degradation mechanisms and improving stability in organic photovoltaics. *Adv. Mater.* **29**, 1603940 (2017).
9. Liu, Z.-X. et al. Molecular insights of exceptionally photostable electron acceptors for organic photovoltaics. *Nat. Commun.* **12**, 3049 (2021).
10. 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).
11. 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).
12. Burlingame, Q. et al. Intrinsically stable organic solar cells under high-intensity illumination. *Nature* **573**, 394–397 (2019).
13. Burlingame, Q., Ball, M. & Loo, Y.-L. It's time to focus on organic solar cell stability. *Nat. Energy* **5**, 947–949 (2020).
14. Kuramoto, N. in *Physico-Chemical Principles of Color Chemistry* (eds Peters, A. T. & Freeman, H. S.) 196–253 (Springer Netherlands, 1996); https://doi.org/10.1007/978-94-009-0091-2_6.
15. Kosco, J. et al. Enhanced photocatalytic hydrogen evolution from organic semiconductor heterojunction nanoparticles. *Nat. Mater.* **19**, 559–565 (2020).
16. Zbyradowski, M. et al. Triplet-driven chemical reactivity of β-carotene and its biological implications. *Nat. Commun.* **13**, 2474 (2022).
17. 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).
18. 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).
19. Kim, J.-H. et al. Hydrophilic/hydrophobic silane grafting on TiO_2 nanoparticles: photocatalytic paint for atmospheric cleaning. *Catalysts* **11**, 193 (2021).
20. 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).
21. 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).
22. 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).
23. 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).
24. 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).
25. 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).
26. Distler, A. et al. Effect of PCBM on the photodegradation kinetics of polymers for organic photovoltaics. *Chem. Mater.* **24**, 4397–4405 (2012).
27. Zheng, Q. & Lavis, L. D. Development of photostable fluorophores for molecular imaging. *Curr. Opin. Chem. Biol.* **39**, 32–38 (2017).
28. Meng, Q.-Y. et al. Longevity gene responsible for robust blue organic materials employing thermally activated delayed fluorescence. *Nat. Commun.* **14**, 3927 (2023).
29. Li, H. et al. Chemical stabilities of the lowest triplet state in aryl sulfones and aryl phosphine oxides relevant to OLED applications. *Chem. Mater.* **31**, 1507–1519 (2019).
30. Zhang, K. et al. New insights about the photostability of DNA/RNA bases: triplet ππ* state leads to effective intersystem crossing in pyrimidinones. *J. Phys. Chem. B* **125**, 2042–2049 (2021).
31. 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).
32. Mukherjee, S., Fedorov, D. A. & Varganov, S. A. Modeling spin-crossover dynamics. *Annu. Rev. Phys. Chem.* **72**, 515–540 (2021).
33. Strieth-Kalthoff, F. et al. Delocalized, asynchronous, closed-loop discovery of organic laser emitters. *Science* **384**, eadk9227 (2024).
34. Wang, X. et al. Bayesian-optimization-assisted discovery of stereoselective aluminum complexes for ring-opening polymerization of racemic lactide. *Nat. Commun.* **14**, 3647 (2023).
35. 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).
36. Li, J. et al. Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).
37. Wang, W. et al. Rapid automated iterative small molecule synthesis. *Nat. Synth.* **3**, 1031–1038 (2024).
38. Blair, D. J. et al. Automated iterative Csp3–C bond formation. *Nature* **604**, 92–97 (2022).
39. Zhang, G. et al. Renewed prospects for organic photovoltaics. *Chem. Rev.* **122**, 14180–14274 (2022).
40. 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).
41. 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).
42. Landrum, G. RDKit: open-source cheminformatics software, version 2021_09_01 (Q3 2021) (accessed 12 August 2024); www.rdkit.org.

43. 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).
44. 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).
45. Abroshan, H., Coropceanu, V. & Brédas, J.-L. Hyperfluorescence-based emission in purely organic materials: suppression of energy-loss mechanisms via alignment of triplet excited states. *ACS Mater. Lett.* **2**, 1412–1418 (2020).
46. Li, Y. et al. Non-fullerene acceptor organic photovoltaics with intrinsic operational lifetimes over 30 years. *Nat. Commun.* **12**, 5419 (2021).
47. Yousif, E. & Haddad, R. Photodegradation and photostabilization of polymers, especially polystyrene: review. *Springerplus* **2**, 398 (2013).
48. Fan, B. et al. Understanding the role of removable solid additives: selective interaction contributes to vertical component distributions. *Adv. Mater.* **35**, 2302861 (2023).

Data availability

The data that support the findings of this study are available online (https://github.com/TheJacksonLab/ClosedLoopTransfer), including experimental photostability measurement data for all molecules synthesized in this work, DFT- and RDKit-derived molecular featurizations and predicted photostabilities across the full space of 2,200 molecules. Regression models and scripts used to train and perform all analysis with the associated data are also provided. Datasets are available at Zenodo (https://doi.org/10.5281/zenodo.11580889)[49]. Source Data are provided with this paper.

Code availability

All supervised learning model codes are available at GitHub (https://github.com/TheJacksonLab/ClosedLoopTransfer). Gryffin is available at GitHub (https://github.com/aspuru-guzik-group/gryffin). Codes are available at Zenodo (https://doi.org/10.5281/zenodo.11580889)[49].

49. TheJacksonLab. TheJacksonLab/ClosedLoopTransfer: ClosedLoopTransfer v1.0. *Zenodo* https://doi.org/10.5281/zenodo.11580889 (2024).

Acknowledgements This work was supported by the Molecule Maker Lab Institute, an AI Research Institutes programme supported by the US National Science Foundation under grant no. 2019897 (to N.E.J., Y.D., C.M.S. and M.D.B.). A.A.-G. and A.H.C. acknowledge support from the Canada 150 Research Chairs Program and the Acceleration Consortium at the University of Toronto, as well as the generous support of A. G. Frøseth. T.C.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 US National Science Foundation.

Author contributions N.H.A., D.M.F., C.H. and S.Y. contributed equally to this project. A.H.C. and T.C.T.-F. contributed equally to this project. 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.C.T.-F., E.R.J. and W.W. Solution testing was conducted by C.H. BO 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.

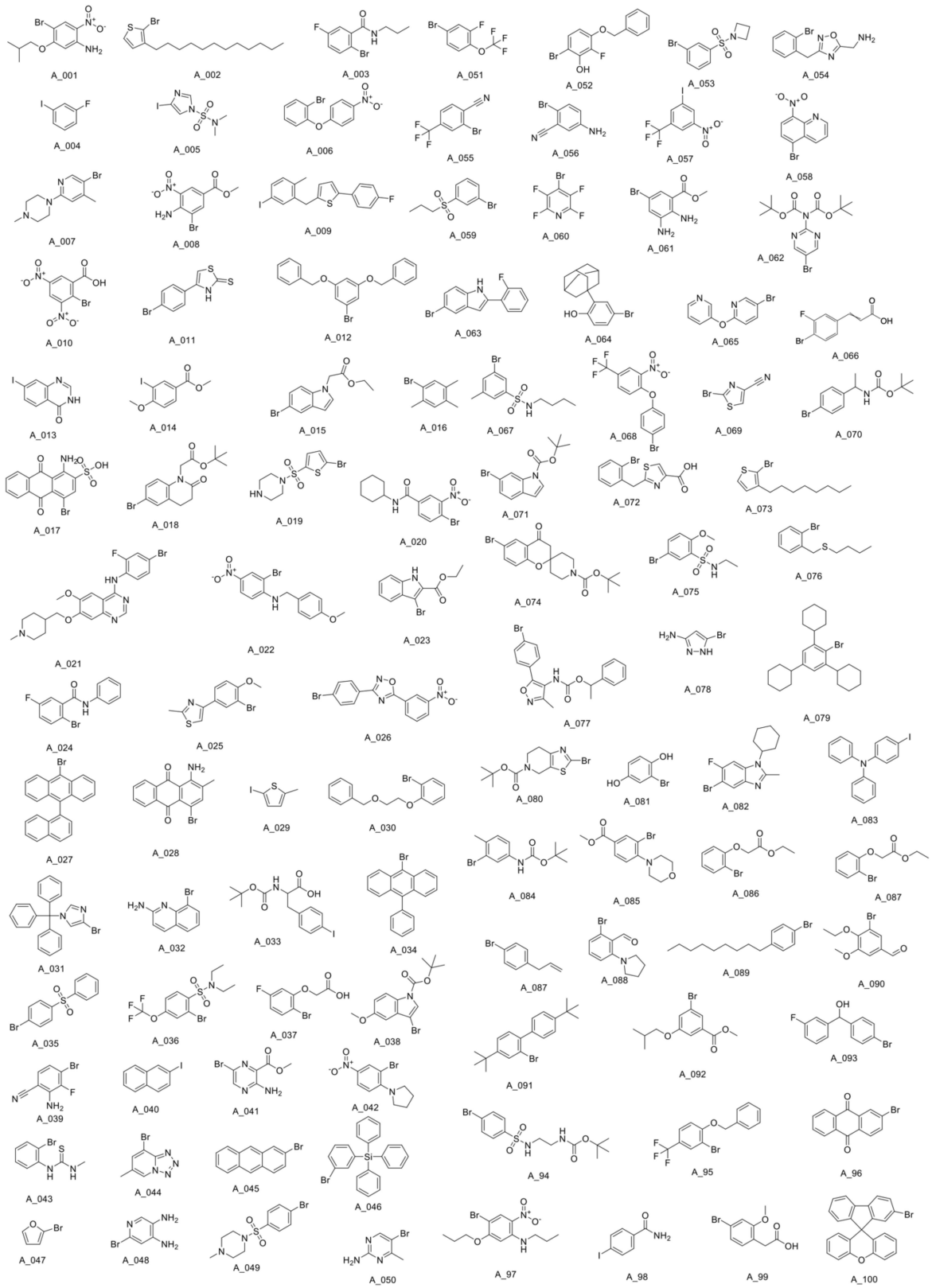
Competing interests The University of Illinois has filed patent applications related to MIDA and TIDA boronates with M.D.B., N.H.A. and W.W. as inventors.

Additional information

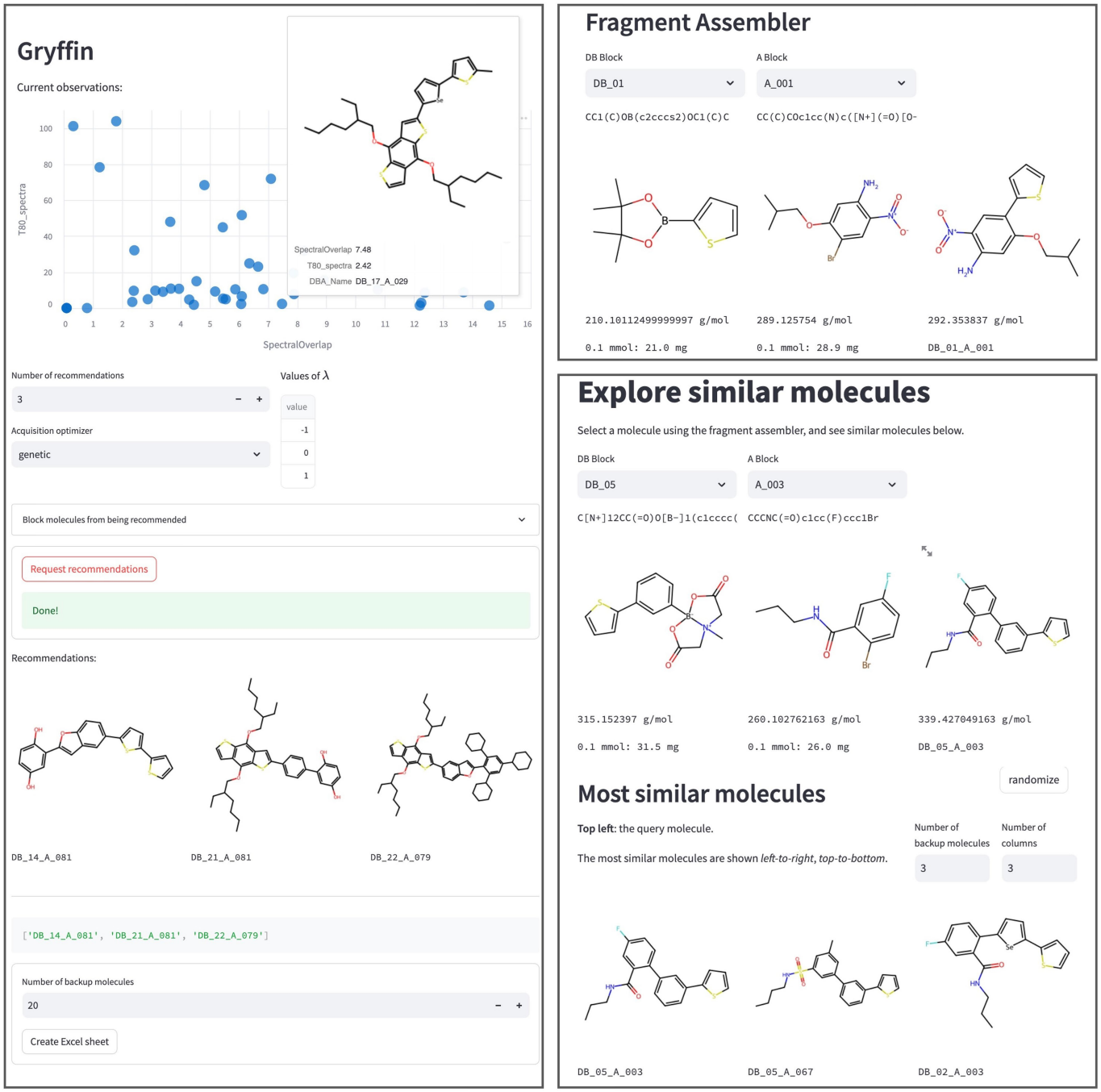
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07892-1.

Correspondence and requests for materials should be addressed to Alán Aspuru-Guzik, Martin D. Burke, Charles M. Schroeder, Ying Diao or Nicholas E. Jackson.

Peer review information *Nature* thanks Brett Savoie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

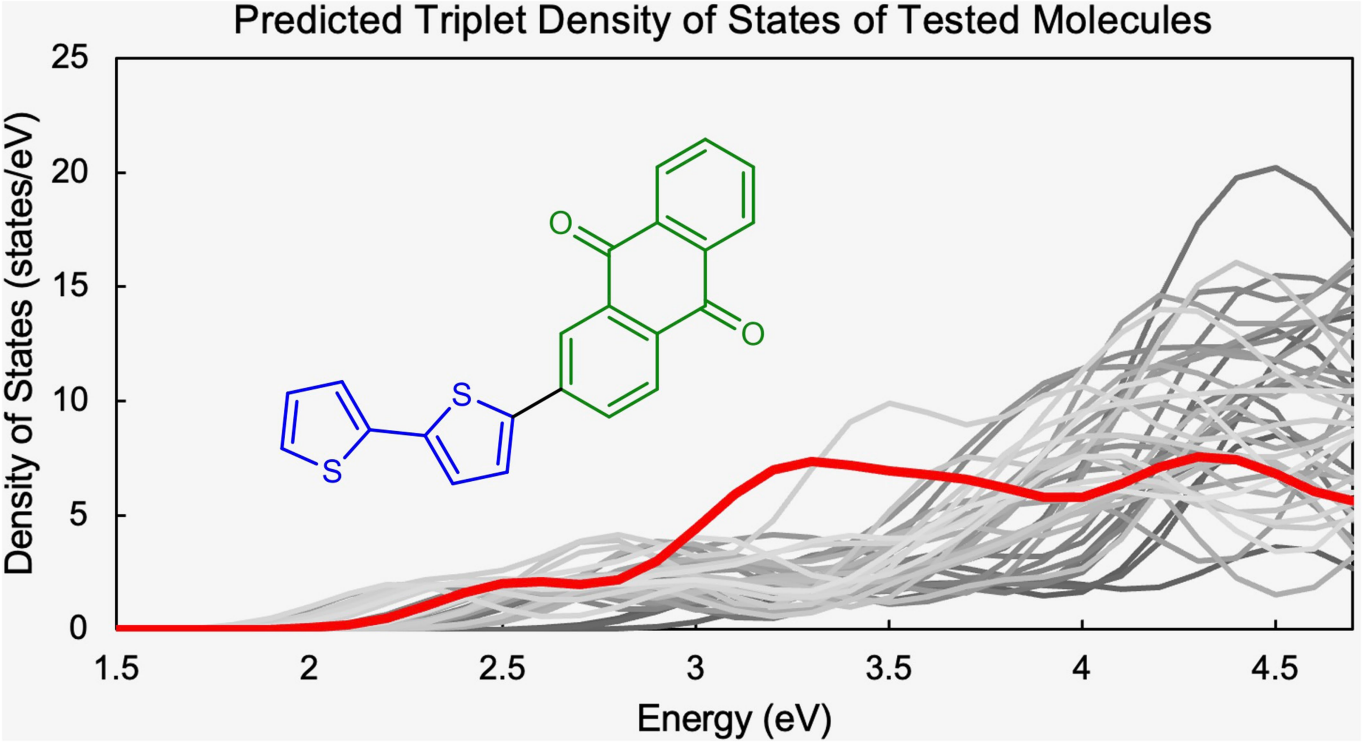


Extended Data Fig. 1 | Chemical design space of acceptor moieties. Chemical diversity down-selected set of acceptor blocks used in populating the design space for the D-B-A motif utilized in this work.

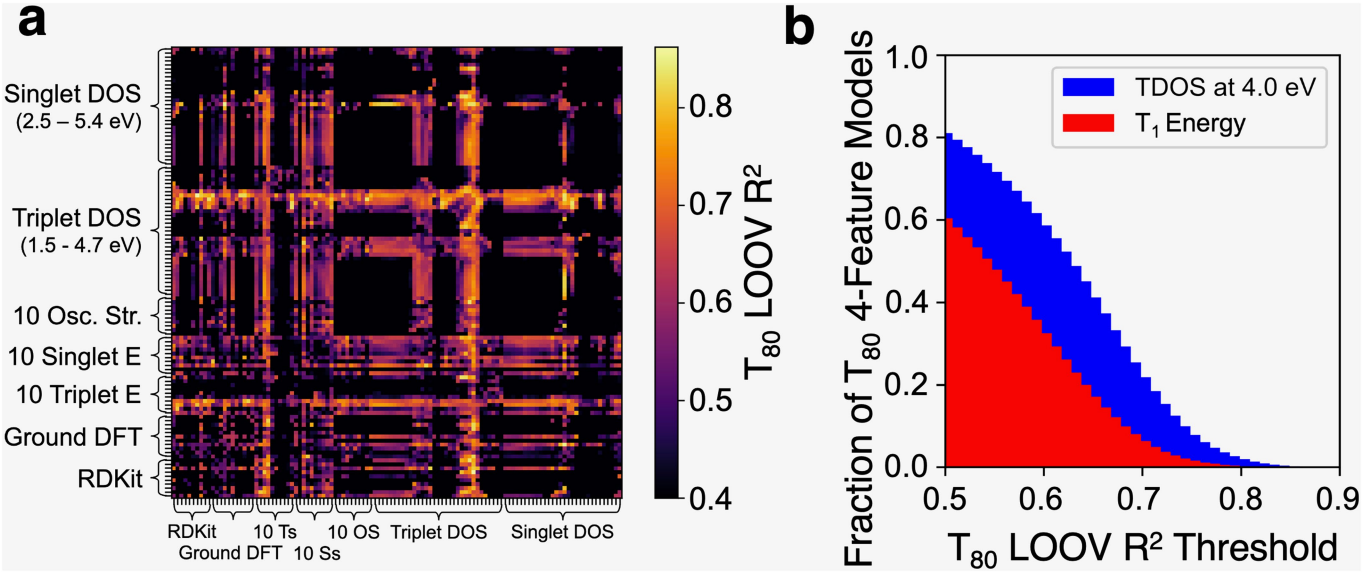


Extended Data Fig. 2 | Digital project manager. Visualization of Streamlit web app used in this work showcasing integration of the Gryffin Bayesian optimizer, building block fragment assembler, and suggestion window for similar molecules as synthetic backups.

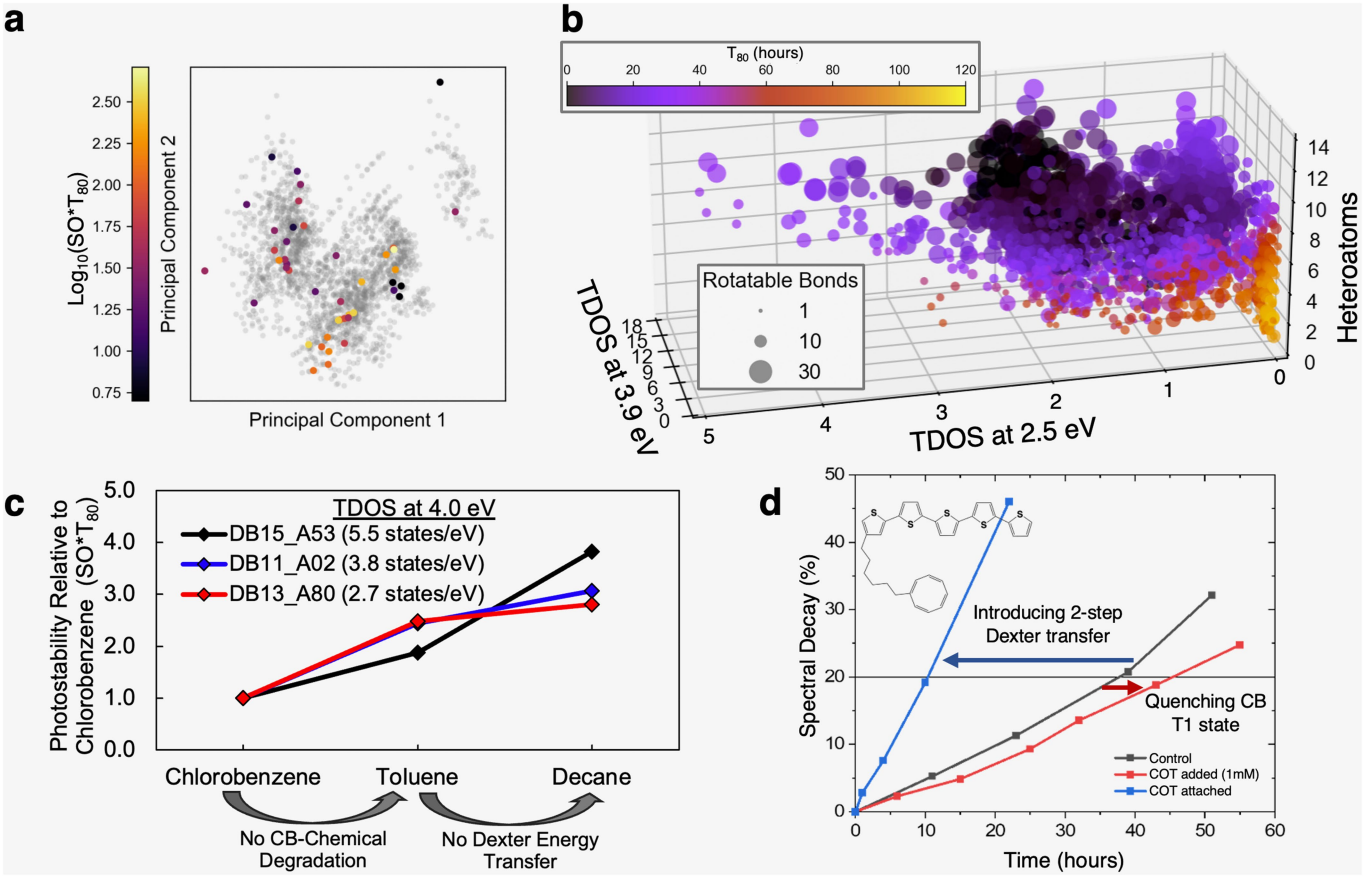
Extended Data Fig. 3 | The small molecule synthesizer used in this work. a, Picture of the hardware. **b**, Design schematic.



Extended Data Fig. 4 | Triplet Density of States (TDOS) for all experimentally measured molecules. DB08_A096 (the high performer in the predicted Bottom 7) 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 Fig. 5 | Results from CLT phase I. a, All support vector regression leave-one-out validation (LOOV) results for predicting T_{80} of the 30 molecules characterized in Phase I from 2-feature combinations. **b**, Comparison of the prediction strength of all possible 4-feature models containing either the TDOS at 4.0 eV or the T_1 energy. Compare to Fig. 5b & c.



Extended Data Fig. 6 | Physics-driven discovery in phase III. a, The distribution of photostabilities of the 44 molecules synthesized through Phase II of the CLT campaign. **b**, The best 4 feature model for predicting T_{80} from Phase II. Note the similarities of the TDOS features to those in the original physics based T_{80} model **c**, The relative photostability of 3 molecules in CB, toluene, and decane, showing the improved photostability for all molecules in the absence of Dexter energy transfer, and that the improvement correlates with the TDOS at 4.0 eV. **d**, The impact of adding cyclooctatetraene (COT) triplet quencher to the CB solution (red) and chemically attaching it to DB_11_A_002 (blue). The structure shown is DB_11_A_002 (the highest performing molecule in Fig. 4a, the control), with the dodecyl side chain replaced with a hexyl-COT side chain. Results support the Dexter energy transfer hypothesis as explained in SI Section 6.

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)

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	12.3	3	36.8
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	8.2
2-1	DB_05_A_089	0.1	n/a	n/a
2-2s11	DB_13_A_100	6.1	51.7	315.4
2-3s3	DB_08_A_018	3.65	48	175.2
2-4	DB_19_A_025	14.6	1.5	21.9
2-d5s1	DB_20_A_027	4.55	15	68.2
2-d8s13	DB_17_A_069	6.84	10.6	72.5
2-d9	DB_12_A_022	5.55	5	27.8
2-d10	DB_16_A_061	4.46	1.9	8.47
3-1	DB_15_A_088	5.19	9.3	48.3
3-2	DB_11_A_002	15.2	38.7	587.7
3-3s1	DB_20_A_012	2.4	9.7	23.3
3-4	DB_06_A_049	1.23	78.3	96.3
3-6	DB_08_A_034	6.08	2.36	14.4
4-1	DB_22_A_046	5.88	10.5	61.7
4-3	DB_22_A_083	6.66	23.1	153.8
4-4	DB_05_A_002	0.1	n/a	n/a
4-5	DB_01_A_002	0.33	101.1	33.4
4-6	DB_06_A_002	1.8	103.9	187.0
5-1	DB_10_A_007	10.8	10.5	112.8
5-2	DB_11_A_007	13.7	8.87	121.5
5-3s4	DB_10_A_091	7.87	19.5	153.3
5-4s7	DB_10_A_084	8.92	15.6	139.5
5-5	DB_10_A_002	12.4	8.64	106.9
5-6s2	DB_09_A_002	7.89	7.86	62.0
Top7-1	DB_04_A_046	6.36	24.9	158.4
Top7-2	DB_04_A_070	8.23	31.9	262.9
Top7-3	DB_13_A_009	2.42	32.2	77.8
Top7-4	DB_13_A_031	3.95	10.8	42.7
Top7-5	DB_13_A_044	5.45	45.0	245.0
Top7-6	DB_13_A_071	4.82	68.4	329.5
Top7-7	DB_13_A_080	6.1	6.67	40.7
Bot7-1	DB_08_A_096	7.1	71.9	510.4
Bot7-2	DB_15_A_066	5.46	5.34	29.2
Bot7-3	DB_17_A_029	7.48	2.42	18.1
Bot7-4	DB_22_A_007	3.4	9.15	31.1
Bot7-5	DB_22_A_018	3.14	9.77	30.7
Bot7-6	DB_22_A_023	4.3	4.8	20.6
Bot7-7	DB_22_A_063	3.67	10.9	39.9

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.