

Towards a comprehensive data infrastructure for redox-active organic molecules targeting non-aqueous redox flow batteries

Rebekah Duke,^{1,2} Vinayak Bhat,^{1,2} Parker Sornberger,^{1,2} Susan A. Odom,¹ Chad Risko^{1,2,*}

¹Department of Chemistry
University of Kentucky
Lexington, Kentucky 40506 USA

²Center for Applied Energy Research
University of Kentucky
Lexington, Kentucky 40511 USA

* Corresponding author: chad.risko@uky.edu

Abstract

The shift of energy production towards renewable, yet at times inconsistent, resources like solar and wind have increased the need for better energy storage solutions. An emerging energy storage technology that is highly scalable and cost-effective is the redox-flow battery comprised of redox-active organic materials. Designing optimum materials for redox flow batteries involves balancing key properties such as the redox potential, stability, and solubility of the redox-active molecules. Here, we present the Data-enabled Discovery and Design to Transform Liquid-based Energy Storage (D³TaLES) database, a curated data collection of more than 43,000 redox-active organic molecules that are of potential interest as the redox-active species for redox flow batteries with the aim to offer readily accessible and uniform data for big data metanalyses. D³TaLES raw data and derived properties are organized into a molecule-centric schema, and the database ontology contributes to the establishment of community reporting standards for electrochemical data. Data are readily accessed and analyzed through an easy-to-use web interface. The data infrastructure is coupled with data upload and processing tools that extract, transform, and load relevant data from raw computation or experimental data files, all of which are available to the public via a D³TaLES API. These processing tools along with an embedded high-throughput computational workflow enable community contributions and versatile data sharing and analyses, not only in redox-flow battery research but also in any field that applies redox-active organic molecules.

Introduction

Increasing use of renewable yet inconsistent energy sources like solar and wind demands better energy storage solutions. An emerging energy storage technology that is highly scalable and cost-effective is the redox flow battery (RFB).¹⁻³ The RFB decouples energy capacity and power by separating the electrochemical reactions from stored electrochemical energy, allowing the battery to store large quantities of energy cheaply and safely.¹ The battery consists of two tanks of solvated redox-active molecules—the catholyte in one tank and the anolyte in the other. During charge, the catholyte and anolyte are pumped through a reaction cell where a membrane separates them. The catholyte is comprised of redox-active molecules that are oxidized at a porous electrode, while the anolyte contains redox-active molecules that are reduced at another porous electrode. At discharge, the oxidized catholyte and reduced anolyte are pumped back through the reaction cell, where the reverse reaction occurs, releasing stored electrochemical energy.

While current commercially available RFB use vanadium, organic-based RFB show promise as organic molecules can be more widely available and cheaper than mined and/or rare metals.⁴⁻⁶ Additionally, redox-active organic molecules are highly tunable and can be synthesized from sustainable materials.^{3, 7-9} While commercial and other promising RFB materials are comprised of aqueous solvents, nonaqueous solvents afford large potential windows, increasing the battery's voltage and thus its potential energy storage.¹⁰ Even so, there is limited research targeting redox-active molecules for nonaqueous solvents in RFB, so-called nonaqueous RFB (NARFB).

Both experimental and computational methods exist for deciphering fundamental material properties for NARFB materials (**Figure 1**), and the computational/simulation-based approaches have been vital in identifying candidates for NARFB by simulating redox potential, stability, and reversibility, to name a few.¹¹⁻¹³ However, while there have been efforts to identify redox-active molecules suitable for NARFB

catholyte and anolyte materials, there remains a lack of fundamental chemical understanding for these systems, especially considerations as to how to appropriately balance critical properties such as redox potential, stability, and solubility.^{3, 9, 14, 15}

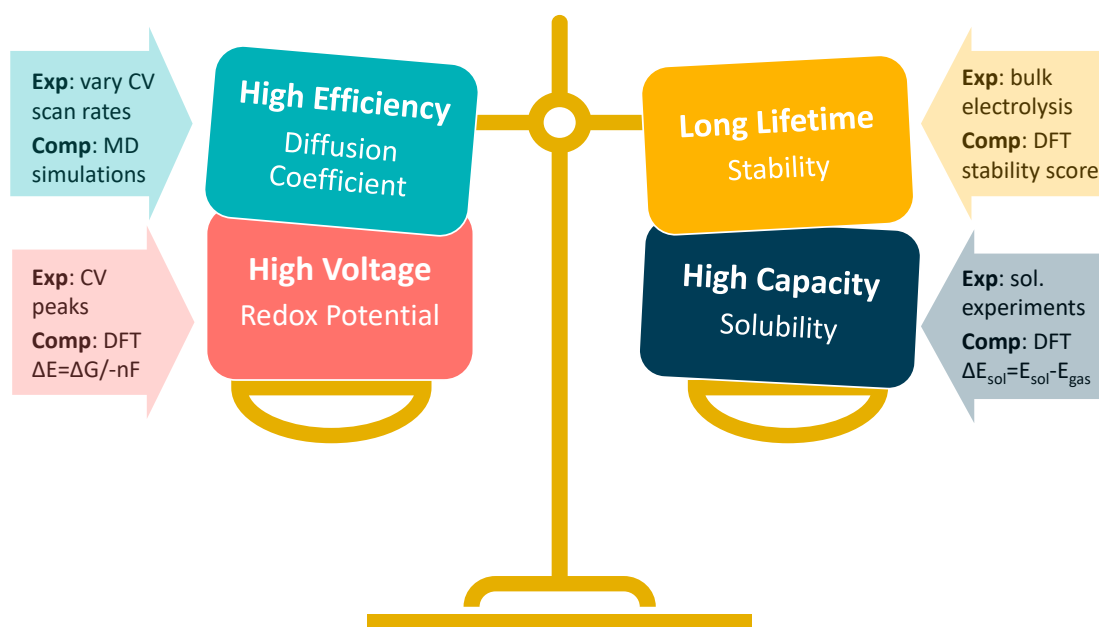


Figure 1. Fundamental redox-active material properties that must be balanced for use as the catholyte or anolyte in an RFB. Each fundamental property can be estimated experimentally via techniques like cyclic voltammetry (CV) or computationally via density functional theory (DFT) and/or molecular dynamics (MD) simulations.

Fortunately, when data are amassed from both computational and experimental sources, big-data analyses can inform structure-property relationships. Previous data-enabled insights have been achieved in similar fields. For example, big data analyses elucidated a “stability cliff” in quinones, a prevalent molecular class in aqueous RFB, encouraging researchers to explore other chemical spaces.¹⁶ Efforts are already underway to develop data-driven pipelines and apply big-data analyses for vanadium RFB^{17, 18} and aqueous organic RFB¹⁹⁻²¹ materials. Some big data approaches have been applied to the search for NARFB materials; for example, data-enabled high-throughput screening of redox-active molecules for NARFB has been demonstrated in a small-scale proof-of-concept study where several theoretically viable molecules

for NARFB analytes were selected from ~1400 quinoxaline-based systems with a funnel-based screening approach focusing on reduction potential, solvation energy, and structural changes with oxidation.²²

Unfortunately, the few studies that examine systems for NARFB are smaller scale and (like those in the field of aqueous RFB) often focus on quinone-based systems alone. Additionally, elucidating structure-property relationships for properties such as solubility in nonaqueous environments for NARFB can be much more challenging than in aqueous environments.²³ Thus, in the field of NARFB, the meta-analyses necessary to elucidate structure-property relationships are often prohibited by the lack of large-scale, broad, accessible, and uniform data.

Here we present a curated database of redox-active organic molecules as a part of a multidisciplinary, collaborative platform entitled *data-enabled discovery and design to transform liquid energy storage* (D³TaLES).²⁴ We collect and curate data from various sources, including computational analyses and original experimentation, and we build the infrastructure to accept data submissions from the community. The data infrastructure includes data upload and processing tools that extract, transform, and load (ETL) relevant data from raw computation or experimental data files and organize it into a molecule-centric schema. Here the data are easily accessed and analyzed, and a layered, redundant database structure provides critical opportunities for manual and automated data curation. A high-throughput computational workflow begins populating the database with DFT computational data. Similar data structures involving high-throughput computation for π -conjugated organic molecules exist,²⁵⁻³⁰ and there also exist databases targeting batteries.^{19, 31, 32} But unlike these existing databases, D³TaLES provides a data infrastructure and framework for multiple data types targeting NARFB. The database enables deeper physiochemical understanding and opportunities for meta-analysis. Though the focus presented here is on identifying systems that hold potential for NARFB, the platform has a broad scope and can be used or expanded to search for characteristics of redox-active organic molecules in other fields of application.

D³TaLES database

Design

The schema, or organizational data structure, provides the foundation for the database. A No-SQL schema was chosen because of its flexibility and scalability.³³ D³TaLES data is segmented into two databases to accommodate the complexity and breadth of the data collected—one for raw data (backend) and the other for processed data (frontend).

The backend database contains data parsed directly from experiment data files. It uses computation- and/or experiment-centric schema where each data instance is a calculation or experiment with associated attributes (**Figure 2A**). Attributes include calculation/experiment identifier, submission information, and a collection of raw data values, including computational/experimental conditions. For example, the backend database might hold raw data values like the computed energies for a molecule's ground and oxidized or reduced states. The backend database can also hold cyclic voltammetry (CV) data extracted from a CV output file. Thus, the backend schema relates directly to the files that supply data and often incorporates features from existing community schema.^{34, 35} Because D³TaLES contains many types of data, the backend schema has several sub-schemas—one for each data type. The sub-schemas share common fields such as "mol_id", "submission_info", and "data". The backend schemas are broad and accommodate different types of data including but not limited to computations at different levels of theory (e.g., molecular dynamics simulations), experiments with various processing or data collection conditions, and literature-extracted data from various learning models (see SI Section 1); efforts are ongoing to add more schema.

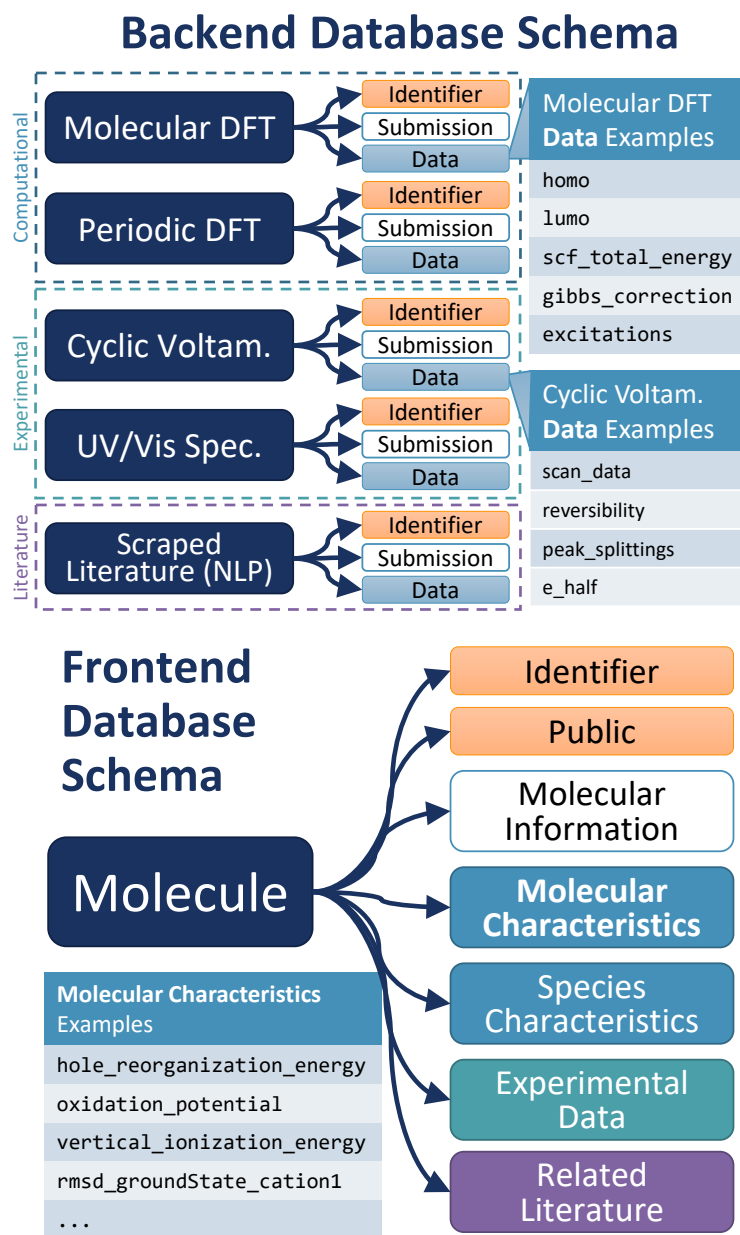


Figure 2. (Top) Depiction of the backend D³TaLES schema and collection types. Note that the figure shows a sampling of the types of collections that exist in the backend database; to view the full D³TaLES database schema, visit the documentation.³⁶ Tables with example “data” attributes are also shown. (Bottom) Schematic showing the first property level for the molecule-centric, frontend D³TaLES schema along with a table showing example attributes in the “molecular characteristics” group.

The frontend database holds data that are more useful for analysis. For example, the frontend database contains ionization potentials calculated from the ground- and oxidized-state energies in the backend database. Likewise, the frontend database contains an estimated redox potential calculated from CV data.

The frontend database uses a molecule-centric schema where each data instance is a molecule (**Figure 2B**). Molecule attributes include a molecule identifier and public/private status, while the remaining attributes are grouped into the following sub-categories: molecule characteristics, species characteristics, raw experiment data (which connects to the backend database), and related literature. Molecular characteristics include properties of the entire molecule (usually involving multiple species), such as oxidation potential or relaxation energies. Species characteristics include properties relating to a single charge species for the molecule, such as ground-state species HOMO or oxidized-state solvation energy. The complete D³TaLES schema is available online.³⁶

Population

A processing workflow populates the database when raw data files and associated metadata are uploaded to the D³TaLES website (**Figure 3**). The raw data files are parsed to extract key values. Existing parsing packages (such as Pymatgen,³⁵ RDKit,³⁷ and SciPy³⁸) are integrated with original code to parse raw computational and experimental data files. These processing tools are packaged in the D³TaLES application programming interface (API; discussed in the D³TaLES Tools section). The raw data files are then compressed and stored, while extracted key values are inserted into the backend database. At this stage, an administrator inspects the backend data to ensure some degree of fidelity. Upon administrator approval, the backend data is transformed into frontend properties. Users may view the frontend database via interactive molecule viewing webpages on the D³TaLES website.²⁴ More information about the database software can be found in the SI Section 2.

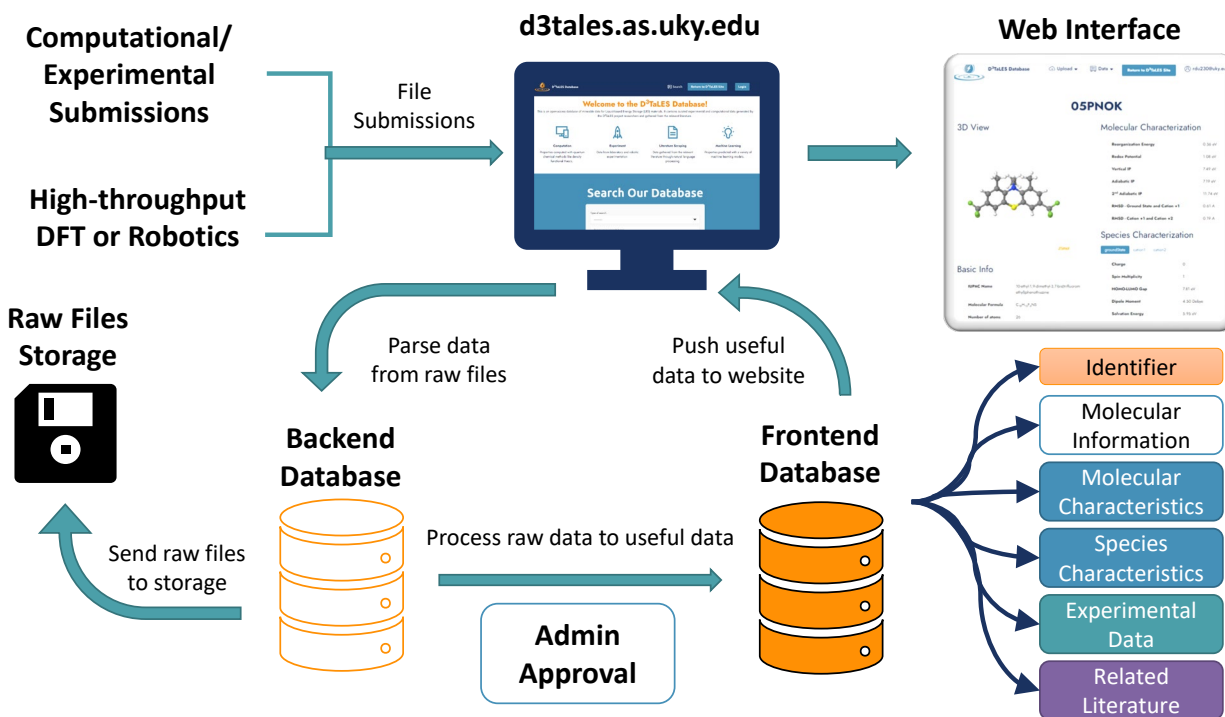


Figure 3. Schematic showing D³TaLES data processing. Data flows from external sources, such as high-throughput computation or robotics, through the D³TaLES website to the backend database. From here, raw data is stored while administrator (admin) approval allows data transformation to the frontend database. Frontend data is displayed through a user interface on the D³TaLES website.

The D³TaLES database contains primarily computational data generated through a high-throughput molecular computational workflow using density functional theory (DFT) carried out at the (IP-tuned) LC- ω HPBE/Def2SVP level of theory via the Gaussian16 (rev A.03) software suite.³⁹⁻⁴² The data produced in this workflow cover several fundamental properties of redox-active molecules, including oxidation and reduction potentials, stability, and solubility; see SI Section 4 for more details.

Data Composition

Molecules in the D³TaLES database are collected from those appearing in the NARFB literature,^{23, 43, 44} scraped from the Cambridge Structural Database (CSD)⁴⁵ and ZINC⁴⁶ datasets (**Figures S4-5**), and

combinatorically generated from fragments of molecules commonly used in NARFB; see SI Section 3 for more details. While these datasets contain inherent biases (e.g., CSD molecules are crystallizable, ZINC molecules are already commercially available, combinatorically generated molecules conform to current conceptions in the field about what structures will work in NARFB, etc.), this collection provides an initial dataset of small organic molecules covering a relatively diverse chemical space. The scraped data number over 600,000 molecules, along with a few dozen experimental molecules from collaborators and a few hundred auto-generated molecules from common motifs used in NARFB (**Figure 4**). The following criteria were then used to filter this extensive molecular dataset: A molecule must have at least one aromatic ring, contain no rings with more than six atoms and no rings with less than five atoms, contain no rings with more than three heteroatoms, and not exist already in the OCELOT²⁶ database (a database of large organic molecule and their corresponding crystal properties targeting organic semiconductors developed by our lab). This narrowed the dataset to approximately 115,000 molecules. Finally, the dataset was narrowed further because of limited computational resources. To ensure diversity of the chemical space, the 33,000 filtered ZINC molecules most different from the rest of the dataset (CSD, generate, and NARFB literature molecules) were chosen. The similarity was determined with the RDKit Tanimoto fingerprint method.^{37, 47} The final chemical space consists of 43,168 molecules, where approximately 3,500 are proprietary and 39,500 are public. Of these structures, 31,583 have a complete oxidation profile.

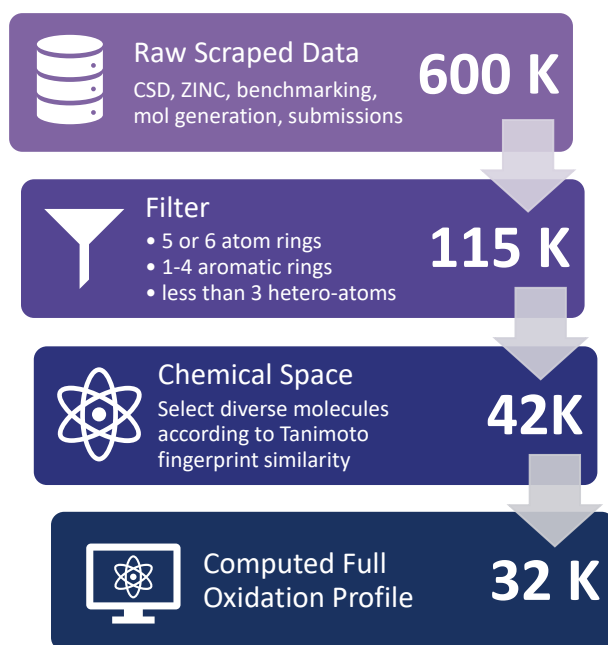


Figure 4. Process for molecule generation for the D³TaLES database.

The 43,168 unique structures in the D³TaLES database have a mean molecular weight of 329 g/mol (**Figure 5A**). All properties generated for the oxidation profile are listed in the D³TaLES database documentation,³⁶ but notable properties include oxidation potential, relaxation energies, vertical and adiabatic ionization potentials, solvation energies, and a radical-cation stability score developed by Sowndarya et. al.⁴⁸ **Figure 5B** shows a UMAP⁴⁹ chemical space plot of the calculated oxidation potentials where groupings of higher and lower potentials are viable. The plot includes 10-ethylphenothiazine (EPT) and (2,2,6,6-tetramethylpiperidin-1-yl)oxyl (TEMPO), two widely-reported molecules of interest for organic RFB.⁷ **Figure 5C** shows the database structures plotted by oxidation potential and the radical-cation stability score.⁴⁸ The marginal histogram depicts a normally distributed radical stability score (RSS), with the highest stability scores observed for larger molecules. In contrast, there exists little correlation between size and oxidation potential, though most oxidation potentials are concentrated just above zero eV (relative to the standard hydrogen electrode, SHE). The database is now being populated with reduction

profiles for many of the structures. These profiles contain the reduction analog for each of the oxidation profile properties. Currently, the database contains over 25,000 reduction profiles.

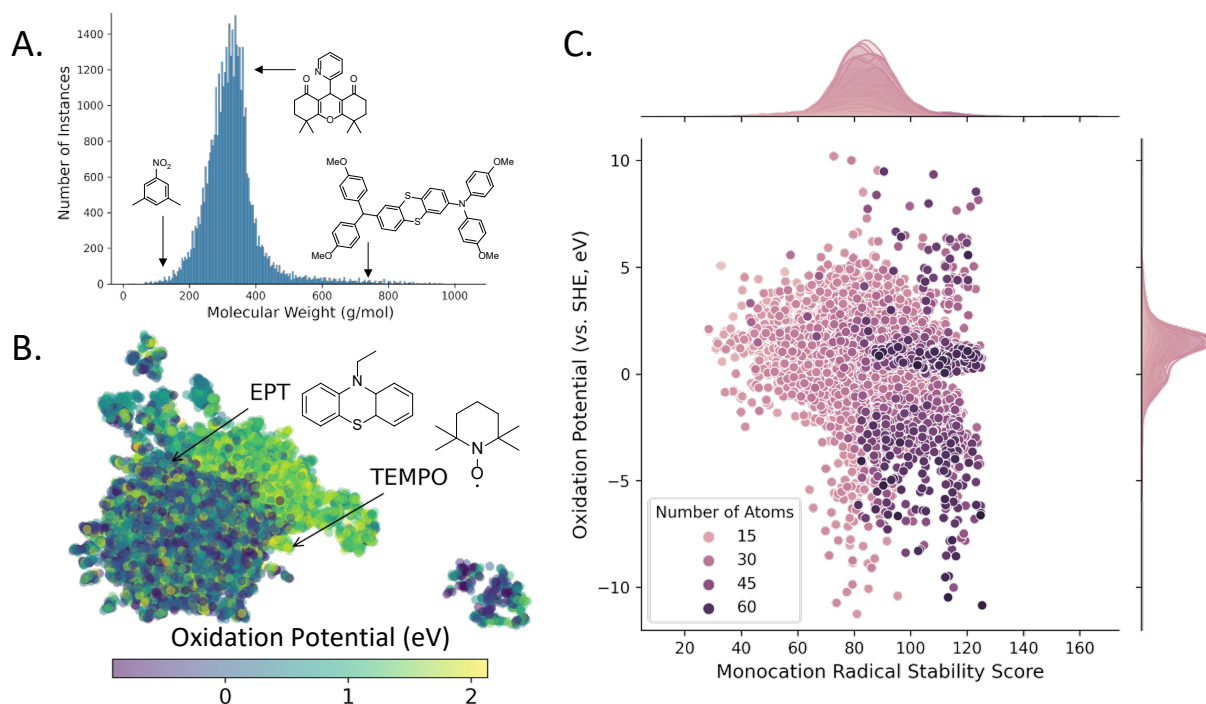


Figure 5. The D³TaLES frontend database contains over 43,000 molecules. (A) Histogram showing molecular weight distribution for the D³TaLES database. (B) The computed values for oxidation potential (a molecular characteristic) are mapped onto a two-dimensional chemical space with ChemPlot⁵⁰ and UMAP⁴⁹ dimension reduction. (C) Scatter plot with marginal histograms showing D³TaLES molecules plotted by calculated oxidation potential (versus the standard hydrogen electrode, SHE) and radical stability score, colored by number of atoms.⁴⁸

D³TaLES Tools

The D³TaLES database is coupled with several data interaction and management tools including the D³TaLES website²⁴ and the D³TaLES API.⁵¹ The D³TaLES website is integral for many of the processes described above. Website features include file upload systems, backend data viewing and approval, database search functions, and molecule viewing pages. All user data submissions and administrator approval of the processed data occur through the website. Users may search the database by molecule name or structure. All data for a given molecule can be viewed on the molecule property viewing page

(Figure 6A). Alternatively, for those wishing to access large quantities of data through code, the D³TaLES REST API allows data access through HTML according to REST (representational state transfer) standards.⁵² Finally, the site contains links to the D³TaLES database documentation,³⁶ D³TaLES API documentation,⁵¹ and the D³TaLES calculators interactive python notebooks.⁵³

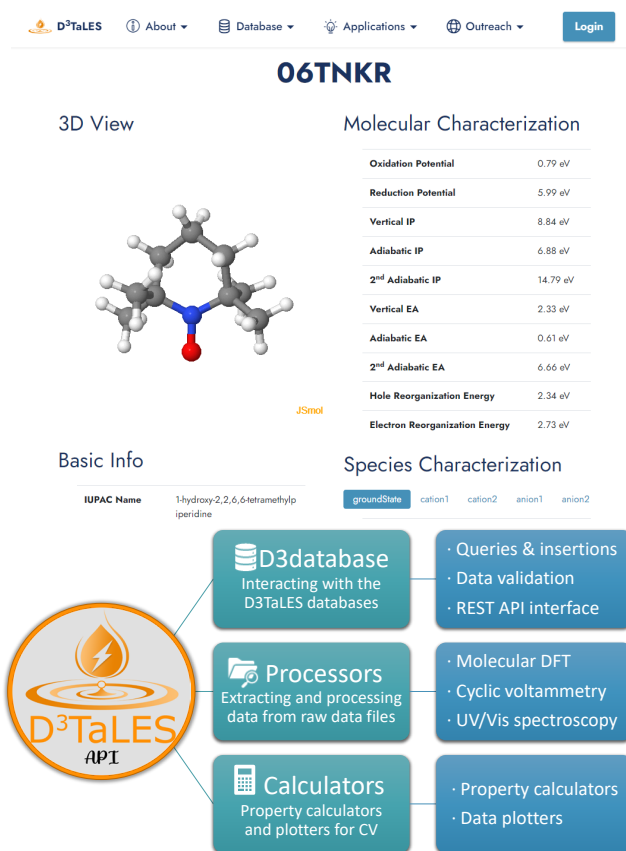


Figure 6. (Top) D³TaLES molecule viewing page.⁵⁴ (Bottom) The organizational structure of the D³TaLES API. Full documentation for the D³TaLES API is available.⁵¹

Several tools for moving, processing, and transforming data accompany the D³TaLES database. These tools are compiled in the D³TaLES API.⁵¹ The D³TaLES API includes three modules: *Processors* for data processing, *D3database* for database access, and *Calculators* for property calculations (Figure 6B). The *Processors* module contains various parsing classes for extracting useful data from instrument-produced computational and experimental data files. Among the database access functions, the *D3database*

module contains a class for accessing the D³TaLES database via Python through the REST API. This module also contains functions for gathering and plotting D³TaLES properties as one- and two-dimensional histograms. Finally, the *Calculators* module, perhaps the most useful module for the general community, allows users to calculate useful computational and experimental properties from nested data. All calculators contain unit conversion features. Useful molecular DFT calculators include redox potential, radical buried volume,⁵⁵ and radical spin density,⁵⁶ while useful CV calculators include diffusion constant using the Randles-Sevcik equation and charge-transfer rate. The D³TaLES API documentation⁵¹ explains basic usage for these calculators, and we also provide interactive Python notebooks that use the calculators to perform calculations without the need for the user to know Python coding.⁵³ For more information about the D³TaLES API, see SI Section 6.

D³TaLES Database Utility

To demonstrate the D³TaLES database utility in identifying candidates for redox flow batteries, we used the compiled computational data to perform a proof-of-concept funnel pipeline (**Figure 7**).⁵⁷⁻⁶⁰ The funnel pipeline iteratively narrows the D³TaLES chemical space through a series of tests to identify candidates for a NARFB catholyte material. The tests are ordered from least to most computationally intensive. The first test (~1 ms) selects molecules with less than 30 atoms. Redox-active systems with fewer atoms per charge event increase the atom economy,⁶¹ and thus the capacity for a RFB. Subsequently, the second test (~1 s) filters out molecules that would be difficult to synthesize by selecting systems with a synthetic accessibility score below 4.1.^{56, 62} The next two tests filter by stability and solvation energy, respectively, relative to the properties of a known candidate for NARFB: N-(2-(2-methoxyethoxy)ethyl)phenothiazine (MEEPT).^{7, 63} MEEPT is known to be soluble, especially in its ground state, and it shows stable cycling of one oxidation event. The third test (~21 core-hrs) filters out molecules with an RSS greater than MEEPT's

81, while the fourth test (~21 core-hrs) identifies molecules with solvation energy lower than MEEPT's – 0.19 eV. The final test (~43 core-hrs) finds molecules with an oxidation potential of approximately 3 V, as higher oxidation potentials are most desirable for catholyte materials. (To view structures from the funnel pipeline and for more information about the core-hour estimations, see SI Section 5.) The funnel pipeline down-selects the 43,168 D³TaLES structures to 364 potential systems for NARFB. While all calculations were performed for all molecules used here, this approach could be employed to explore a large chemical space without performing all resource-intensive calculations for all systems. Additionally, the existing D³TaLES data can be used to train machine learning (ML) models that quickly estimate resource-intensive properties such as oxidation potential; these models could be added as an upper level of the funnel pipeline.⁶⁴

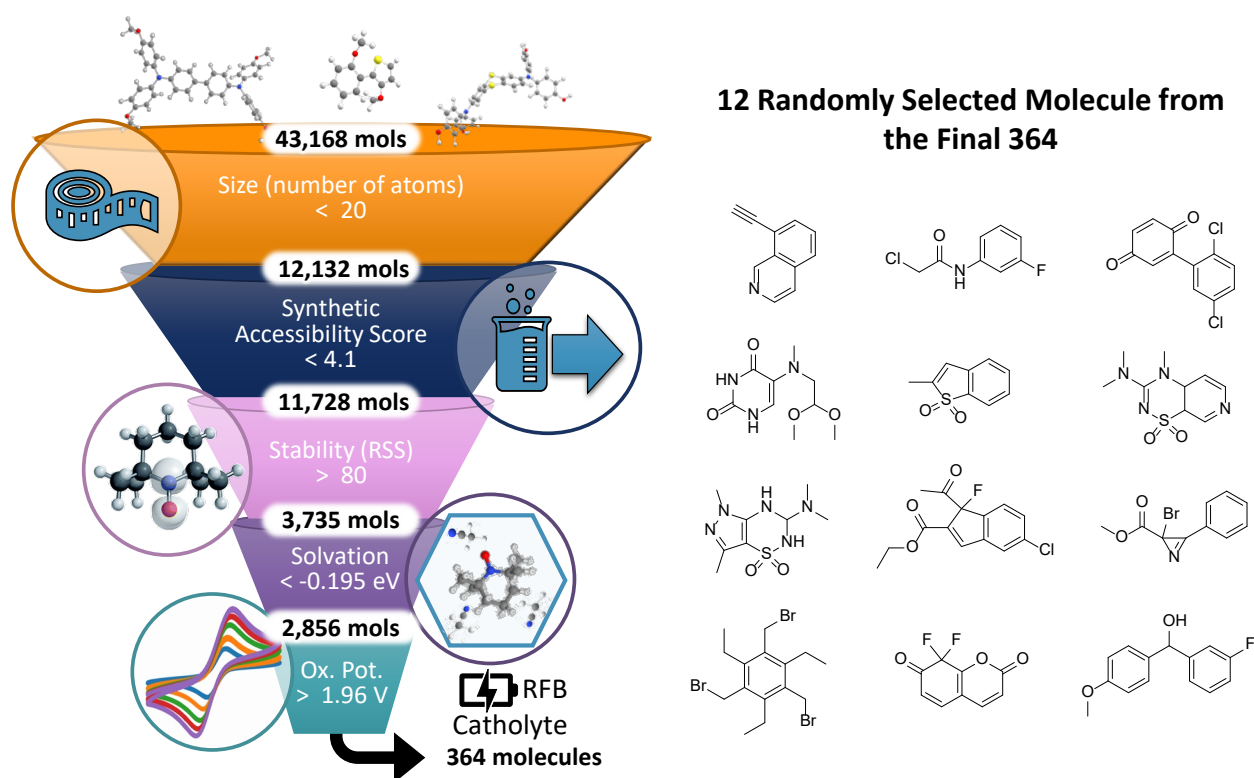


Figure 7. (Left) Schematic demonstrating the proof-of-concept funnel pipeline using D³TaLES computational data. The five tests narrow the chemical space by number of atoms, synthetic accessibility score, radical stability score (RSS), solvation energy, and oxidation potential, respectively. (Right) Twelve randomly selected structures from the final 364 structures that emerged from the funnel pipeline.

Conclusion

We demonstrate a comprehensive data infrastructure for redox-active small molecules for use in NARFBs. For the over 43,000 molecules currently in the D³TaLES database, a high-throughput computational workflow has determined over 31,000 oxidation profiles and other properties of interest to date. While the database currently consists almost exclusively of DFT computational data, the schema and processing infrastructure exist for incorporating experimental and literature-reported data. Future work will focus on exploiting the data processing tools and data storage infrastructure to continue populating the D³TaLES database, especially in areas outside of molecular DFT, such as periodic DFT, molecular dynamics simulations, and cyclic voltammetry and UV-Vis spectroscopy experiments.

We demonstrate the utility of the D³TaLES infrastructure by screening the over 43,000 molecules in the database for NARFB application. This preliminary screening predicts 364 candidates with characteristics superior to the current standard MEEPT. We note that a thorough analysis is warranted to confirm these predictions. The D³TaLES database and data infrastructure will enable integrated meta-analytical and machine-learning-based evaluation in the NARFB field, with the aim to expedite materials discovery and pave the way for predictive models for properties such as redox potentials and radical cation stability. The uniform and accessible D³TaLES data will enable machine learning and robotic experimentation towards better exploring relevant chemical space for application-suitable redox molecules.

Data Availability Statement

The data presented here are accessible via the D³TaLES website (<https://d3tales.as.uky.edu/>), and the public portion of the dataset (~39,500 molecules) can be downloaded at <https://d3tales.as.uky.edu/datasets>. The D³TaLES website also includes documentation for the database

structure and more information about the data composition (<https://d3tales.as.uky.edu/docs/>). The processing tools associated with the D³TaLES API exist in an open-access Python package documented at https://d3tales.github.io/d3tales_api/. The Fireworks-based⁶⁵ code used for the high-throughput quantum chemical calculations is available publicly at https://github.com/D3TaLES/d3tales_fw. Additional details and information can be found in the accompanying Supplementary Information.

Author Contributions

R. D.: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, writing – review & editing. **V. B.:** conceptualization, software, supervision, writing – review & editing. **P. S.:** data curation, writing – review & editing. **S. A. O.:** conceptualization, funding acquisition. **C. R.:** conceptualization, supervision, funding acquisition, writing – review & editing.

Supporting Information

Document containing details about molecular generation, computational methods, funnel workflow details, more information about the D³TaLES API, etc.

Acknowledgments

This work was generously supported by the National Science Foundation (NSF) under Cooperative Agreement Number 2019574. Computational resources were provided through an NSF Extreme Science

and Engineering Discovery Environment (XSEDE) Resource Allocation Award (CHE200119) and Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) DISCOVER Allocation Award (PHY220121). We further acknowledge the University of Kentucky (UK) Center for Computational Sciences and Information Technology Services Research Computing for their fantastic support and collaboration, and use of the Lipscomb Compute Cluster and associated research computing resources. Finally, we wholeheartedly thank the entire D³TaLES team for their insights into the development of this data architecture.

References

- (1) Luo, J.; Hu, B.; Hu, M.; Zhao, Y.; Liu, T. L. Status and Prospects of Organic Redox Flow Batteries toward Sustainable Energy Storage. *ACS Energy Letters*. **2019**, *4*, 2220-2240. DOI: 10.1021/acsenerylett.9b01332.
- (2) Pan, F.; Wang, Q. Redox Species of Redox Flow Batteries: A Review. *Molecules*. **2015**, *20*, 20499-20517. DOI: 10.3390/molecules201119711.
- (3) Li, M.; Odom, S. A.; Pancoast, A. R.; Robertson, L. A.; Vaid, T. P.; Agarwal, G.; Doan, H. A.; Wang, Y.; Suduwella, T. M.; Bheemireddy, S. R.; et al. Experimental Protocols for Studying Organic Non-aqueous Redox Flow Batteries. *ACS Energy Letters*. **2021**, 3932-3943. DOI: 10.1021/acsenerylett.1c01675.
- (4) Scott, R. A.; Hu, B.; Luo, J.; DeBruler, C.; Hu, M.; Wu, W.; Liu, T. L. Redox-Active Inorganic Materials for Redox Flow Batteries. *Encyclopedia of Inorganic and Bioinorganic Chemistry*. **2019**, 1-25. DOI: 10.1002/9781119951438.eibc2679.
- (5) Viswanathan, V.; Crawford, A.; Stephenson, D.; Kim, S.; Wang, W.; Li, B.; Coffey, G.; Thomsen, E.; Graff, G.; Balducci, P.; et al. Cost and performance model for redox flow batteries. *Journal of Power Sources*. **2014**, *247*, 1040-1051. DOI: <https://doi.org/10.1016/j.jpowsour.2012.12.023>.
- (6) Wang, W.; Luo, Q.; Li, B.; Wei, X.; Li, L.; Yang, Z. Recent Progress in Redox Flow Battery Research and Development. *Advanced Functional Materials*. **2013**, *23*, 970-986, <https://doi.org/10.1002/adfm.201200694>. DOI: <https://doi.org/10.1002/adfm.201200694>.
- (7) Li, M.; Rhodes, Z.; Cabrera-Pardo, J. R.; Minter, S. D. Recent advancements in rational design of non-aqueous organic redox flow batteries. *Sustainable Energy & Fuels*. **2020**, *4*, 4370-4389. DOI: 10.1039/d0se00800a.
- (8) Fang, X.; Li, Z.; Zhao, Y.; Yue, D.; Zhang, L.; Wei, X. Multielectron Organic Redoxmers for Energy-Dense Redox Flow Batteries. *ACS Materials Letters*. **2022**, 277-306. DOI: 10.1021/acsmaterialslett.1c00668.
- (9) Wei, X.; Pan, W.; Duan, W.; Hollas, A.; Yang, Z.; Li, B.; Nie, Z.; Liu, J.; Reed, D.; Wang, W.; et al. Materials and Systems for Organic Redox Flow Batteries: Status and Challenges. *ACS Energy Letters*. **2017**, *2*, 2187-2204. DOI: 10.1021/acsenerylett.7b00650.
- (10) Shin, S.-H.; Yun, S.-H.; Moon, S.-H. A review of current developments in non-aqueous redox flow batteries: characterization of their membranes for design perspective. *RSC Advances*. **2013**, *3*, 9095. DOI: 10.1039/c3ra00115f.
- (11) Montoto, E. C.; Cao, Y.; Hernández-Burgos, K.; Sevov, C. S.; Braten, M. N.; Helms, B. A.; Moore, J. S.; Rodríguez-López, J. Effect of the Backbone Tether on the Electrochemical Properties of Soluble Cyclopropenium Redox-Active Polymers. *Macromolecules*. **2018**, *51*, 3539-3546. DOI: 10.1021/acs.macromol.8b00574.
- (12) Yan, Y.; Robinson, S. G.; Sigman, M. S.; Sanford, M. S. Mechanism-Based Design of a High-Potential Catholyte Enables a 3.2 V All-Organic Nonaqueous Redox Flow Battery. *Journal of the American Chemical Society*. **2019**, *141*, 15301-15306. DOI: 10.1021/jacs.9b07345.

- (13) Yan, Y.; Vaid, T. P.; Sanford, M. S. Bis(diisopropylamino)cyclopropenium-arene Cations as High Oxidation Potential and High Stability Catholytes for Non-aqueous Redox Flow Batteries. *Journal of the American Chemical Society*. **2020**, *142*, 17564-17571. DOI: 10.1021/jacs.0c07464.
- (14) Goulet, M.-A.; Tong, L.; Pollack, D. A.; Tabor, D. P.; Odom, S. A.; Aspuru-Guzik, A.; Kwan, E. E.; Gordon, R. G.; Aziz, M. J. Extending the Lifetime of Organic Flow Batteries via Redox State Management. *Journal of the American Chemical Society*. **2019**, *141*, 8014-8019. DOI: 10.1021/jacs.8b13295.
- (15) Zhong, F.; Yang, M.; Ding, M.; Jia, C. Organic Electroactive Molecule-Based Electrolytes for Redox Flow Batteries: Status and Challenges of Molecular Design. *Frontiers in Chemistry*. **2020**, *8*. DOI: 10.3389/fchem.2020.00451.
- (16) Tabor, D. P.; Gómez-Bombarelli, R.; Tong, L.; Gordon, R. G.; Aziz, M. J.; Aspuru-Guzik, A. Mapping the frontiers of quinone stability in aqueous media: implications for organic aqueous redox flow batteries. *Journal of Materials Chemistry A*. **2019**, *7*, 12833-12841. DOI: 10.1039/c9ta03219c.
- (17) Cheng, Z.; Tenny, K. M.; Pizzolato, A.; Forner-Cuenca, A.; Verda, V.; Chiang, Y.-M.; Brushett, F. R.; Behrou, R. Data-driven electrode parameter identification for vanadium redox flow batteries through experimental and numerical methods. *Applied Energy*. **2020**, *279*, 115530. DOI: <https://doi.org/10.1016/j.apenergy.2020.115530>.
- (18) Li, R.; Xiong, B.; Zhang, S.; Zhang, X.; Li, Y.; Lu, H.; Fernando, T. A novel U-Net based data-driven vanadium redox flow battery modelling approach. *Electrochimica Acta*. **2023**, *444*, 141998. DOI: <https://doi.org/10.1016/j.electacta.2023.141998>.
- (19) Sorkun, E.; Zhang, Q.; Khetan, A.; Sorkun, M. C.; Er, S. RedDB, a Computational Database of Electroactive Molecules for Aqueous Redox Flow Batteries. American Chemical Society (ACS): 2021.
- (20) Gao, P.; Andersen, A.; Sepulveda, J.; Panapitiya, G. U.; Hollas, A.; Saldanha, E. G.; Murugesan, V.; Wang, W. SOMAS: a platform for data-driven material discovery in redox flow battery development. *Scientific Data*. **2022**, *9*. DOI: 10.1038/s41597-022-01814-4.
- (21) Zhang, Q.; Khetan, A.; Sorkun, E.; Niu, F.; Loss, A.; Pucher, I.; Er, S. Data-driven discovery of small electroactive molecules for energy storage in aqueous redox flow batteries. *Energy Storage Materials*. **2022**, *47*, 167-177. DOI: 10.1016/j.ensm.2022.02.013.
- (22) Cheng, L.; Assary, R. S.; Qu, X.; Jain, A.; Ong, S. P.; Rajput, N. N.; Persson, K.; Curtiss, L. A. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. *The Journal of Physical Chemistry Letters*. **2015**, *6*, 283-291. DOI: 10.1021/jz502319n.
- (23) Perera, A. S.; Suduwella, T. M.; Attanayake, N. H.; Jha, R. K.; Eubanks, W. L.; Shkrob, I. A.; Risko, C.; Kaur, A. P.; Odom, S. A. Large variability and complexity of isothermal solubility for a series of redox-active phenothiazines. *Materials Advances*. **2022**, *3*, 8705-8715. DOI: 10.1039/d2ma00598k.
- (24) D3TaLES. <https://d3tales.as.uky.edu/>.
- (25) Gallarati, S.; Van Gerwen, P.; Laplaza, R.; Vela, S.; Fabrizio, A.; Corminboeuf, C. OSCAR: an extensive repository of chemically and functionally diverse organocatalysts. *Chemical Science*. **2022**, *13*, 13782-13794. DOI: 10.1039/d2sc04251g.

- (26) Ai, Q.; Bhat, V.; Ryno, S. M.; Jarolimek, K.; Sornberger, P.; Smith, A.; Haley, M. M.; Anthony, J. E.; Risko, C. OCELOT: An infrastructure for data-driven research to discover and design crystalline organic semiconductors. *The Journal of Chemical Physics*. **2021**, *154*, 174705. DOI: 10.1063/5.0048714.
- (27) Curtarolo, S.; Setyawan, W.; Hart, G. L. W.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O.; et al. AFLOW: An automatic framework for high-throughput materials discovery. *Computational Materials Science*. **2012**, *58*, 218-226. DOI: 10.1016/j.commatsci.2012.02.005.
- (28) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*. **2013**, *1*, 011002. DOI: 10.1063/1.4812323.
- (29) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *The Journal of Physical Chemistry Letters*. **2011**, *2*, 2241-2251. DOI: 10.1021/jz200866s.
- (30) Omar, Ö. H.; Nematiram, T.; Troisi, A.; Padula, D. Organic materials repurposing, a data set for theoretical predictions of new applications for existing compounds. *Scientific Data*. **2022**, *9*. DOI: 10.1038/s41597-022-01142-7.
- (31) Huang, S.; Cole, J. M. A database of battery materials auto-generated using ChemDataExtractor. *Scientific Data*. **2020**, *7*. DOI: 10.1038/s41597-020-00602-2.
- (32) Ward, L.; Babinec, S.; Dufek, E. J.; Howey, D. A.; Viswanathan, V.; Aykol, M.; Beck, D. A. C.; Blaiszik, B.; Chen, B.-R.; Crabtree, G.; et al. Principles of the Battery Data Genome. *Joule*. **2022**, *6*, 2253-2271. DOI: 10.1016/j.joule.2022.08.008.
- (33) Duke, R.; Bhat, V.; Risko, C. Data storage architectures to accelerate chemical discovery: data accessibility for individual laboratories and the community. *Chemical Science*. **2022**, *13*, 13646-13656. DOI: 10.1039/d2sc05142g.
- (34) Andriuc, O.; Siron, M.; Montoya, J. H.; Horton, M.; Persson, K. A. Automated Adsorption Workflow for Semiconductor Surfaces and the Application to Zinc Telluride. *Journal of Chemical Information and Modeling*. **2021**, *61*, 3908-3916. DOI: 10.1021/acs.jcim.1c00340.
- (35) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*. **2013**, *68*, 314-319. DOI: 10.1016/j.commatsci.2012.10.028.
- (36) D3TaLES Database Documentation. <https://d3tales.as.uky.edu/docs/>.
- (37) Landrum, G. RDKit. **2010**.
- (38) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. **2020**, *17*, 261-272. DOI: 10.1038/s41592-019-0686-2.

- (39) Izmaylov, A. F.; Scuseria, G. E.; Frisch, M. J. Efficient evaluation of short-range Hartree-Fock exchange in large molecules and periodic systems. *The Journal of Chemical Physics*. **2006**, *125*, 104103. DOI: 10.1063/1.2347713.
- (40) *Gaussian 16 Rev. A.03*; Wallingford, CT, 2016. (accessed).
- (41) Henderson, T. M.; Izmaylov, A. F.; Scalmani, G.; Scuseria, G. E. Can short-range hybrids describe long-range-dependent properties? *The Journal of Chemical Physics*. **2009**, *131*, 044108. DOI: 10.1063/1.3185673.
- (42) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*. **2005**, *7*, 3297. DOI: 10.1039/b508541a.
- (43) Preet Kaur, A.; Neyhouse, B. J.; Shkrob, I. A.; Wang, Y.; Harsha Attanayake, N.; Kant Jha, R.; Wu, Q.; Zhang, L.; Ewoldt, R. H.; Brushett, F. R.; et al. Concentration-dependent Cycling of Phenothiazine-based Electrolytes in Nonaqueous Redox Flow Cells. *Chemistry – An Asian Journal*. *n/a*, e202201171. DOI: <https://doi.org/10.1002/asia.202201171>.
- (44) Casselman, M. D.; Kaur, A. P.; Narayana, K. A.; Elliott, C. F.; Risko, C.; Odom, S. A. The fate of phenothiazine-based redox shuttles in lithium-ion batteries. *Physical Chemistry Chemical Physics*. **2015**, *17*, 6905-6912. DOI: 10.1039/c5cp00199d.
- (45) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*. **2016**, *72*, 171-179. DOI: 10.1107/s2052520616003954.
- (46) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*. **2015**, *55*, 2324-2337. DOI: 10.1021/acs.jcim.5b00559.
- (47) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*. **2010**, *50*, 742-754. DOI: 10.1021/ci100050t.
- (48) Sowndarya S. V., S.; St. John, P. C.; Paton, R. S. A quantitative metric for organic radical stability and persistence using thermodynamic and kinetic features. *Chemical Science*. **2021**, *12*, 13158-13166. DOI: 10.1039/d1sc02770k.
- (49) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. **2020**.
- (50) Cihan Sorkun, M.; Mullaj, D.; Koelman, J. M. V. A.; Er, S. ChemPlot, a Python Library for Chemical Space Visualization**. *Chemistry-Methods*. **2022**, *2*. DOI: 10.1002/cmt.202200005.
- (51) D3TaLES API Docs. https://d3tales.github.io/d3tales_api/.
- (52) Fielding, R. T.; Taylor, R. N. Principled design of the modern Web architecture. *ACM Transactions on Internet Technology*. **2002**, *2*, 115-150. DOI: 10.1145/514183.514185.
- (53) D3TaLES Google Collaboratory Calculators. <https://d3tales.as.uky.edu/tools/calculators>.

(54) <https://d3tales.as.uky.edu/database/06TNKR/>.

(55) Poater, A.; Cosenza, B.; Correa, A.; Giudice, S.; Ragone, F.; Scarano, V.; Cavallo, L. SambVca: A Web Application for the Calculation of the Buried Volume of N-Heterocyclic Carbene Ligands. *European Journal of Inorganic Chemistry*. **2009**, 2009, 1759-1766. DOI: 10.1002/ejic.200801160.

(56) S. V., S. S.; Law, J. N.; Tripp, C. E.; Duplyakin, D.; Skordilis, E.; Biagioni, D.; Paton, R. S.; St. John, P. C. Multi-objective goal-directed optimization of de novo stable organic radicals for aqueous redox flow batteries. *Nature Machine Intelligence*. **2022**. DOI: 10.1038/s42256-022-00506-3.

(57) Peng, J.; Schwalbe-Koda, D.; Akkiraju, K.; Xie, T.; Giordano, L.; Yu, Y.; Eom, C. J.; Lunger, J. R.; Zheng, D. J.; Rao, R. R.; et al. Human- and machine-centred designs of molecules and materials for sustainability and decarbonization. *Nature Reviews Materials*. **2022**. DOI: 10.1038/s41578-022-00466-5.

(58) Omar, Ö. H.; Del Cueto, M.; Nematiram, T.; Troisi, A. High-throughput virtual screening for organic electronics: a comparative study of alternative strategies. *Journal of Materials Chemistry C*. **2021**. DOI: 10.1039/d1tc03256a.

(59) Kunkel, C.; Margraf, J. T.; Chen, K.; Oberhofer, H.; Reuter, K. Active discovery of organic semiconductors. *Nature Communications*. **2021**, 12. DOI: 10.1038/s41467-021-22611-4.

(60) Pyzer-Knapp, E. O.; Suh, C.; Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Aspuru-Guzik, A. What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery. *Annual Review of Materials Research*. **2015**, 45, 195-216. DOI: 10.1146/annurev-matsci-070214-020823.

(61) Kowalski, J. A.; Casselman, M. D.; Kaur, A. P.; Milshtein, J. D.; Elliott, C. F.; Modekrutti, S.; Attanayake, N. H.; Zhang, N.; Parkin, S. R.; Risko, C.; et al. A stable two-electron-donating phenothiazine for application in nonaqueous redox flow batteries. *Journal of Materials Chemistry A*. **2017**, 5, 24371-24379. DOI: 10.1039/c7ta05883g.

(62) Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*. **2009**, 1, 8. DOI: 10.1186/1758-2946-1-8.

(63) Milshtein, J. D.; Kaur, A. P.; Casselman, M. D.; Kowalski, J. A.; Modekrutti, S.; Zhang, P. L.; Harsha Attanayake, N.; Elliott, C. F.; Parkin, S. R.; Risko, C.; et al. High current density, long duration cycling of soluble organic active species for non-aqueous redox flow batteries. *Energy & Environmental Science*. **2016**, 9, 3531-3543. DOI: 10.1039/c6ee02027e.

(64) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials*. **2016**, 15, 1120-1127. DOI: 10.1038/nmat4717.

(65) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G. M.; Hautier, G.; et al. FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurrency and Computation: Practice and Experience*. **2015**, 27, 5037-5059. DOI: 10.1002/cpe.3505.

TOC

