

Supplementary Information

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. Odom¹ and 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

Table of Contents

1. Database Schema Flexibility	2
2. Database Technology Stack	2
3. Database Molecular Structures	3
Molecule Generation	3
Database Structure Samplings	5
4. Computational Data	6
Properties of interest from DFT	6
High-throughput computational workflow	9
Benchmark studies	10
5. Funnel Workflow	11
Calculation times	11
Structures resulting from the funnel workflow	12
6. D ³ TaLES API: More Information	16
7. References	17

1. Database Schema Flexibility

Both the D³TaLES frontend and backend schema are broad and can accommodate various data types with varying conditions including but not limited to computations at different levels of theory or considering different conditions, a broad range of experiments with various processing or data collection conditions, and literature-extracted data from learning models.

While the molecule-centric frontend schema is convenient for machine and human interpretability and searchability, it is difficult for a molecule-centric schema alone to accommodate all the conditions and variables for computational and wet lab experiments, especially when there may be multiple experiments with different conditions producing multiple values for the same property. For this reason, the D³TaLES data infrastructure consists of two databases: An experiment-centric backend database and a molecule-centric frontend database. The backend database includes all metadata and all calculated/measured values for each experiment instance. Importantly, each “data” attribute in a backend database instance has an attribute titled “conditions.” The “conditions” attribute contains all identifying conditions metadata that are likely to affect the experiment calculated/measured properties. For example, “conditions” for molecular DFT backend instances include code name and version, functional, basis set, tuning parameter, solvent, etc. “Conditions” for an electrochemistry experiment may include concentrations, supporting electrolytes, electrodes, solvents, temperature, etc. Complete documentation of D³TaLES conditions can be found in the D³TaLES database documentation.

When an experiment-derived property is pushed to the frontend database, the property value remains connected with the backend experiment ID, so a user can always find all the metadata for the original experiment. Additionally, the parent experiment’s “conditions” attribute is stored with the property value in the frontend database. Finally, a frontend property for a given molecule (e.g., oxidation potential) is not a single value. It is an array, where each item in the array has a value and associated conditions. Thus, for example, the property oxidation potential for a molecule might have three items, each one measured with different experimental conditions.

For more detailed discussion of the D³TaLES data schema, readers are encouraged to see reference the full documentation including an interactive visualization of the data schema.¹

2. Database Technology Stack

Both the frontend and backend D³TaLES databases are implemented with MongoDB.² Both databases use JSON Schema³ style schema that are publicly available.⁴ The D³TaLES website is built and hosted through Django⁵ and WSGI.⁶ The databases and website are located on OpenStack Virtual Machines at the University of Kentucky. Compressed raw data files and database backups are stored on a storage cluster at the University of Kentucky. All Database management and processing code are compiled in the D³TaLES API⁷ (discussed more in SI section 6).

3. Database Molecular Structures

Molecule Generation

The molecule generation process created potential redox-active molecules from fragments of molecules common in the field of NARFB. 10 backbones and 20 substituents commonly occurring in literature-reported materials for NARFB were selected (**Figure S3**).⁸⁻¹⁰ To create the synthetic dataset, backbones and substituents were rendered as graphs via RDKit¹¹ and NetworkX.¹² The nodes on the generated graphs had a one-to-one mapping to the atom IDs on the corresponding RDKit molecules. Node attributes were added to denote atom type, if RDKit flagged the atom as aromatic, and if the atom appeared in a ring system within a molecule. The edges had a one-to-one mapping to the molecular bonds. Likewise, edge attributes denoted bond type.

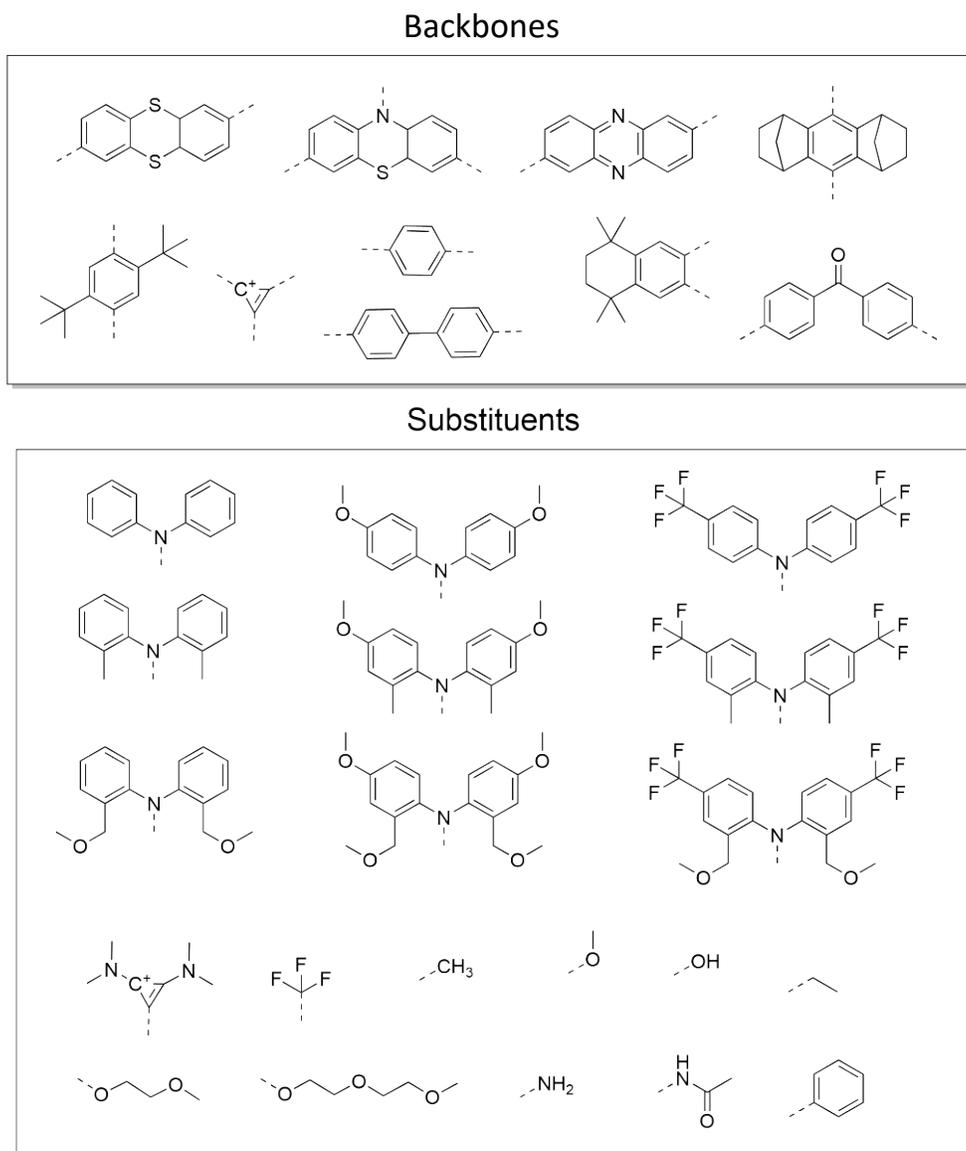


Figure S3. Ten backbones and twenty substituents commonly occurring in literature-reported materials for NARFB.

When joining substituent and backbone fragments, we created a relabeling map between the node IDs in the fragment and the sum of each node ID with the number of nodes in the scaffold it would be joined to. The fragment was then relabeled with this new map to create an edge between the relabeled node ID on the fragment with the labeled node ID on the scaffold. The resulting edge was denoted as a single bond type. This process can be repeated for an arbitrary number of fragments to be added on a given scaffold.

For the combination here, we manually selected the indices on which to combine molecular fragments with molecular scaffolds and iteratively combined each pair of indices. Rendering a molecule from the generated graph begins with an empty RDKit molecule. For each node in the generated graph, an atom is added to the new molecule with the type denoted in the node's attributes. Likewise, for each edge in the graph, a bond is added to the new molecule between the corresponding atom IDs. To verify the validity of the new molecule, the RDKit molecule is converted to SMILES.

Database Structure Samplings

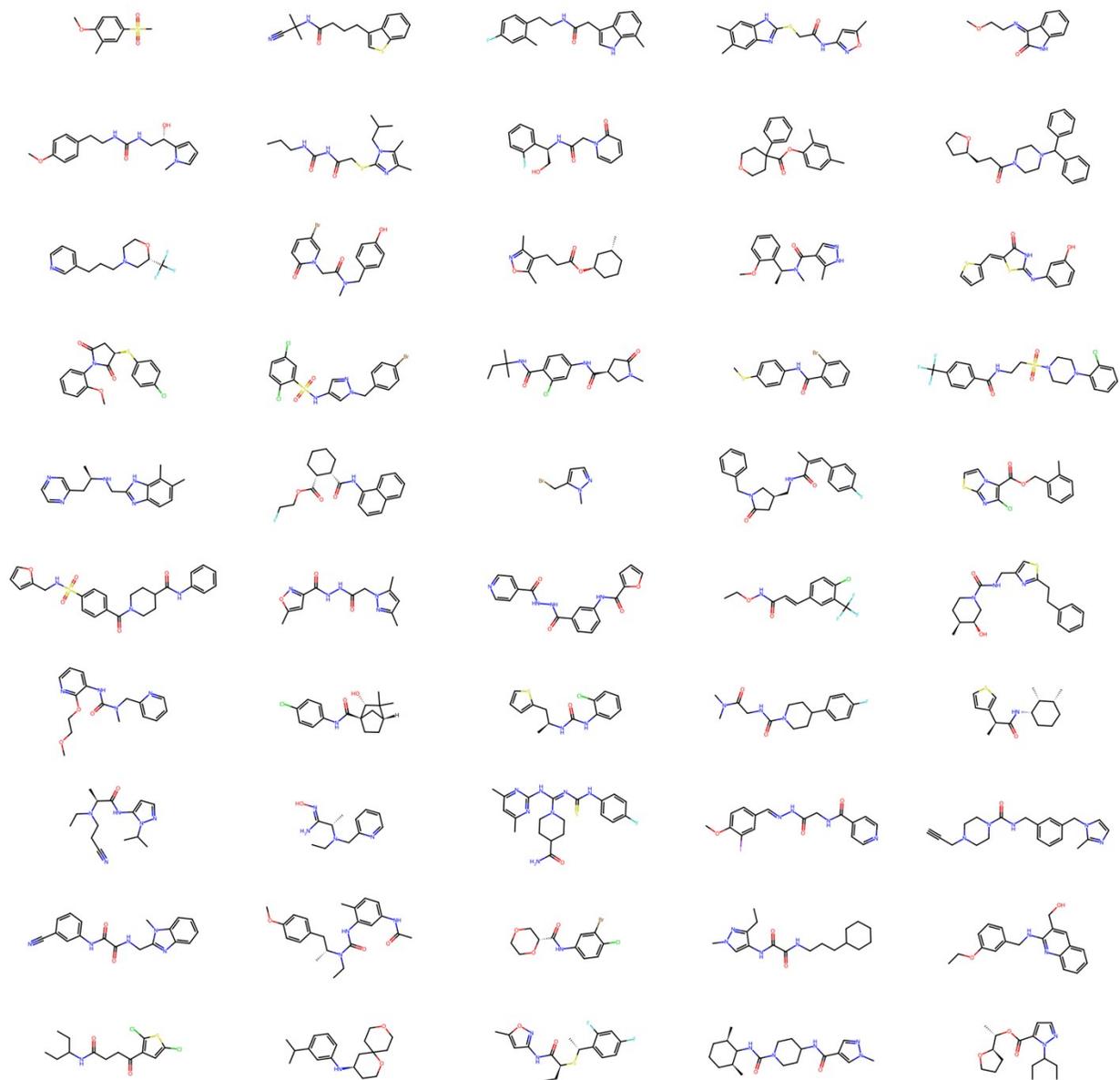


Figure S4. Fifty randomly selected molecules from the cleaned ZINC dataset.

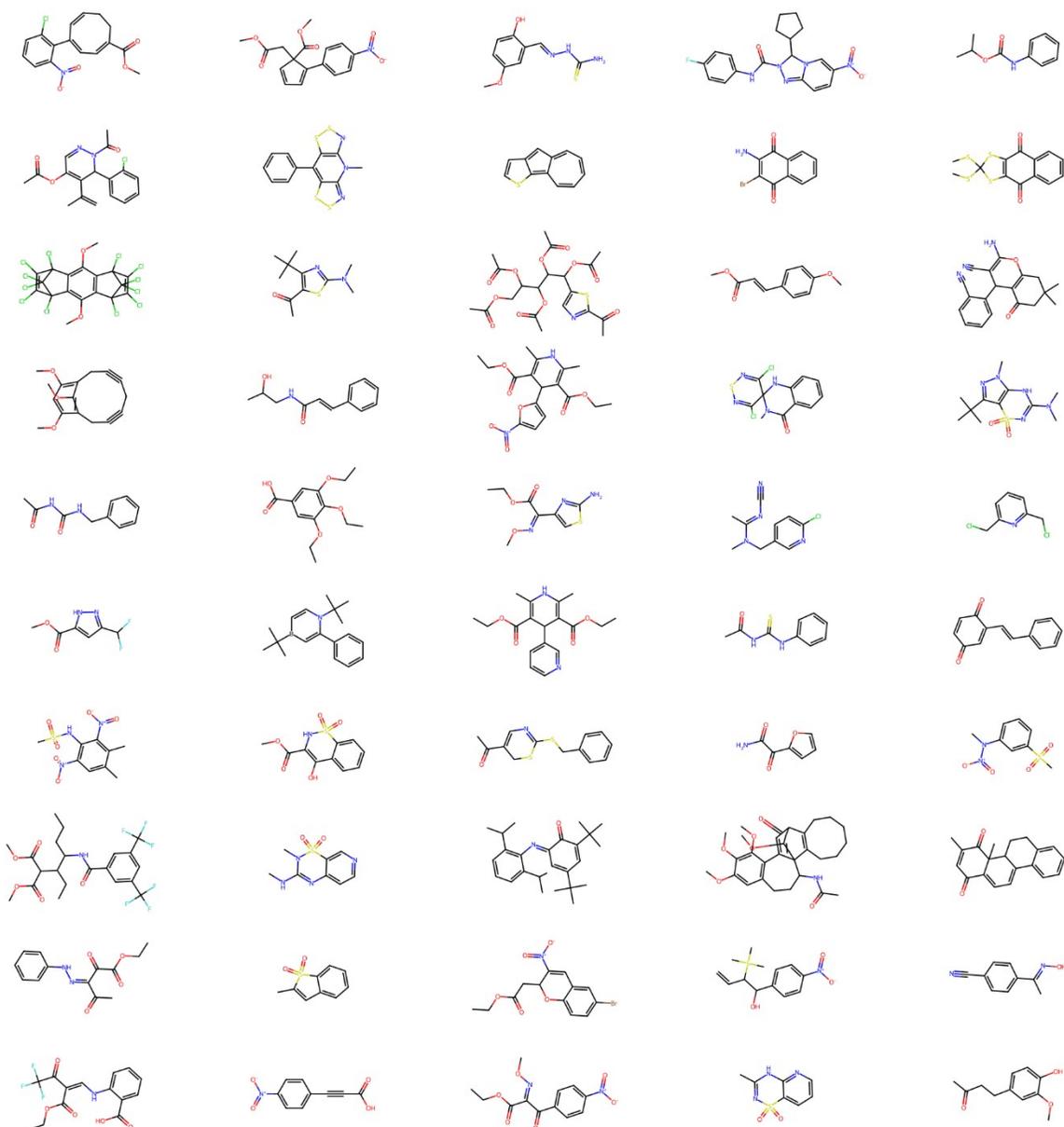


Figure S5. Fifty randomly selected molecules from the cleaned CSD dataset.

4. Computational Data

Properties of interest from DFT

To establish the D³TaLES database of redox-active small organic molecules, a high-throughput density functional theory (DFT) workflow was created and implemented to populate the database with quantum-chemical information about each molecule. The following section discusses several properties of interest that were determined through this process.

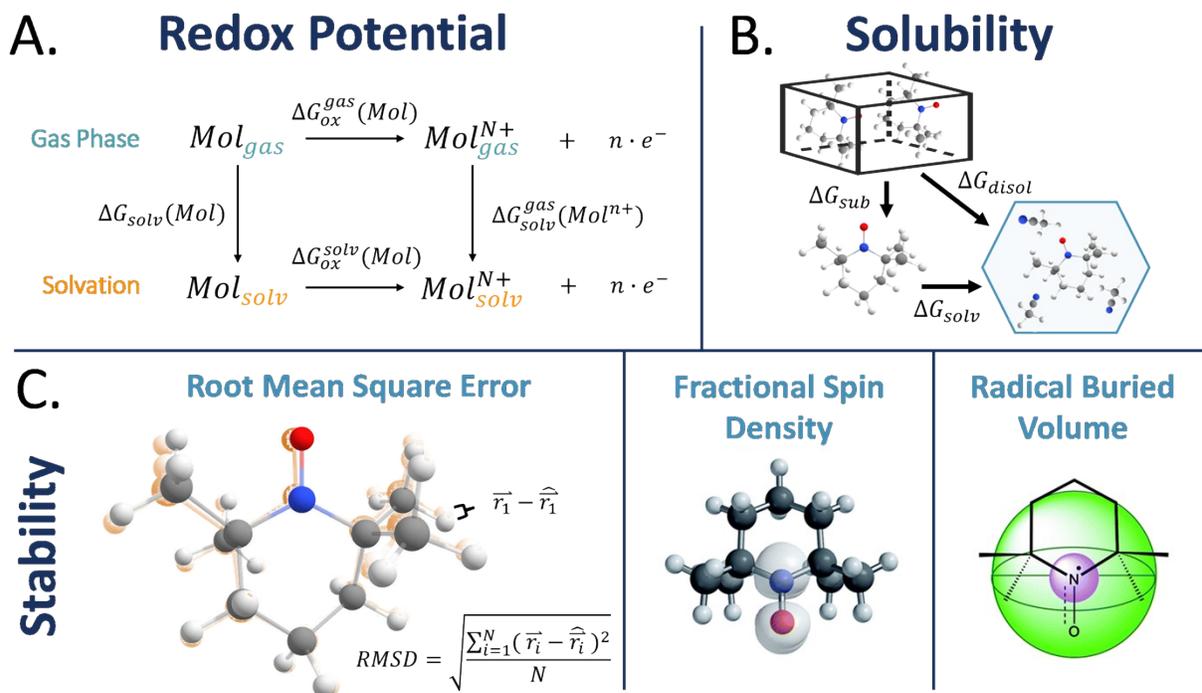


Figure S1. Methods for calculating various properties of interest with DFT, specifically (A) redox potential, (B) solubility (adapted from Cheng et. al.¹³), and (C) stability with three different methods (fraction spin density and radical buried volume adapted from Sowndarya et. al.¹⁴).

Oxidation and Reduction Potentials

Molecular oxidation and reduction potentials convey a molecule's tendency to lose and gain an electron, respectively.¹⁵ In designing molecules for NARFB, molecules should be relatively resistant to redox events, having a low reduction potential (analyte) or high oxidation potential (catholyte). This increases the voltage window in which the battery can operate. Especially in NARFB, it is crucial to have a large potential window since the increased solvent potential window is one of the primary reasons for using nonaqueous solvents over water.

D³TaLES uses a detailed approximation for oxidation (and reduction) potentials involving six DFT calculations, based on the Born-Haber cycle. For a given molecule, DFT geometry optimizations are performed for the ground-state molecule and the first oxidation (or reduction) event, both in the gas phase. Frequency calculations follow and serve two purposes: (1) confirm the optimized geometry is the true minimum energy geometry and (2) determine the Gibbs free energy correction for the calculated energies. For each of the gas phase ground state and oxidized (reduced) molecules, the free energy is the enthalpy found in the molecule's optimization calculations plus its Gibbs free energy correction. Then, single-point energy calculations are performed for each of the optimized molecules in an acetonitrile solution simulated by a polarized continuum implicit solvent model. Acetonitrile is among the most common solvents used for non-aqueous redox flow batteries because of its stability, relatively wide voltage range, and availability.¹⁶⁻²⁴ However, we note that (as explained in section 1) the D³TaLES data schema is broad and versatile, meaning that in the future the database could include computational and/or experimental data using other solvents. The Gibbs free energy corrections for the ground state

and cation gas phase calculations are assumed to be equivalent to the corrections for their respective solvent calculations.

As the Born-Haber cycle dictates, these four free energies produce the change in Gibbs free energy for oxidation in solution ($\Delta G_{soln,ox}^\circ$) (**Figure S1A**). With this energy, the molecule's oxidation potential ($\Delta E_{cell,ox}^\circ$) relative to the standard hydrogen electrode is found with Equation S1 where 4.42 eV is the potential of the standard hydrogen electrode.²⁵ The analogous equation is used for reduction potential.

$$\Delta E_{cell,ox}^\circ = \frac{\Delta G_{soln,ox}^\circ}{-nF} + 4.42 \text{ eV} \#S1$$

Solubility

High solubility for catholyte and analyte materials in RFB is a critical characteristic. Redox-active molecules that have larger solubility can achieve greater charge per volume in a battery, increasing the battery capacity. Solubility matters for all states – i.e., the ground state and oxidized/reduced states. Unfortunately, solubility is difficult to estimate computationally.¹³ The thermodynamic method for predicting solubility involves relating the dissolution change in Gibbs free energy with intrinsic solubility, S (**Figure S1B**).

$$\Delta G_{disol.} = \Delta G_{sub} + \Delta G_{solv} = -RT \ln S V_m \#S2$$

In Equation S2, R and T are the ideal gas constant and temperature, ΔG_{disol} is Gibbs free energy of dissolution, ΔG_{sub} is Gibbs free energy of sublimation, ΔG_{solv} is Gibbs free energy of solvation, and V_m is the molar volume of the molecule in crystalline form. While R and T are known and ΔG_{solv} can be estimated as the difference between gas phase and solvated energies, ΔG_{sub} and ΔV_m are unknown without the molecule's crystal structure, which is very difficult to predict for novel molecules. While it is possible to estimate solubility with the energy of solvation alone, this is often inaccurate. To mitigate these challenges, several efforts have been reported to develop machine learning (ML) methods to predict solubility.²⁶⁻²⁸

Stability

Among the most challenging properties to predict and optimize is stability. Like solubility, stability matters for all states of charge. In RFB, any unstable species will degrade and prevent reversible electrochemical processes. Additionally, unstable intermediates can react with solvent molecules or other redox-active species. These factors impede cyclability and shorten battery lifetime. Detailed computational analysis can calculate reaction energies for bond dissociations and various side reactions. However, it is difficult to consider every possible side reaction and, given the molecule-specific nature of the side reactions, it is nearly impossible to automate the calculations to run in a high-throughput manner. Other methods roughly estimate the stability of a system (**Figure S1C**). For example, it is possible to calculate the root mean squared error (RMSE) between the ground state atomic coordinates and those of a redox species. This method assumes that a notable change in geometry occurring with a redox event indicates destabilization, which is often the case. Often, RMSE above 0.1 Å indicate unstable redox species.¹³ This, of course, does not consider the original stability of the ground state species or potential side reactions.

The highest fraction of electronic spin density that is concentrated on one atom, another measure of stability, estimates thermodynamic stability. This metric quantified the highest fraction of electronic spin density that is concentrated on one atom.¹⁴ Alternatively, the kinetic stability can be estimated with the percent radical buried volume, which captures the extent to which the radical (the atom with the highest spin) is sheltered sterically. Sowndarya et. al. combined the fractional spin density and buried volume metrics to produce a radical stability score.¹⁴

Synthetic Accessibility Score

Here we incorporate the synthetic accessibility score developed by Ertl and Schuffenhauer²⁹ and implemented by RDKit¹¹. While there is some debate about the efficacy of this score and users are encouraged to use the score carefully,³⁰ it is widely used in the molecular screening.³¹⁻³⁵ For a more detailed explanation of the synthetic accessibility score, see Ertl and Schuffenhauer's original work.²⁹

High-throughput computational workflow

The workflow first initializes a molecule and performs an initial DFT geometry optimization before completing (ionization-potential) IP fitting to tune the ω parameter for the LC- ω HPBE functional, tuning the amount of long and short-range exchange (**Figure S2**).³⁶ This tuned functional is used for all subsequent calculations. Next, the workflow splits into five species workflows: one each for the ground state, mono- and di-cation, and mono- and di-anion states (assuming in the discussion here, for the sake of clarity, that we start from a neutral, closed-shell ground state). Di-ion calculations are included to identify systems with potential for multiple redox events, which would increase battery energy density.^{18, 37} Each species geometry is optimized and confirmed to be a minimum on the potential energy surface with vibrational normal mode analysis. The ground state, mono-cation state, and mono-anion state workflows include single-point energy calculations of the ground state geometry with a mono-cation/anion charge and mono-cation/anion geometry with the ground state charge. The workflow also contains a single-point energy calculation performed in an implicit solvent (most frequently in acetonitrile). All solvation calculations are performed with the implicit polarized continuum model (PCM), estimating acetonitrile with a dielectric constant ϵ of 35.688. Oxidation potentials estimated with the Born-Haber cycle method³⁸ were benchmarked with experimentally measured oxidation potentials. (See SI for more details.) All five species workflows include time-dependent DFT (TDDFT) calculations to evaluate the properties of the low-lying electronic excited states. The TDDFT calculations produce absorption spectra, which are useful in validating several experimental observables (e.g., identifying molecular species, determining stability over time, assessing aggregation in the context of solubility, etc.). While some calculations fail for various reasons (most often unconverged geometry optimizations), approximately 31,000 molecules have completed the oxidative computational workflow.

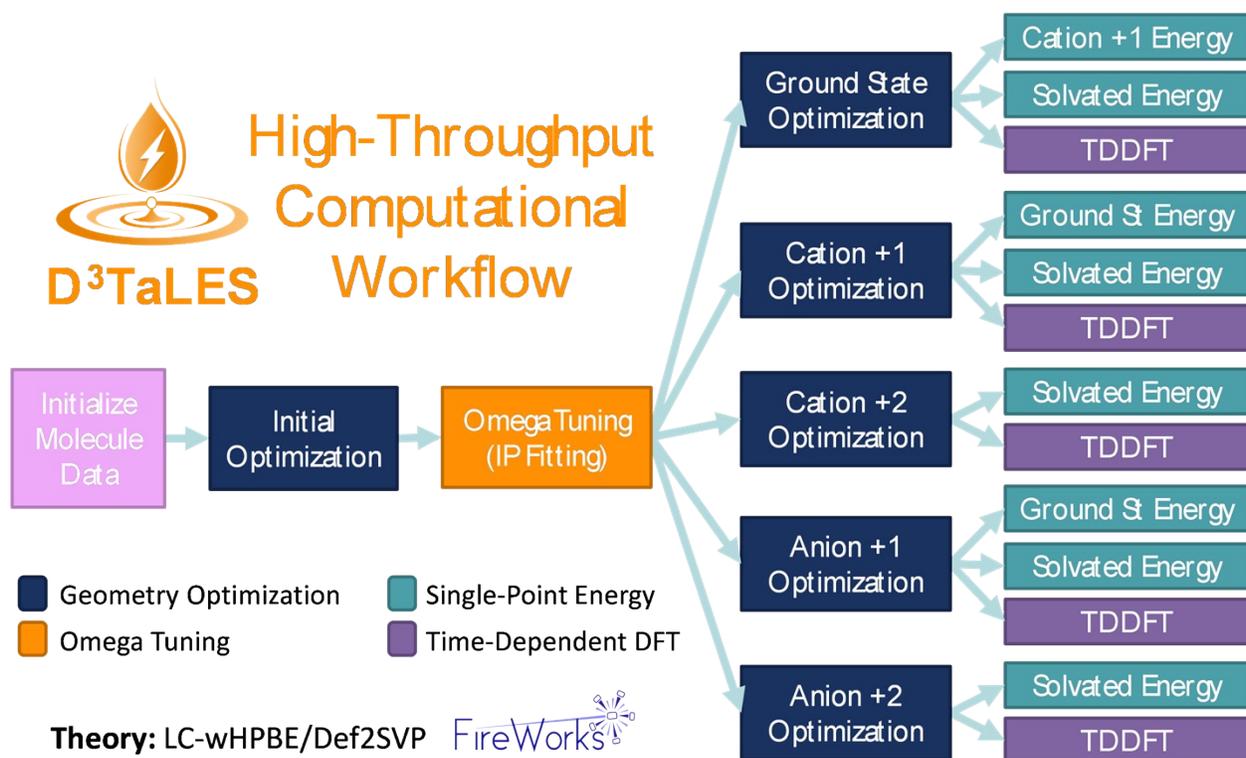


Figure S2. Computational workflow for molecular DFT on D³TaLES molecules. Pink indicates molecular initiation, navy indicates optimization, orange indicates IP fitting, teal green indicates a single-point energy calculation, and light blue indicates time-dependent DFT (TD-DFT).

Benchmark studies

After construction, the workflow was implemented and evaluated. First, calculations were performed for several known molecules, some published and some unpublished. These calculations were manually inspected to ensure the workflow ran the appropriate calculations.

For the second benchmark stage, the computational oxidation potential estimations were benchmarked with a dataset already benchmarked in the literature.³⁸ The study contained a set of small organic molecules (~1-2 rings) with experimental oxidation potentials. Calculations were completed for 118 molecules. Overall, the experimentally known and computationally predicted redox potentials show good agreement except for a few outliers (**Figure S6A**). Next, we benchmarked a series of 20 more complex systems developed for NARFB applications. In these systems, the predicted oxidation potential does not necessarily align with the measured value (**B**). This is unsurprising, especially considering different electrodes and solvents were used. However, clear trends emerge. As the predicted oxidation potential increases, so do the experimental potentials relative to potentials with the same conditions.

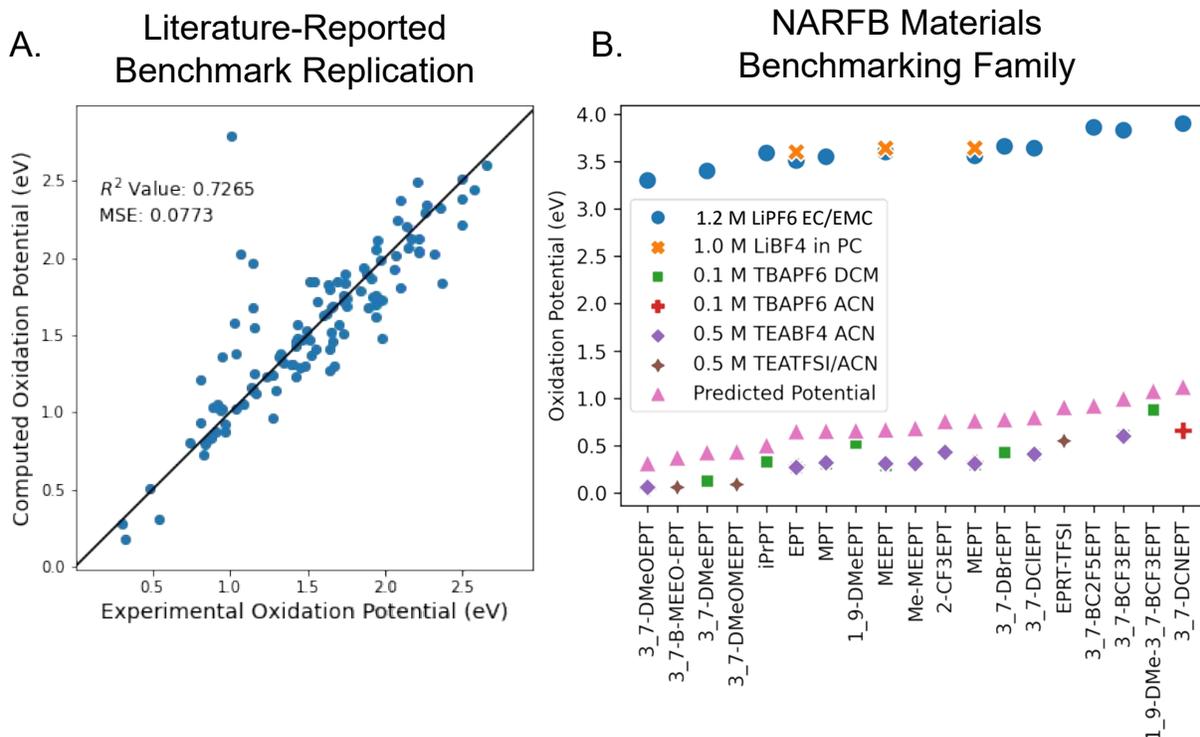


Figure S6. Benchmarking data for oxidation potentials. (A) Computationally predicted vs measured oxidation potential for the published benchmarking dataset.³⁸ (B) Measured and calculated oxidation potentials vs molecule name for 20 more complex systems developed for use in NARFB. Notice pink triangles show computationally predicted values, while other points show measured values for various electrodes and solvents. Abbreviations include EC for ethylene carbonate, EMC for ethyl methyl carbonate, PC for propylene carbonate, and ACN for acetonitrile.

5. Funnel Workflow

Calculation times

All core-hour estimations for the funnel workflow properties are estimated based on computing times for jobs performed on the TACC Stampede2 Intel Xeon Skylake (SKX) nodes. More information about these computing nodes can be found at <https://docs.tacc.utexas.edu/hpc/stampede2/>.

Structures resulting from the funnel workflow

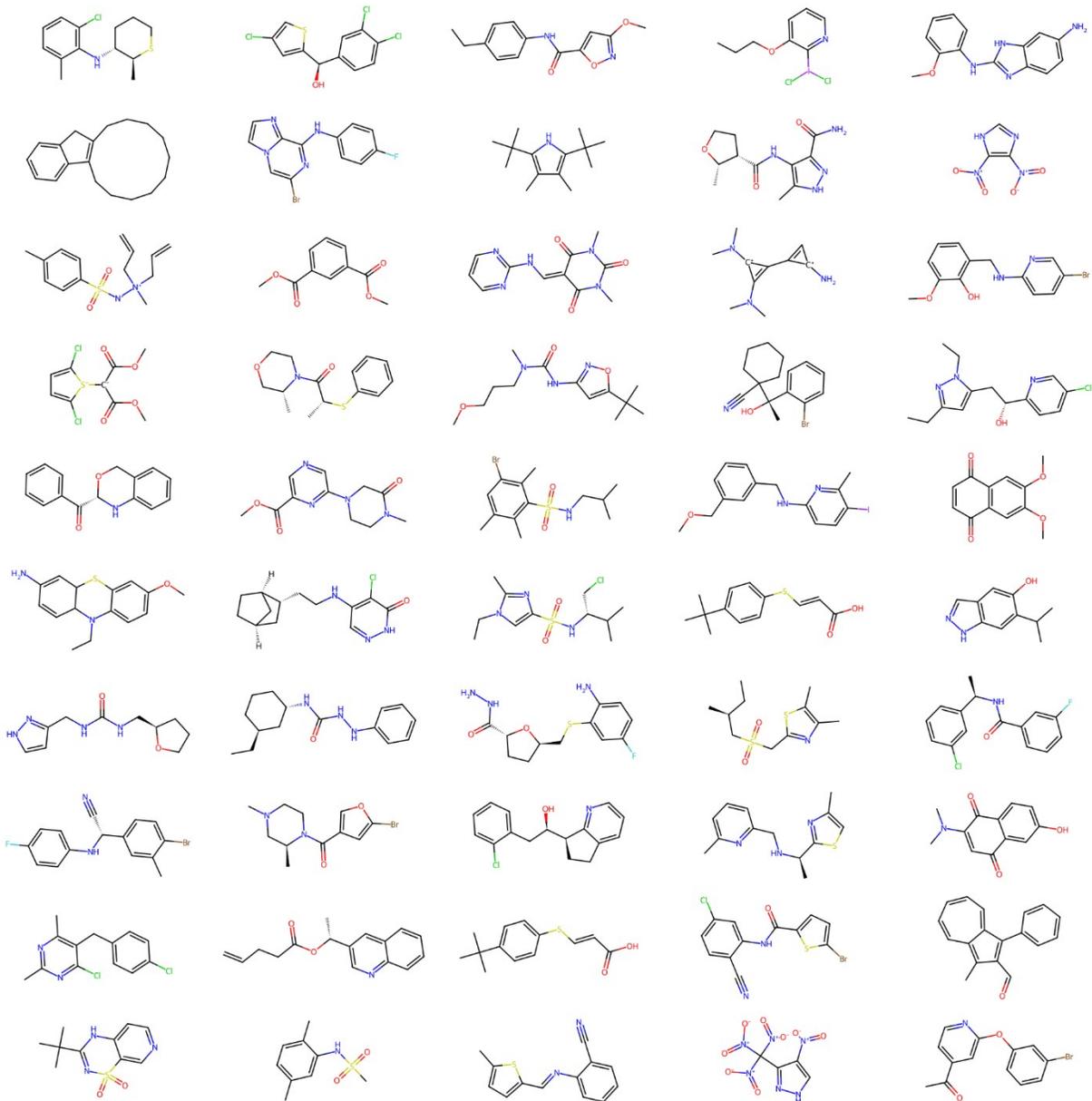


Figure S7. Fifty randomly selected structures from the 12,132 molecules that passed the size test.

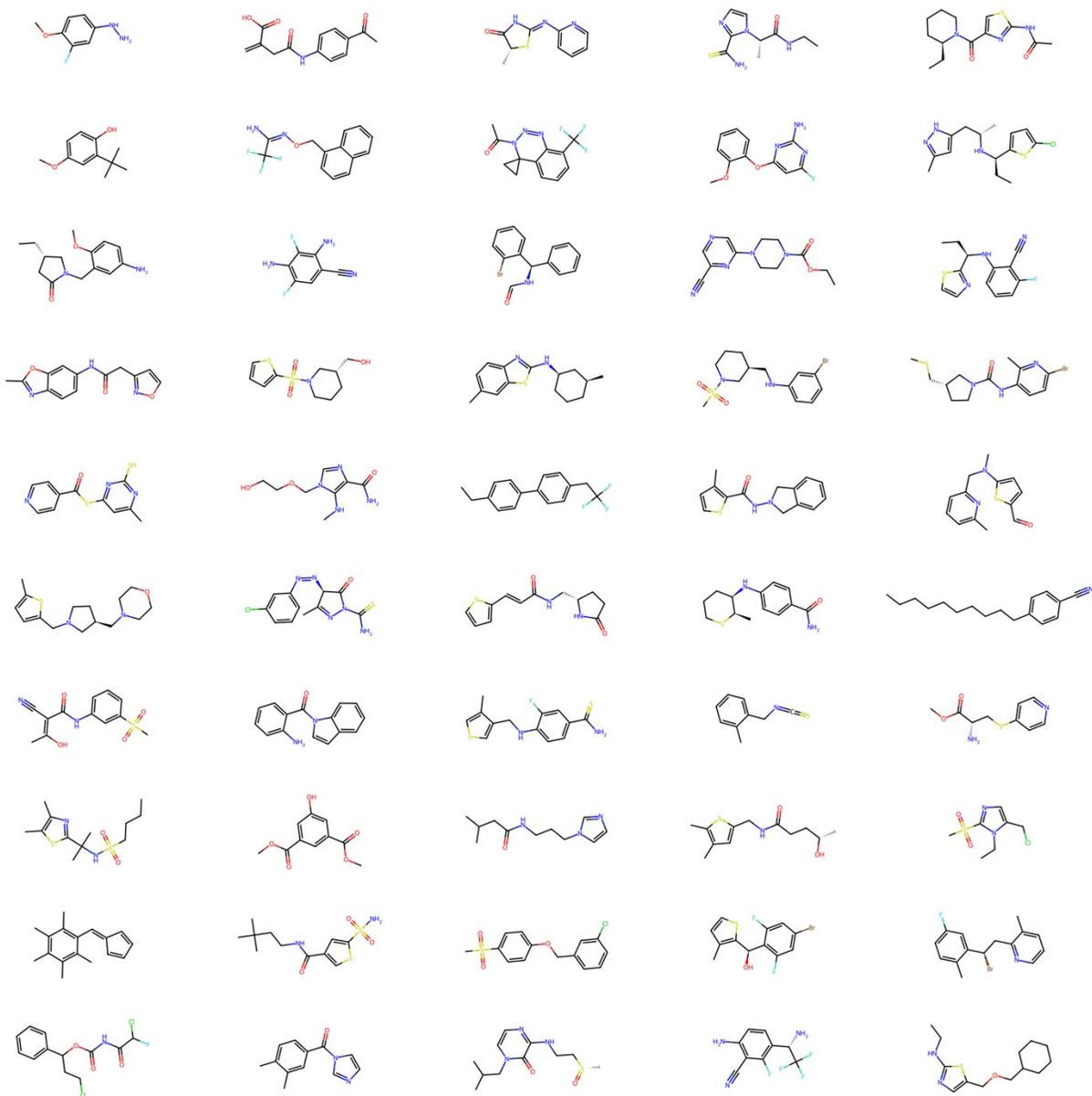


Figure S8. Fifty randomly selected structures from the 11,728 molecules that passed the size and synthetic stability score tests.

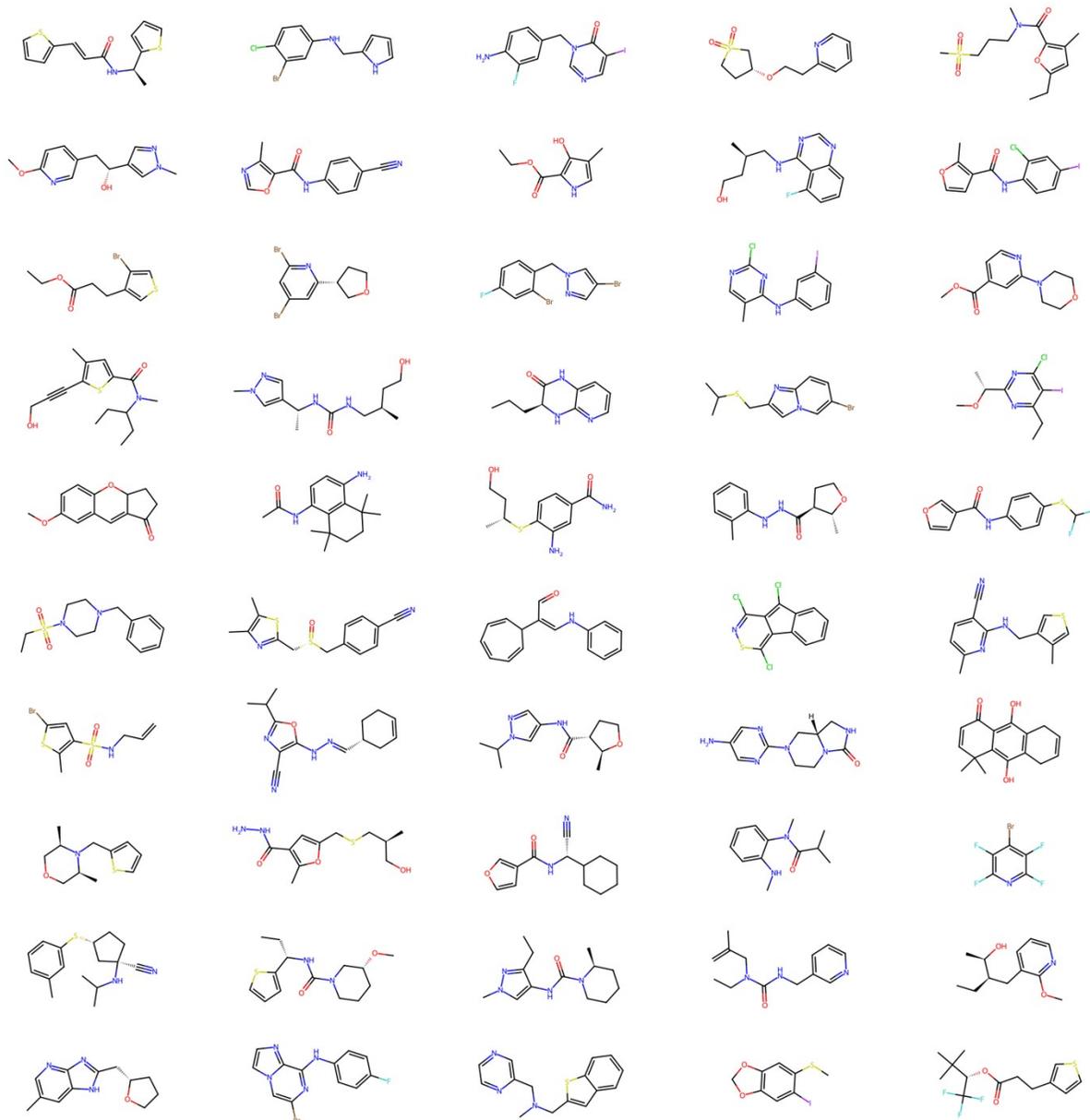


Figure S9. Fifty randomly selected structures from the 3,735 molecules that passed the size, synthetic stability score, and stability tests.

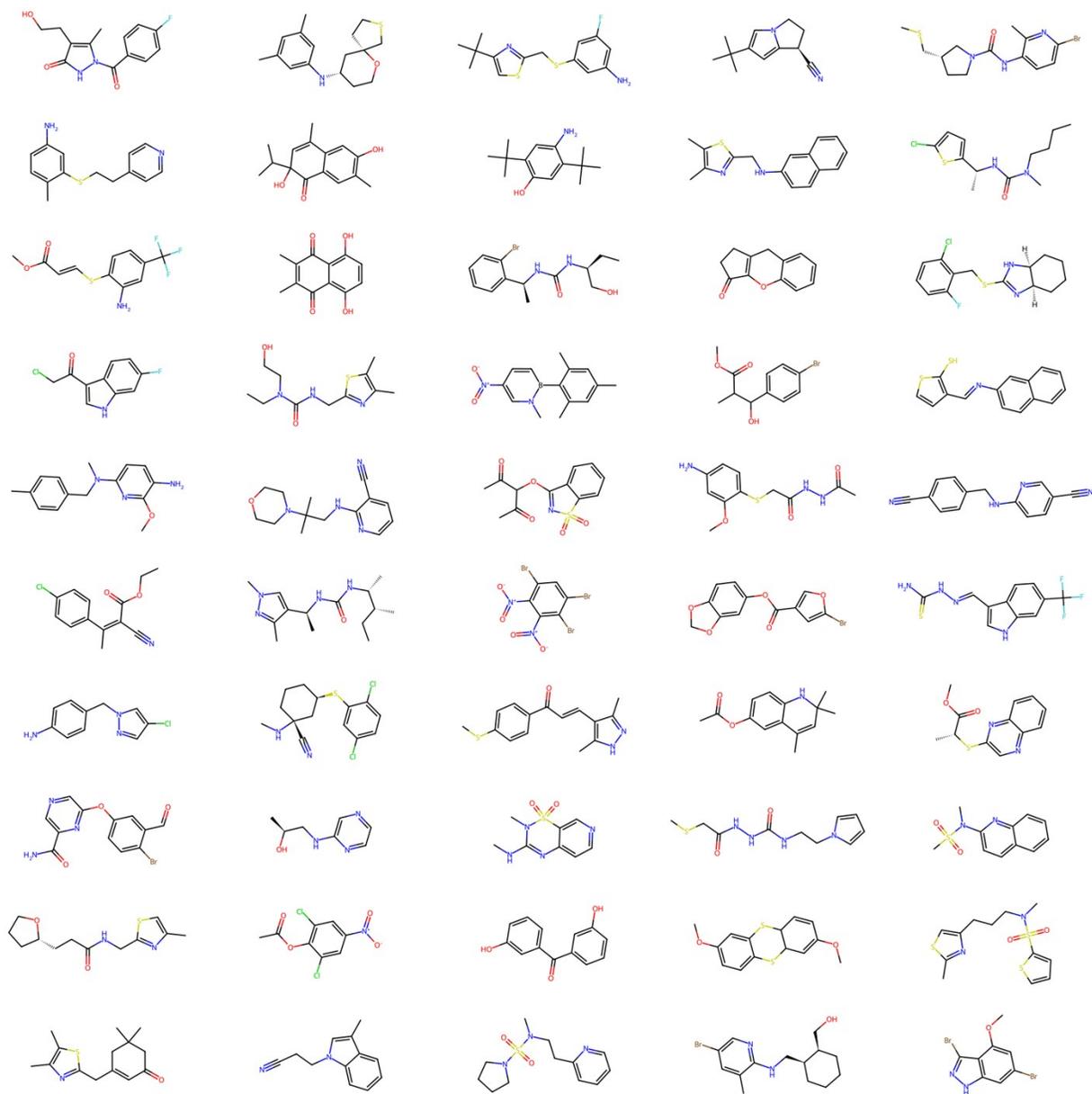


Figure S10. Fifty randomly selected structures from the 2,856 molecules that passed the size, synthetic stability score, stability, and solvation tests.

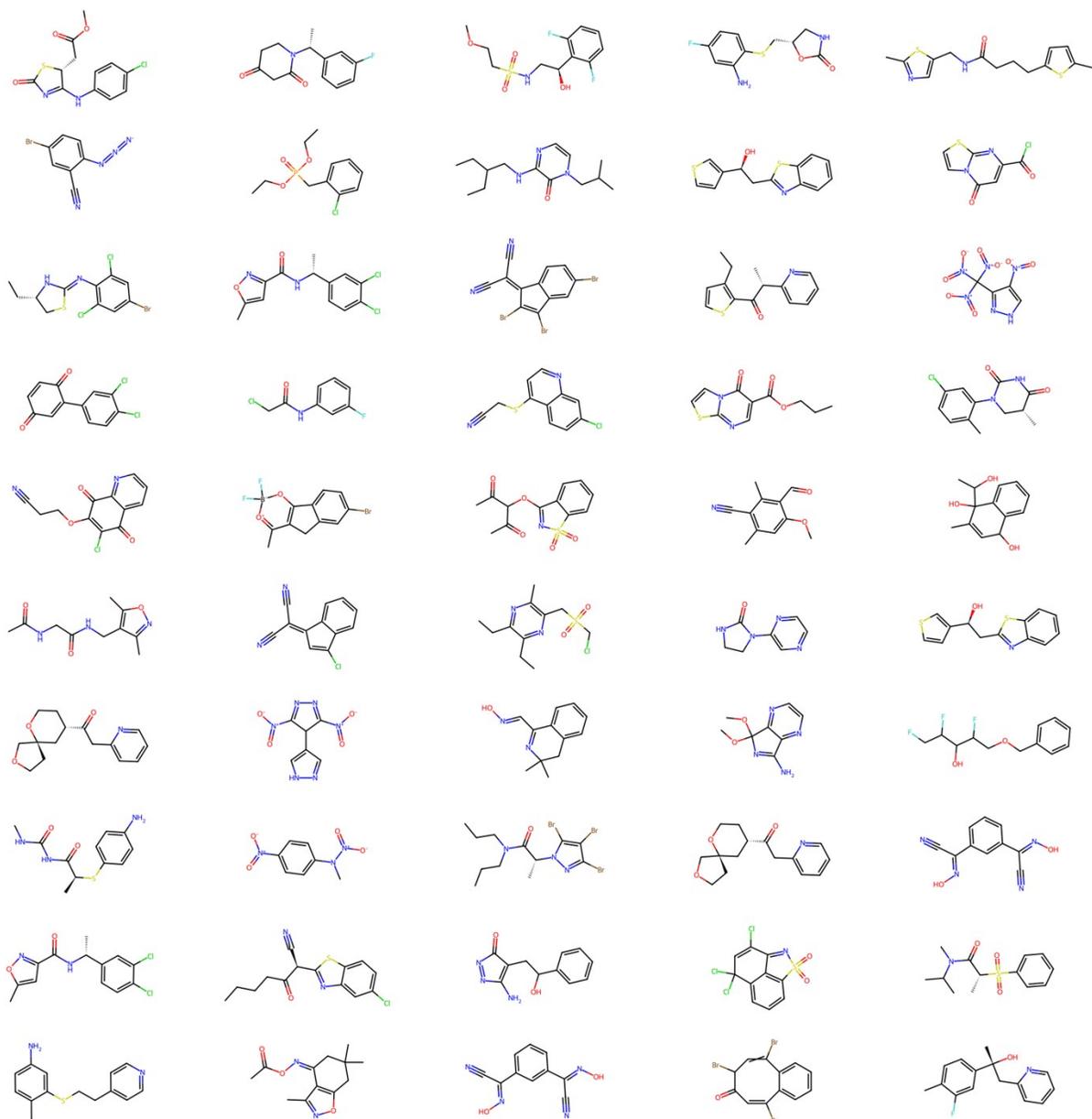


Figure S11. Fifty randomly selected structures from the 364 molecules that passed all five (size, synthetic stability score, stability, solvation, and oxidation) tests.

6. D³TaLES API: More Information

The full documentation for the D³TaLES API is available at https://d3tales.github.io/d3tales_api/.

The Processors module contains various parsing classes for extracting useful data from instrument-produced computational and experimental data files. It currently supports cyclic voltammetry data from Pine WaveNow instruments and UV/Vis data in Excel format. The module also supports computational data from Gaussian log files. Extracted properties include all listed properties in the D3TaLES schema. All parsing classes can format extracted data to fit the D3TaLES schema.¹ This module also contains a class

for transforming data from the backend D³TaLES database to the format for the frontend D³TaLES database; however, this is likely not as useful to the general community.

The D3database module contains features most useful to D³TaLES database administrators such as data insertions/validation and direct database queries (both of which require user authentication). However, it does contain 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.⁷

The Calculators module provides tools that allow users to calculate useful experimental and computational properties from nested data, and all calculators contain unit conversion features. Useful molecular DFT calculators include reorganization energy, root mean squared error between atoms in two geometries, change in Gibbs energy for solvation, redox potential, radical buried volume³⁹, and radical spin density.³³ Useful cyclic voltammetry calculators include solution concentration general CV properties (peaks, reversibility, E_{1/2}, peak splitting, etc.), diffusion constant using the Randles-Scidwick equation and charge transfer rate. While the D³TaLES API documentation explains basic usage for these calculators, we also provide Google Collaboratory notebooks that use the calculators to perform calculations without the need for the user to know python coding.⁴⁰

7. References

- (1) D3TaLES Database Documentation. <https://d3tales.as.uky.edu/docs/>.
- (2) MongoDB. <https://www.mongodb.com/> (accessed June 25 2022).
- (3) JSONSchema. <https://json-schema.org/> (accessed June 6 2023).
- (4) D3TaLES Schema. <https://github.com/D3TaLES/schema>.
- (5) Django Software Foundation. Django 2.2. 2019. <https://djangoproject.com>.
- (6) WSGI. <https://wsgi.readthedocs.io/>.
- (7) D3TaLES API Docs. https://d3tales.github.io/d3tales_api/.
- (8) 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/acseenergylett.9b01332.
- (9) 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.
- (10) Armstrong, C. G.; Toghiani, K. E. Stability of molecular radicals in organic non-aqueous redox flow batteries: A mini review. *Electrochemistry Communications*. **2018**, *91*, 19-24. DOI: 10.1016/j.elecom.2018.04.017.
- (11) Landrum, G. RDKit. **2010**.
- (12) Aric A. Hagberg, D. A. S. a. P. J. S. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Pasadena, CA USA, Aug 2008, pp 11–15.
- (13) 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.

- (14) 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.
- (15) Elgrishi, N.; Rountree, K. J.; Mccarthy, B. D.; Rountree, E. S.; Eisenhart, T. T.; Dempsey, J. L. A Practical Beginner's Guide to Cyclic Voltammetry. *Journal of Chemical Education*. **2018**, *95*, 197-206. DOI: 10.1021/acs.jchemed.7b00361.
- (16) D3TaLES Schema Conditions. https://d3tales.as.uky.edu/docs/properties/conditions_and_more.html.
- (17) De La Garza, G. D.; Kaur, A. P.; Shkrob, I. A.; Robertson, L. A.; Odom, S. A.; Mcneil, A. J. Soluble and stable symmetric tetrazines as anolytes in redox flow batteries. *Journal of Materials Chemistry A*. **2022**, *10*, 18745-18752. DOI: 10.1039/d2ta04515j.
- (18) 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.
- (19) Huo, Y.; Xing, X.; Zhang, C.; Wang, X.; Li, Y. An all organic redox flow battery with high cell voltage. *RSC Advances*. **2019**, *9*, 13128-13132. DOI: 10.1039/c9ra01514k.
- (20) Kwon, G.; Lee, S.; Hwang, J.; Shim, H.-S.; Lee, B.; Lee, M. H.; Ko, Y.; Jung, S.-K.; Ku, K.; Hong, J.; et al. Multi-redox Molecule for High-Energy Redox Flow Batteries. *Joule*. **2018**, *2*, 1771-1782. DOI: 10.1016/j.joule.2018.05.014.
- (21) Liu, B.; Tang, C. W.; Zhang, C.; Jia, G.; Zhao, T. Cost-Effective, High-Energy-Density, Nonaqueous Nitrobenzene Organic Redox Flow Battery. *Chemistry of Materials*. **2021**, *33*, 978-986. DOI: 10.1021/acs.chemmater.0c04118.
- (22) 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.
- (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) 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.
- (25) Batista, J. L. P. a. V. S. Tutorial on Ab Initio Redox Potential Calculations. Department of Chemistry, Yale University, P.O. Box 208107, New Haven CT 06520-8107.
- (26) Sorkun, M. C.; Koelman, J. M. V. A.; Er, S. Pushing the limits of solubility prediction via quality-oriented data selection. *iScience*. **2021**, *24*, 101961. DOI: 10.1016/j.isci.2020.101961.
- (27) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. Predicting aqueous solubilities from aqueous free energies of solvation and experimental or calculated vapor pressures of pure substances. *The Journal of Chemical Physics*. **2003**, *119*, 1661-1670. DOI: 10.1063/1.1579474.
- (28) Ye, Z.; Ouyang, D. Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms. *Journal of Cheminformatics*. **2021**, *13*. DOI: 10.1186/s13321-021-00575-3.
- (29) 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.
- (30) Foscatto, M.; Jensen, V. R. Automated in Silico Design of Homogeneous Catalysts. *ACS Catalysis*. **2020**, *10*, 2354-2377. DOI: 10.1021/acscatal.9b04952.
- (31) Moraes, A. S.; Pinheiro, G. A.; Lourenço, T. C.; Lopes, M. C.; Quiles, M. G.; Dias, L. G.; Da Silva, J. L. F. Screening of the Role of the Chemical Structure in the Electrochemical Stability Window of Ionic Liquids:

- DFT Calculations Combined with Data Mining. *Journal of Chemical Information and Modeling*. **2022**, *62*, 4702-4712. DOI: 10.1021/acs.jcim.2c00748.
- (32) Nigam, A.; Pollice, R.; Aspuru-Guzik, A. Parallel Tempered Genetic Algorithm Guided by Deep Neural Networks for Inverse Molecular Design. *Digital Discovery*. **2022**. DOI: 10.1039/d2dd00003b.
- (33) 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.
- (34) Bagal, V.; Aggarwal, R.; Vinod, P. K.; Priyakumar, U. D. MolGPT: Molecular Generation Using a Transformer-Decoder Model. *Journal of Chemical Information and Modeling*. **2021**. DOI: 10.1021/acs.jcim.1c00600.
- (35) 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.
- (36) Stein, T.; Kronik, L.; Baer, R. Reliable Prediction of Charge Transfer Excitations in Molecular Complexes Using Time-Dependent Density Functional Theory. *Journal of the American Chemical Society*. **2009**, *131*, 2818-2820. DOI: 10.1021/ja8087482.
- (37) Yan, Y.; Vogt, D. B.; Vaid, T. P.; Sigman, M. S.; Sanford, M. S. Development of High Energy Density Diaminocyclopropenium-Phenothiazine Hybrid Catholytes for Non-Aqueous Redox Flow Batteries. *Angewandte Chemie International Edition*. **2021**, *60*, 27039-27045. DOI: 10.1002/anie.202111939.
- (38) Guerard, J. J.; Arey, J. S. Critical Evaluation of Implicit Solvent Models for Predicting Aqueous Oxidation Potentials of Neutral Organic Compounds. *Journal of Chemical Theory and Computation*. **2013**, *9*, 5046-5058. DOI: 10.1021/ct4004433.
- (39) 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.
- (40) D3TaLES Google Collaboratory Calculators. <https://d3tales.as.uky.edu/tools/calculators>.