



Cite this: DOI: 10.1039/d6ta00771f

# From machine learning to chemical insight: Darwinian chance and the stability of charge carriers in flow batteries

Jingjing Zhang,<sup>ab</sup> Lily A. Robertson,<sup>ab</sup> Ilya A. Shkrob,<sup>\*ab</sup> Rajeev S. Assary<sup>ac</sup> and Lu Zhang<sup>ab</sup>

In chemistry, rationalizing and predicting reaction behavior in complex environments is challenging, but this knowledge is required for practical applications and materials advancement. Here, we focus on one such example that uses organic molecules as redox-active materials in electrochemical energy storage. In flow batteries, these molecules (also known as redoxmers) serve as charge carriers, and exceptional chemical stability in all states of charge is required for long-term use. Here, we show how machine learning combined with chemist's knowledge can be used to reveal patterns in the reactivity of charged redoxmers by providing mechanistically tractable clues.

Received 26th January 2026  
Accepted 17th May 2026

DOI: 10.1039/d6ta00771f

rsc.li/materials-a

## Introduction

Darwin famously observed that “chance...serves to acknowledge plainly our ignorance of the cause of each particular variation”. This maxim applies equally to biology and electrochemical storage. In complex environments, charged molecules can react in multiple ways, and it is frequently impossible to isolate these chemical pathways for greater scrutiny under more controlled conditions. This hurdle makes predicting and/or rationalizing the reaction patterns difficult as they appear erratic. The complex interplay of multiple factors causes large variations in properties across a trial set; chemists select the desired properties and modify materials for further trials. This iterative approach can be successful, but it is also time- and labor-consuming.

Redox flow batteries (RFBs) are scalable electrochemical devices in which charge separation and charge storage are compartmentalized.<sup>1</sup> Such devices can be used in the electric grid to integrate intermittent power sources<sup>2</sup> or provide backup and peak-shaving for energy-intensive operations such as data centers.<sup>3</sup> During battery charging, the positive and negative charge carriers are separated in electrochemical stacks and dispatched to storage tanks. During battery discharging, the stored charges are flowed back to the stacks, and the charge is collected. The positive and negative charge carriers are referred to as catholytes (posolytes) and anolytes (negolytes), respectively.

Redox-active organic molecules (redoxmers) can serve as anolytes and catholytes for nonaqueous flow cells (NRFBs). Redoxmers and metal complexes involving abundant elements are among the few economic charge-storage materials for grid-scale devices.<sup>4</sup>

The chief attraction of such flow cells is an extended electrochemical window compared with aqueous cells, which potentially allows operation at a higher voltage. To translate this advertised advantage into higher energy density, one needs both higher redox potentials and a high concentration of the carriers. Importantly, the charge carrier must have high solubility in the battery electrolyte; *i.e.*, the high concentration of supporting salts used to balance the charge often affects redoxmer solubility. Thus, catholyte molecules need to have high oxidation potentials, possess high solubility in all states of charge, and be exceptionally stable both during storage of charged fluids in a reservoir (calendar life stability) and electrochemical cycling (cycle life stability). As these properties can be at odds with each other, only a few families of redoxmers have been identified as attractive flow cell catholyte candidates, most notably cyclopropenium cations,<sup>5</sup> phenothiazines,<sup>6</sup> and dialkoxyarenes.<sup>7</sup>

In recent years, numerous derivatives of these redoxmers have been synthesized and tested to optimize electrochemical performance. Navigating the complex trade-offs inherent to this optimization would be improved by the ability to predict properties of interest. Unfortunately, many of these properties are difficult or impossible to compute from first principles.

Machine learning (ML) modeling is used across different branches of chemistry for statistical discovery of trends underlying the interplay of multiple factors.<sup>8</sup> There have been successful examples of using this methodology for flow cells, *e.g.*, for optimization of reactivity<sup>9</sup> and solubility within a redoxmer family.<sup>10,11</sup> The properties of interest for this study are the ones that “make or break” active materials as potential

<sup>a</sup>Joint Center for Energy Storage Research, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL 60439, USA

<sup>b</sup>Chemical Sciences and Engineering Division, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL 60439-4837, USA. E-mail: robertla@anl.gov; shkrob@anl.gov

<sup>c</sup>Materials Science Division, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL 60439-4837, USA



candidates: the electrochemical potentials (that define the energy density), the calendar life, and the cycling stability. The latter two properties can show high variability, appearing erratic and complicating modeling. Another roadblock is data scarcity: the existing training sets for homologous redoxmers are small (10–50 compounds). For such small sets, using multiple structural descriptors and fingerprints routinely used in ML modeling cannot be justified; on the other hand, using too few descriptors can be insufficient to capture variability in the data. Compounding this difficulty is the fact that molecular descriptors are typically designed for closed-shell molecules, whereas singly charged redoxmers are open-shell species, whose properties only somewhat correlate with the properties of neutral precursors. While there are many known approaches to constructing descriptors for neutral molecules,<sup>12</sup> there is a dearth of such methods for reactive, charged molecules such as radical ions.<sup>13,14</sup> Here, we demonstrate how more intuitive descriptors to characterize such species can be constructed and used in ML models using the strategy illustrated in Fig. 1.

## Results and discussion

### Data collection: calendar and cycle life

For this study, we explored non-symmetric and symmetric dialkoxarenes types 1 and 2 shown in Fig. 2, building on the original library studied in our group.<sup>15</sup> Several of these are new

compounds—complete characterization is given in the SI. The compounds are denoted as MeR or DiR for non-symmetric and symmetric molecules, respectively, and *R* is the alkyl substituent (Table 1).

In liquid electrolyte (here, 0.5 M lithium bistriflimide, LiTFSI, in acetonitrile at room temperature in a glovebox), these molecules undergo reversible one-electron oxidation (see cyclic voltammograms and tables in Section S3), and the half-wave potentials  $E_{1/2}$  of this electrochemical reaction vs. an Ag/Ag<sup>+</sup> quasi-reference electrode are given in Table 1.

Another useful metric is the half-life  $t_{1/2}$  of the radical cation (the oxidation product) that was determined in dilute redoxmer solutions (5–20 mM) to minimize cross reactions. Electron paramagnetic resonance (EPR) spectroscopy was used to observe the decay of the dialkoxarene radical cations over time, and these kinetics were numerically analyzed to estimate reaction times (Section S5). In this way, the calendar lifetime was obtained to characterize the chemical stability of charged redoxmer molecules in bulk electrolyte during charge storage in a tank.

The cycling stability was determined in symmetric stationary H-cells where the dialkoxarene is present in both compartments but at opposite states of charge (Section S4). This method allows study of the effects of parasitic reactions: in one compartment, the neutral molecule was oxidized to the radical cation, while in the other compartment, the radical cation was

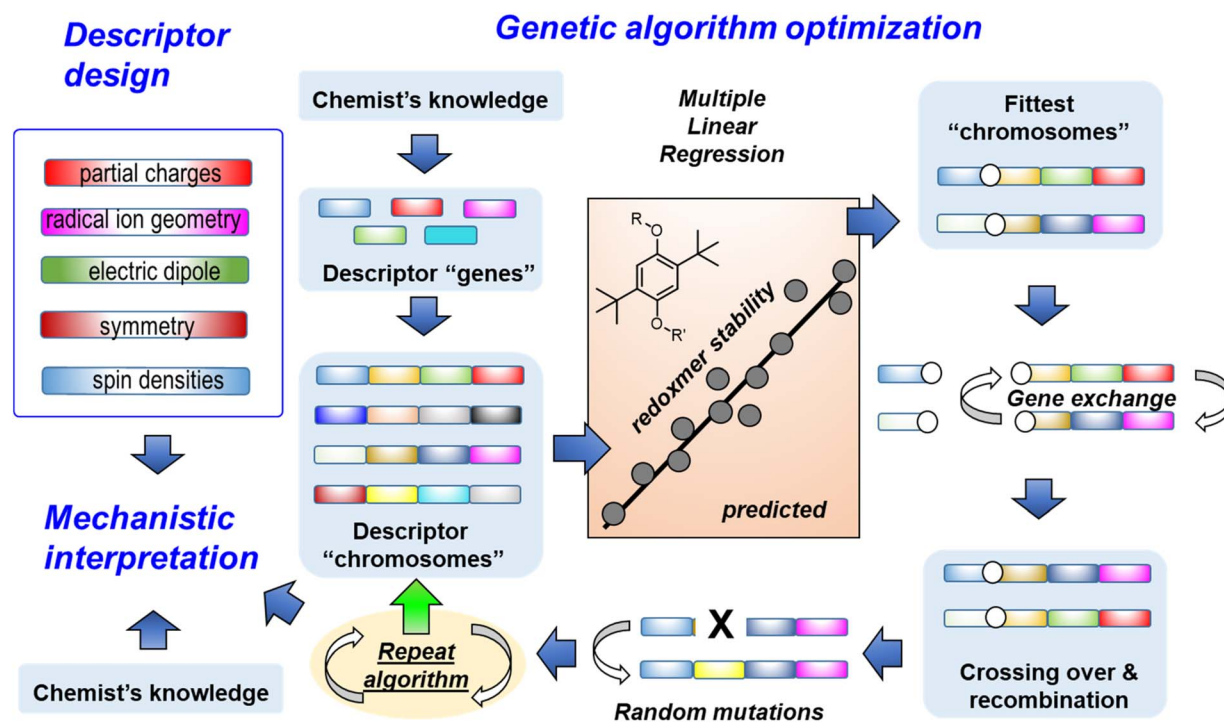


Fig. 1 Schematic representation of data analyses using the combination of qualitative structure–property relationship (QSPR) analyses and genetic algorithm (GA) optimization. The loop begins with the chemist's choice of descriptors reflecting the suspected reaction mechanisms. The chemist's knowledge informs the descriptor design and the initial "chromosome" set of these descriptors. QSPR analysis is performed for each "chromosome", and GA optimization is used to direct the evolution of the "chromosomes" by selective breeding and random mutations. When the algorithm finds the optimum sets of descriptors, the chemist's mechanistic knowledge is used to interpret these optimal sets using the fact that the descriptors reflect different aspects of reactivity.



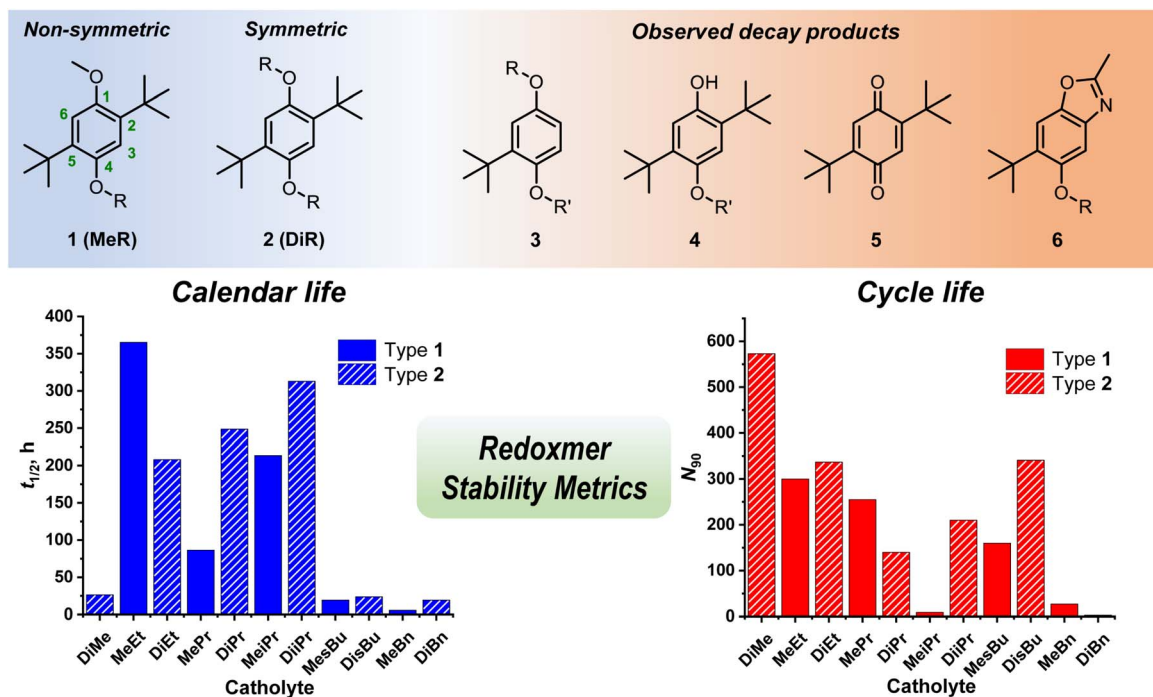


Fig. 2 The upper row: non-symmetric (type 1) and symmetric (type 2) dialkoxyarenes and volatile products 3 to 6 observed after electrolysis of these redoxmers in acetonitrile-based electrolytes. The lower row: bar plots of the calendar and cycle lives for redoxmers in Table 1.

Table 1 Experimental measurements for selected dialkoxyarenes

Catholyte	Type <sup>a</sup>	R <sup>b</sup>	E <sub>1/2</sub> , V vs. Ag/Ag <sup>+</sup> <sup>c</sup>	t <sub>1/2</sub> , h <sup>d</sup>	N <sub>90</sub> <sup>e</sup>
DiMe	2	Methyl	0.690	26.5	573
MeEt	1	Ethyl	0.673	365.2	300
DiEt	2	Ethyl	0.657	208	337
MePr	1	<i>n</i> -Propyl	0.702	86.3	255
DiPr	1	<i>n</i> -Propyl	0.661	248.6	140
MeiPr	1	<i>iso</i> -Propyl	0.644	213.5	9
DiiPr	2	<i>iso</i> -Propyl	0.598	312.9	210
MesBu	1	<i>sec</i> -Butyl	0.648	19.5	160
DisBu	2	<i>sec</i> -Butyl	0.594	23.7	341
MeBn	1	Benzyl	0.710	5.6	27
DiBn	2	Benzyl	0.731	19.5	3

<sup>a</sup> Symmetry type. <sup>b</sup> Substitution groups in Fig. 2. <sup>c</sup> Measured half-wave potential. <sup>d</sup> 50% decay lifetime for the radical cation. <sup>e</sup> Cycle number corresponding to 10% capacity loss at 50% SOC for 3C rate cycling at 20 mM.

reduced to the precursor. As the electrochemically active material is converted to products *via* parasitic reactions, the capacity decreases, and the cycling stability can be characterized by this capacity fade. In such diagnostic experiments, the cell is charged and discharged at a constant rate until a set capacity is reached or the potential limits are exceeded. These experiments were limited to 50% theoretical capacity with a 3C rate (1C means full discharge in 1 h) and a potential range of 0–1.8 V vs. Ag/Ag<sup>+</sup>. The number of full cycles N<sub>90</sub> that lead to 10% decay of the discharge capacity, calculated from initial discharge capacity (see Fig. 1 in ref. 16), was used as a metric for chemical stability during charge or discharge; thus, the two

metrics reflect the two basic regimes of operation in a flow cell. As shown in Fig. 2 and Table 1, these two stability metrics show high variability and do not correlate in an obvious way.

In these cycling experiments, except for the benzyl derivatives, for which N<sub>90</sub> was very low (Table 1), >10% of the capacity faded when >90% of the redoxmer still remained in the cell, as suggested by gas-chromatography mass spectrometry (GCMS), proton nuclear magnetic resonance spectroscopy (<sup>1</sup>H NMR) and high-performance liquid chromatography (HPLC) post-mortem analyses of the cycled electrolyte solutions (Section S7); thus, the loss of cyclability was not a simple matter of loss of the redox-active material to parasitic reactions. Rather, the cycling was terminated due to a rapid increase in the potential of the working electrode as illustrated for MeEt in Fig. 3, which suggests fouling of the electrodes by insoluble reaction products. Initially, the charging half-cycle terminated at a charging plateau at ~1.0 V vs. Ag/Ag<sup>+</sup> (Fig. 3a), but over time the end-cycle potentials increased, and a new plateau started to form at ~1.5 V vs. Ag/Ag<sup>+</sup> (Fig. 3b and c). In Fig. 3d, we plot the maximum potential at the end of each cycle. The initial slow growth that was linear with the cycle number was replaced with a rapid increase in the potential, at which time the potential control was lost, and the discharge capacity rapidly decreased. In some systems, the potential control was recovered after 1–3 h of rest at open circuit; the cycling resumed to once again halt in the same manner.

Such behavior would be consistent with the formation of poorly soluble insulating deposits on the electrode at the end of the cycle that slowly dissolved during rest.



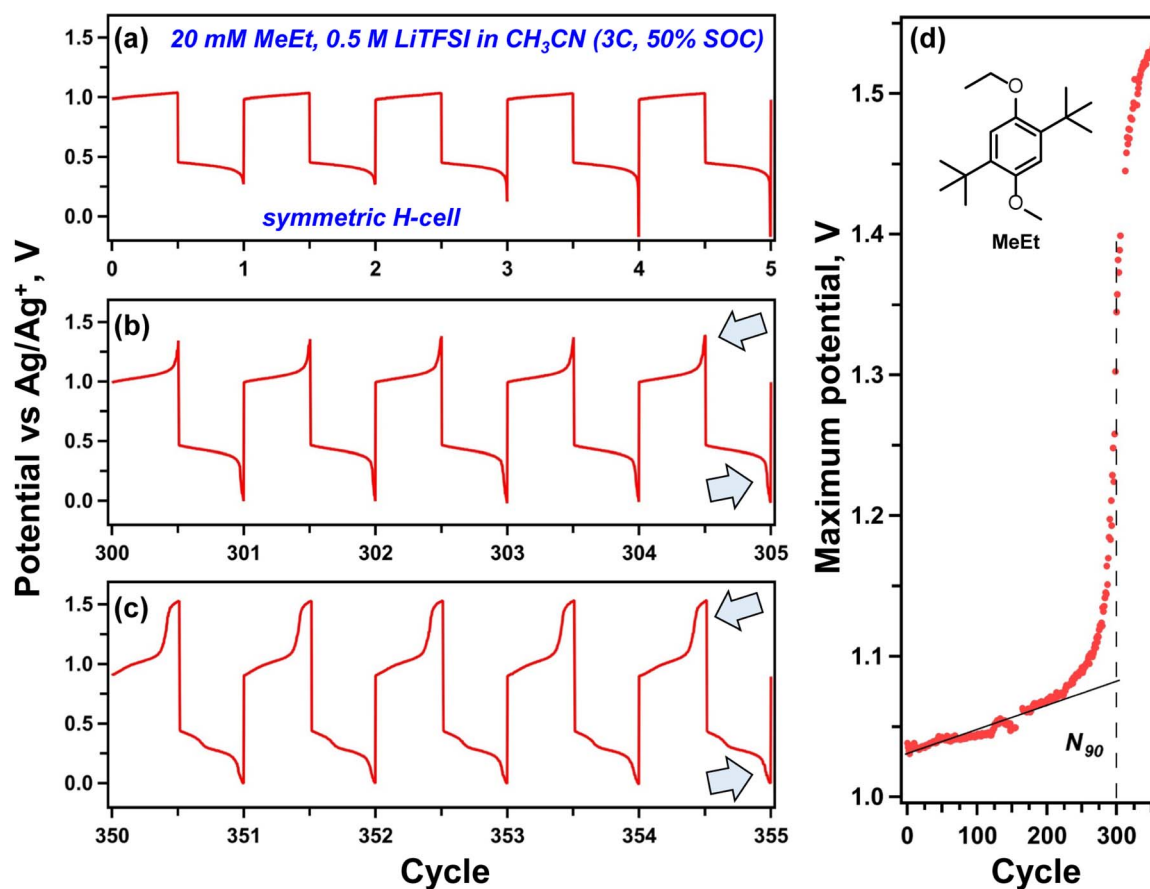


Fig. 3 Panels a to c show charge–discharge profiles (vs. Ag/Ag<sup>+</sup>) for the working electrode in a symmetric H-cell containing 20 mM MeEt in 0.5 M LiTFSI in acetonitrile (~30 °C, 3C cycling, 50% SOC). The arrows indicate features that appear abruptly at around cycle 300 when discharge capacity becomes abruptly lost, and the maximum potential during a cycle suddenly increases. Panel d demonstrates these voltage increases graphically with a rough trend line extrapolated to the  $N_{90}$  point based on initial discharge capacity.

### Parasitic reactions

Poor correlation between the calendar and cycling stability can arise for at least two reasons.<sup>16</sup> First, for many of the dialkoxyarene redoxmers, the main channel for the decay of their radical cation in the solvent bulk is charge neutralization. This reaction, if it is sufficiently slow on the time scale of a cycle, has relatively little impact on the cycle life as it simply regenerates the starting material. However, this neutralization reaction can have a significant impact on the radical cation lifetime in the solvent bulk.

Second, as observed above, for many dialkoxyarenes, the cyclability is not limited by redoxmer decomposition *per se* but rather by a rapid increase in the working electrode potential, as illustrated in Fig. 3. This overshoot occurs even when the redoxmer is still present in the electrolyte at the end of the cycle. We attribute this increase to the formation of a resistive film consisting of insoluble products whose overall yield is small. For this reason, their formation has little effect on the decay kinetics of a radical cation in the bulk, yet it greatly affects the cycle life.

We note that the molecules in Table 1 all share a 2,5-di-*tert*-butyl substitution.<sup>17</sup> These protective groups play an important

role in the stabilization of the radical cation by preventing proton loss coupled with radical addition to the arene ring of a neutral molecule.<sup>17–20</sup> The bulky groups in the ring provide steric hindrance to this reaction without inducing excessive strain in the radical cation. When the arene ring is tri- or tetra-substituted, the stability decreases as the resulting strain eases *O*-dealkylation of the radical cation.<sup>17,21</sup>

The problem with the 2,5-di-*tert*-butyl substitution is that the *tert*-butyl groups can become eliminated as carbocations, leaving the unprotected fragment 3 (Fig. 2).<sup>22</sup> The typical gas chromatogram showing product 3 generated in electrolysis of MeBn can be seen in Fig. S7.1 in Section S7. Adducts of deprotected 3 and precursor molecules were found by chromatography among the products of electrolysis in ref. 23; it is likely that still larger oligomers are formed as there is no impediment to ring addition and polymerization once the protecting groups become lost. The eliminated carbocation can subsequently undergo the Ritter reaction<sup>24,25</sup> with acetonitrile solvent (Fig. 4a and S7.1 and S7.2 in Section S7).<sup>23</sup> While the deprotection is a low-yield reaction (<2–5%), it can have dramatic consequences for the cycling lifetime as it leads to the



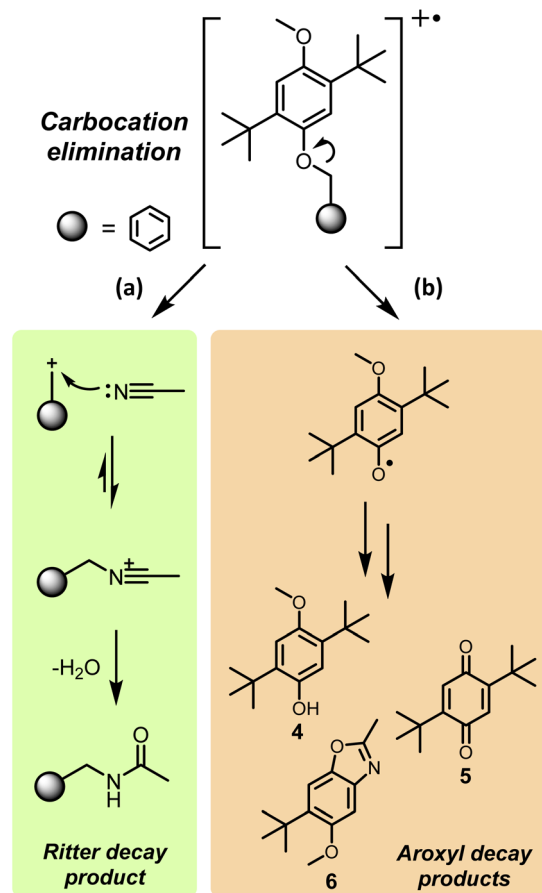


Fig. 4 Decomposition of a dialkoxyarene MeBn radical cation that fragments into (a) a benzyl carbocation and (b) an aroxyl radical. Further transformations of these species yield the chromatographically detected products shown in the panels.

formation of macromolecules near the electrodes, which can form deposits on the electrode surface.

The above applies to the radical cations whose lifetimes exceed the duration of a half-cycle. When the opposite is true, the short lifetime affects both stability metrics. For benzyl derivatives, the radical cations undergo rapid *O*-dealkylation by eliminating the benzyl carbocation and leaving the aroxyl radical (Fig. 4b). Our product analyses suggest that these aroxyl radicals abstract hydrogen (yielding phenols 4) or disproportionate (yielding quinones 5), while the eliminated carbocation ( $R^+$ ) undergoes the Ritter reaction yielding the corresponding nitrilium ion ( $RN=C^+Me$  or  $RN^+≡CMe$ ). During wet analysis, this ion hydrolyzes to *N*-benzylacetamide, which is observed chromatographically among the volatile products along with the phenol and quinone (see Fig. S7.1 and S7.2). A less expected reaction is the formation of a benzoxazole derivative 6, which is a product of the addition of the solvent molecule (acetonitrile) concerted with the removal of the *tert*-butyl group. This product can be identified through mass spectrometry and  $^1H$  NMR spectroscopy (Fig. S7.1 and S7.3). The multiplicity of reaction pathways and their different impacts on bulk and electrode

reactions explains the need for ML modeling in balancing different structural factors on these reactions.

### Data analyses

To analyze these kinetic data, genetic algorithm driven multiple multivariate linear modeling (MMLR) was used (see Section S6 for more detail). The experimental measurements  $Y = \{y_{ik}\}$  ( $i = 1, \dots, N_y$ , variables for  $k = 1, \dots, n$  measurements) are correlated with the explanatory variables  $X = \{x_{jk}\}$  ( $j = 1, \dots, N_x$  variables) that can include structural descriptors. Our goal was to find the smallest set of explanatory variables  $X$  that accounts for most of the variation in the data  $Y$ . In MMLR for large data sets, this task can be achieved using ridge, Lasso,<sup>26</sup> or elastic net regularization.<sup>27</sup> However, these common methods proved to be ineffective for small data sets, and we used the genetic algorithm (GA) introduced in ref. 28–30.

In this approach (Fig. 1), multiple subsets  $\chi$  of  $m < N_x$  variables (“chromosomes”, with each explanatory variable representing a “gene”) are generated at random, and the MMLR with the optional ridge regularization is performed for each set. The root mean square deviation (rms) for each “chromosome” serves as the measure of its “fitness”. At each iteration, 50 “chromosomes” are sorted by rms, and the 10 most fit “chromosomes” are selected at random according to the precedence of their fitness. Two “mutations” (random substitutions of the “genes”) are introduced per “chromosome”, and random fragments of these chromosomes are interchanged (Fig. 1). These variations are introduced at each generation as the population evolves, searching for the global minimum on the fitness landscape. Typically, 100–2000 iterations were sufficient to locate the optimal subset.

### Inadequacy of generic descriptors for justifying stability trends

The success of ML modeling is predicated on capturing variability in the data set using a small number of structural descriptors. A common approach in cheminformatics is to use generic descriptors that can be obtained in the hundreds and thousands with the reasonable expectation that the property of interest correlates with some of these descriptors. Unfortunately, in many instances such descriptors lack intuitive content. Their main function is to allow interpolation and extrapolation of data rather than providing insight into the chemical and structural origins of the variability captured by these molecular descriptors.

When these generic descriptors were used to fit well-defined redoxmer properties closely associated with their electronic structure, such as redox potential ( $E_{1/2}$ ) or hyperfine coupling constants (spin densities) for radical cations, they performed quite well (Section S6). For the  $E_{1/2}$  of dialkoxyarenes, even for an extended set as in ref. 16 (which includes the molecules examined in this paper, see Fig. S6.1), it was always possible to find two descriptors modeling the data with a coefficient of determination  $r^2 > 0.97$ . These trends are illustrated in Fig. S6.1, where  $E_{1/2}$  is correlated with a linear combination of Mollog  $P$  (an estimator of lipophilicity)<sup>31</sup> and ATSC7c (lag-7 centered



Moreau–Broto autocorrelator weighted by Gasteiger charge)<sup>32</sup> with  $r^2 \approx 0.98$ . The generic descriptors were also suitable for rationalizing more complex properties directly related to molecular structure. For example, the MMLR model for the proton hyperfine coupling constants in the radical cations that quantify the unpaired electron spread across the molecule (see Table S5.1) performed nearly as well as the model for redox potentials. For dialkoxyarene molecules with unbranched R-groups (16 molecules, 4 hyperfine coupling constants for each), a set of 64 measurements can be fit using just six generic descriptors ( $r^2 > 0.95$ , see Fig. S6.2). This number of descriptors is justified given the number of measurements. The accuracy of this model was better than the accuracy of our density functional theory (DFT) estimates. Thus, certain properties of radical cations readily lend themselves to ML analysis. However, many of the descriptors not only are not intuitive but also lack an obvious connection to the properties of charged redoxmers.

The approach was less successful in capturing trends for the redoxmer stability metrics in Table 1 despite the large number of generic descriptors available for such analyses (>1200 were used). Fitting each metric with  $r^2 > 0.9$  required 5–6 generic descriptors for only 11 measurements (see, for example, Fig. S6.3). We sought to ameliorate this deficiency by using a logistic regression (Section S6) to separate charged dialkoxyarenes into short- and long-lived by their half-life times  $t_{1/2}$  (setting 30 h as the stability threshold). In this way, a quadratic 2-descriptor model was found with a pseudo- $r^2$  of 0.75. Thus, even such a low-fidelity classifier proved challenging to construct. We hypothesized that more chemically intuitive and relevant descriptors representing the reactivity of radical ions are required.

### Tailored descriptors for reactive radical ions

Our first step was to optimize the geometry of the neutral and charged molecules using DFT and compare (i) the geometries of neutral molecules with the crystallographic parameters (Table S2.2) and (ii) the computed and experimental hyperfine coupling constants for radical cations (Table S5.1). As seen from Table S2.2, there is good matching of the bond lengths and bond angles for all molecules examined, while the torsion angles for the alkoxy group in the crystal can deviate significantly from the ones computed in solution due to packing of the side groups in these crystals. For the radical cations, the  $C_\alpha$ –O bond is always in the plane of the arene ring away from the *tert*-butyl groups to minimize steric hindrance. Fig. S6.4a shows a linear correlation of the computed and experimental  $E_{1/2}$ . As seen from this plot, the DFT model provides closely correlated estimates ( $r^2 = 0.88$ ) for the redox potentials. The same applies to the computed hyperfine coupling constants shown in Fig. S6.4b. Thus, our DFT calculations provided reliable estimates for geometries, energetics, and spin delocalization in the radical cations, and we based our descriptors on these calculations.

As stipulated above, the known fragmentation reactions of a dialkoxyarene radical cation involve the quinoid core. As the reactivity is defined by distribution of charge and unpaired

electron density through this fragment, Mulliken partial atomic charges and spin densities on carbon atoms, oxygens, and  $\alpha$ -hydrogens obtained in DFT calculations were chosen as 1D descriptors (the entire list of these descriptors is given in the Excel spreadsheet in the SI). To these quantities, the average charge and spin density and their variance over the arene rings and the quinoid fragment were added. These variances served as heuristic metrics for charge and spin delocalization. Other descriptors included all-molecule H and C counts,  $H_\alpha$  atom counts, bond lengths and bond angles for H, C, and O atoms in the quinoid fragment, and dihedral angles involving the  $C_\alpha$ –O and C(arene)–C(*tert*-butyl) bonds. The minimal-approach distances between the aromatic protons and  $C_\alpha$  carbons were also included to quantify steric strain.

Another insight from previously recognized trends was a possible role of symmetry in the radical cation stabilization.<sup>21,33</sup> Several descriptors were computed to quantify this symmetry, where the electric dipole was the most straightforward metric and other parameters are described in Section S6. In total, 100 descriptors were used as “genes” in the GA-optimized MMLR models (Fig. 1). As observed above, the redox potential  $E_{1/2}$  is a relatively simple property that can be fit with high fidelity using only two descriptors. The computed  $E_{1/2}$  (which was one of the descriptors) already correlates well with the experimental estimates, as seen from Fig. S6.4a. The goodness-of-fit can be increased to  $r^2 = 0.96$  by adding a second descriptor ( $C_6$ – $H_{Ar}$  bond length); still greater accuracy ( $r^2 = 0.98$ ) is obtained by using the spin density in the oxygen and the C(arene)–O bond length. These examples suggest that there is sufficient variability in our descriptor set to account for variations in the redox potentials.

### Modeling with the tailored descriptors

Since there is no *a priori* reason to expect that the two stability metrics in Table 1 are controlled by the same structural parameters, we fit them separately. These fits and the values for all descriptors are given in the Excel spreadsheet in the SI, which also contains cross-validation of the fit (see Section 6). Fig. 5a shows the three-descriptor fit of  $N_{90}$ . This model has  $r^2 = 0.967$  (Table 2). One of the model descriptors is the C(arene)–O bond length (with a negative coefficient) that was also a descriptor for  $E_{1/2}$  (with a positive coefficient, see Table 2). This result accords with the intuitive expectation that a more energetic radical cation will be involved in more types of decay reactions. However, this is only one such factor: two other descriptors are the charges in the  $C_\alpha$  and *tert*-butyl carbons. This is remarkable as the competition between the *O*-dealkylation (*i.e.*, the elimination of the carbocation from the alkoxy group, Fig. 4) and deprotection is strongly implied by the product analyses: the former reaction causes decomposition of the radical cation that does not result in polymerization, whereas the second reaction leads to macromolecular products (Fig. 5c).

As suggested by these analyses, the cycling stability for short-lived radical cations is mainly determined by the efficiency of the *O*-dealkylation, whereas for long-lived radical cations, the



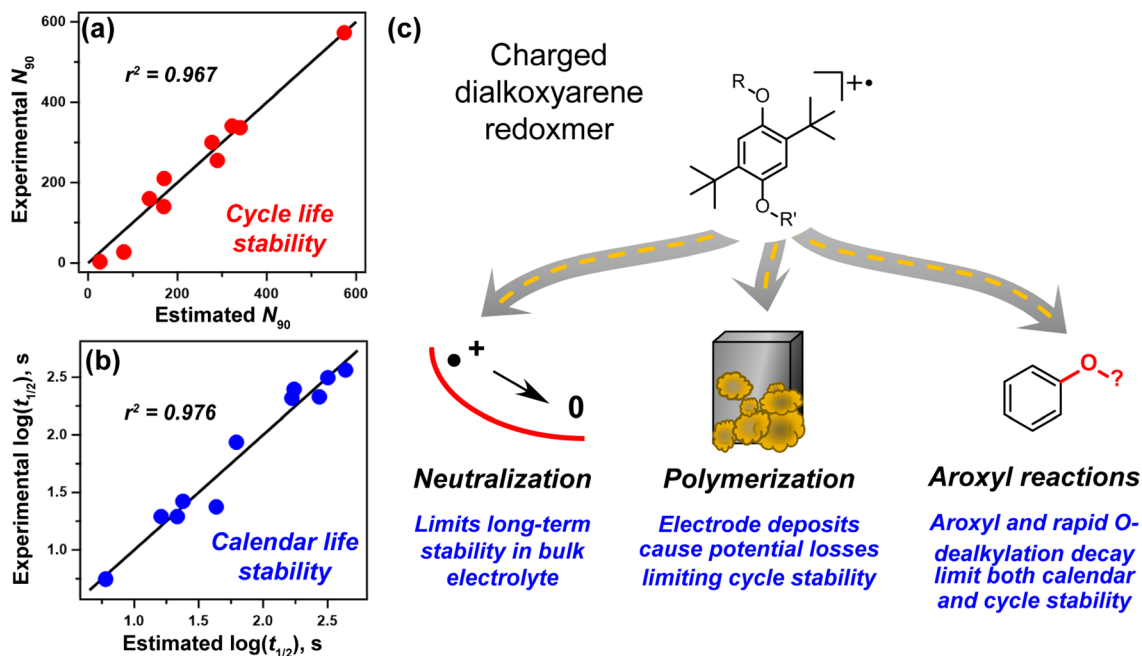


Fig. 5 Correlations of the experimental and estimated (a) cycling life ( $N_{90}$ ) and (b) calendar life ( $\log t_{1/2}$ ) stabilities for dialkoxyarenes in Table 1. The coefficients of determination are shown in the plots. (c) A dialkoxyarene radical cation at the crossroads. Partial charges in *tert*-butyl and  $\alpha$ -carbons are good predictors for elimination of the corresponding carbocations. While rapid *O*-dealkylation limits both of the stability metrics, if this reaction is slow, the cyclability is limited mainly by deprotection of the radical cation leading to the formation of macromolecules, causing the eventual loss of potential control, while the calendar stability is limited by slow neutralization in the bulk.

Table 2 Best fit linear coefficients for reduced descriptors<sup>a</sup>

Descriptors	Fit variable & coeff.	Descriptor type
<b>Redox potential</b>	$E_{1/2}$	
Spin density, O <sub>16</sub>	-0.92	Spin density in the oxygen atom
$r(C_3-O_{10})$	+0.78	C(arene)-O bond length
<b>Cycling life stability</b>	$N_{90}$	
Charge, C <sub>12</sub>	-0.73	Mulliken charge on <i>tert</i> -butyl carbon
Charge, C <sub>17</sub>	-0.84	Mulliken charge on $\alpha$ -carbon
$r(C_3-O_{10})$	-1.00	C(arene)-O bond length
<b>Calendar life stability</b>	$\log(t_{1/2})$	
$I_2$	+0.83	Median semiaxis of the mass weighted gyration ellipsoid
Charge, C <sub>11</sub>	+0.73	Mulliken charge on $\alpha$ -carbon
Spin <sub>ring</sub> _stdev	+1.23	Standard deviation of spin density in the arene ring

<sup>a</sup> Reduced values are shifted by their mean and normalized by their standard deviation over the data set. The atoms are numbered as shown in the upper right panel of Fig. 1. A positive coefficient means positive correlation, while a negative coefficient means negative correlation.

cycling stability is mainly determined by the efficiency of deprotection. As both reactions are carbocation elimination, the charges in the leaving carbon are good quantifiers for the corresponding reaction barriers. The cycling life of dialkoxyarene molecules is predicated on (i) the energy stored in the radical cation and (ii) its proclivity towards carbocation elimination.

For the calendar life stability,  $\log(t_{1/2})$  was chosen as a variable. The best three-descriptor fit we found (with  $r^2 = 0.98$ ) is shown in Fig. 5b. One of the selected descriptors is the C $_{\alpha}$  carbon charge (which was also implicated in our fit for  $N_{90}$ ): the

inclusion of this descriptor can be rationalized through its correlation with the *O*-dealkylation rate (for short-lived radical cations) as was the case with  $N_{90}$ . Another descriptor is the standard deviation for the spin densities in the arene ring, which is an *ad hoc* metric of spin delocalization in the ring.

The most surprising descriptor is the third one ( $I_2$ ), which corresponds to the intermediate principal moment of inertia of the molecule. Its principal axis lies in the arene plane and points perpendicular to the main axis of the molecule. It is unlikely that this descriptor reflects steric hindrance, as steric



effects primarily influence  $I_3$ , whose principal axis is perpendicular to the arene plane. As charge neutralization is the main reaction of long-lived radical cations in the bulk, solvent access to a reactive site plays an important role. For every radical cation that is converted to a neutral molecule, there should be an oxidation. It is unlikely that the oxidized species is a salt ion in the electrolyte. This suggests that the acetonitrile solvent itself becomes oxidized during the reaction. In this respect, our observation of product 6 can be relevant mechanistically as it suggests that the neutralization could be initiated by the formation of a C–O bond between the solvent and the charged oxygen in the alkoxy group of the molecule (similar to the initial step of the Pinner reaction).<sup>34</sup> We hypothesize that the descriptor  $I_2$  is a predictor of a reaction barrier for this reaction, controlling solvent access to the alkoxy oxygens.

Recently, this speculation has been boosted by a ML-driven high throughput study from Argonne's RAPID-200 autonomous discovery laboratory,<sup>35</sup> in which the stability of *N*-substituted phenothiazinium radical cations was studied. It was shown that the slowest reactions of this radical cation involved homolysis of the solvents (including acetonitrile) with the formation of sulfonium carbocations, some of which eventually decayed back to the parent compound. It seems likely that an analogous reaction can occur with the formation of oxonium ions that also partially revert to the parent molecule. It is this multistep reaction that has been recognized as bulk neutralization of the radical cations.

## Conclusions

Inferring causation from apparent randomness has been a fundamental challenge throughout human history. The specific case analyzed in this study concerns redoxmers for grid-scale energy storage devices, namely redox flow batteries. However, it illustrates the general concerns with engineering novel electrolytes: reactive intermediates frequently have multiple reaction pathways, some of which could be poorly understood or unknown. In our case, this multiplicity complicates performance assessment of redoxmer stability in realistic settings: the lifetime data can look random even when measurements are conducted in a uniform way and include homologous species.<sup>16</sup>

Combining product analyses with chemically intuitive descriptors of radical cation reactivity and ML methodology, we obtained a model strongly suggesting that the stability of charged dialkoxyarene redoxmers is controlled by the interplay of three competing reactions: charge neutralization, *O*-dealkylation, and arene ring deprotection (Fig. 5c). Of these three reactions, the deprotection (which is a low-yield reaction) has an especially strong impact on electrochemical cyclability as it enables polymerization of deprotected molecules that causes fouling of the electrodes and the loss of potential control. This detrimental reaction, however, has little effect on the calendar life of charged dialkoxyarenes in the electrolyte bulk. Conversely, the charge neutralization has little effect on the cycle life, as it recovers the active material, while having a major

effect on the calendar life. Expectedly, the *O*-dealkylation affects both stability metrics when it becomes sufficiently fast.

The “erratic” data shown in Fig. 2 and Table 1 reflect competition between these three reactions and their unequal effect on the two metrics of electrochemical performance. In this way, the seeming randomness has been reduced to the interplay of simpler causes, conforming to the Darwinian view on chance.

The broader impact of this study is that the chemical processes we discuss—fragmentation, polymerization, and neutralization—are very common for radical ions, so competing reactions of this kind are the norm. Different reaction pathways respond to molecular structure in distinct ways, making it generally difficult to discern clear patterns in the data. This variability and the challenge of interpreting seemingly random results have been a shared experience in the field for many years. The use of laboratory automation and AI/ML-directed methods may make disentangling these reactions increasingly feasible in the future, but this ideal has not yet been realized.

## Author contributions

The manuscript was written through contributions from all authors. All authors have given approval to the final version of the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

CCDC 645127 and 1938226–1938230 contain the supplementary crystallographic data for this paper.<sup>36a–f</sup>

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information: Sections S1 to S9 contain synthetic procedures, electrochemical, crystallographic, and EPR protocols, NMR spectra, crystallographic and EPR data, EPR kinetics, electrolysis and ML plots, and high-performance liquid chromatography (HPLC) spectra. An Excel spreadsheet is provided containing descriptor names, meanings, and values and ML fits. See DOI: <https://doi.org/10.1039/d6ta00771f>.

## Acknowledgements

We dedicate this study to the memory of Prof. Susan Odom (1980–2021) of the University of Kentucky. IAS thanks Richard Wilson for his help with the crystallographic structure determination and Anna Pluzhnikov for help with statistical analyses and logistic regression. LAR thanks Trevor Dzwiniel and the Manufacturing Engineering Research Facility (MERF) for assistance with liquid chromatography characterization. This work was supported as part of the Joint Center for Energy Storage Research (JCESR), an Energy Innovation Hub funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences. The submitted manuscript has been created by



UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

## Notes and references

- W. Sharmoukh, *RSC Adv.*, 2025, **15**, 10106–10143.
- U. A. Dodo, B. A. Salami, F. M. Bashir, H. Y. Hamdoun, I. S. Rashed Alsadun, Y. A. Dodo, A. G. Usman and S. I. Abba, *Energy*, 2024, **304**, 132140.
- D. Chang, *J. Comput. Methods Sci. Eng.*, 2025, **25**, 1782–1792.
- M. Shoaib, P. Vallayil, N. Jaiswal, P. Iyapazham Vaigunda Suba, S. Sankararaman, K. Ramanujam and V. Thangadurai, *Adv. Energy Mater.*, 2024, **14**, 2400721.
- R. Walser-Kuntz, Y. Yan, M. Sigman and M. S. Sanford, *Acc. Chem. Res.*, 2023, **56**, 1239–1250.
- N. H. Attanayake, J. A. Kowalski, K. V. Greco, M. D. Casselman, J. D. Milshtein, S. J. Chapman, S. R. Parkin, F. R. Brushett and S. A. Odom, *Chem. Mater.*, 2019, **31**, 4353–4363.
- J. Zhang, Z. Yang, I. A. Shkrob, R. S. Assary, S. o. Tung, B. Silcox, W. Duan, J. Zhang, C. C. Su, B. Hu, B. Pan, C. Liao, Z. Zhang, W. Wang, L. A. Curtiss, L. T. Thompson, X. Wei and L. Zhang, *Adv. Energy Mater.*, 2017, **7**, 1701272.
- K. Roy, S. Kar and R. N. Das, *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*, Springer, New York, 2015.
- C. S. Sevov, D. P. Hickey, M. E. Cook, S. G. Robinson, S. Barnett, S. D. Minter, M. S. Sigman and M. S. Sanford, *J. Am. Chem. Soc.*, 2017, **139**, 2924–2927.
- A. S. Perera, Y. Wang, S. Ruhunage, L. A. Robertson, T. M. Suduwella, S. R. Parkin, R. H. Ewoldt, C. Risko and A. P. Kaur, *Energy Fuels*, 2025, **39**, 10649–10658.
- S. G. Robinson, Y. Yan, K. H. Hendriks, M. S. Sanford and M. S. Sigman, *J. Am. Chem. Soc.*, 2019, **141**, 10171–10176.
- Handbook of Chemoinformatics: from Data to Knowledge in 4 Volumes*, ed. J. Gasteiger, Wiley-VCH Verlag GmbH & Co., Weinheim, 2003.
- B. Kalita, R. Zubatyuk, D. M. Anstine, M. Bergeler, V. Settels, C. Stork, S. Spicher and O. Isayev, *Angew. Chem., Int. Ed.*, 2026, **65**, e16763.
- M. R. Dooley and S. Vyas, *Phys. Chem. Chem. Phys.*, 2025, **27**, 6867–6874.
- J. Huang, L. Cheng, R. S. Assary, P. Wang, Z. Xue, A. K. Burrell, L. A. Curtiss and L. Zhang, *Adv. Energy Mater.*, 2015, **5**, 1401782.
- B. Silcox, J. Zhang, I. A. Shkrob, L. Thompson and L. Zhang, *J. Phys. Chem. C*, 2019, **123**, 16516–16524.
- J. Huang, B. Pan, W. Duan, X. Wei, R. S. Assary, L. Su, F. R. Brushett, L. Cheng, C. Liao, M. S. Ferrandon, W. Wang, Z. Zhang, A. K. Burrell, L. A. Curtiss, I. A. Shkrob, J. S. Moore and L. Zhang, *Sci. Rep.*, 2016, **6**, 32102.
- J. Holcman and K. Sehested, *J. Phys. Chem.*, 1976, **80**, 1642–1644.
- P. O'Neill, D. Schulte-Frohlinde and S. Steenken, *Faraday Discuss. Chem. Soc.*, 1977, **63**, 141–148.
- P. O'Neill, S. Steenken and D. Schulte-Frohlinde, *J. Phys. Chem.*, 1975, **79**, 2773–2779.
- J. Zhang, I. A. Shkrob, R. S. Assary, S. o. Tung, B. Silcox, L. A. Curtiss, L. T. Thompson and L. Zhang, *J. Phys. Chem. C*, 2017, **121**, 23347–23358.
- J. Huang, I. A. Shkrob, P. Wang, L. Cheng, B. Pan, M. He, C. Liao, Z. Zhang, L. A. Curtiss and L. Zhang, *J. Mater. Chem. A*, 2015, **3**, 7332–7337.
- J. Zhang, J. Huang, L. A. Robertson, R. S. Assary, I. A. Shkrob and L. Zhang, *J. Phys. Chem. C*, 2018, **122**, 8116–8127.
- J. J. Ritter and P. P. Minieri, *J. Am. Chem. Soc.*, 1948, **70**, 4045–4048.
- J. J. Ritter and J. Kalish, *J. Am. Chem. Soc.*, 1948, **70**, 4040–4050.
- R. Tibshirani, *J. Roy. Stat. Soc. B Stat. Methodol.*, 1996, **58**, 267–288.
- H. Zou and T. Hastie, *J. Roy. Stat. Soc. B Stat. Methodol.*, 2005, **67**, 301–320.
- D. Broadhurst, R. Goodacre, A. Jones, J. J. Rowland and D. B. Kelp, *Anal. Chim. Acta*, 1997, **348**, 71–86.
- A. Yasri and D. Hartsough, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1218–1227.
- J. K. Wegner and A. Zell, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1077–1084.
- S. A. Wildman and G. M. Crippen, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 868–873.
- G. Moreau and P. Broto, *Nouv. J. Chim.*, 1980, **4**, 359–360.
- W. Weng, J. Huang, I. A. Shkrob, L. Zhang and Z. Zhang, *Adv. Energy Mater.*, 2016, **6**, 1600795.
- R. Roger and D. G. Neilson, *Chem. Rev.*, 1961, **61**, 179–211.
- L. A. Robertson, I. A. Shkrob, R. Lewis, L. Ward, R. Vescovi and B. T. Diroll, *J. Am. Chem. Soc.*, 2025, **147**, 37211–37222.
- (a) CCDC 645127: Experimental Crystal Structure Determination, 2026, DOI: [10.5517/ccdc.csd.ccpn9k5](https://doi.org/10.5517/ccdc.csd.ccpn9k5); (b) CCDC 1938226: Experimental Crystal Structure Determination, 2026, DOI: [10.5517/ccdc.csd.cc231wfh](https://doi.org/10.5517/ccdc.csd.cc231wfh); (c) CCDC 1938227: Experimental Crystal Structure Determination, 2026, DOI: [10.5517/ccdc.csd.cc231wgj](https://doi.org/10.5517/ccdc.csd.cc231wgj); (d) CCDC 1938228: Experimental Crystal Structure Determination, 2026, DOI: [10.5517/ccdc.csd.cc231whk](https://doi.org/10.5517/ccdc.csd.cc231whk); (e) CCDC 1938229: Experimental Crystal Structure Determination, 2026, DOI: [10.5517/ccdc.csd.cc231wj](https://doi.org/10.5517/ccdc.csd.cc231wj); (f) CCDC 1938230: Experimental Crystal Structure Determination, 2026, DOI: [10.5517/ccdc.csd.cc231wkm](https://doi.org/10.5517/ccdc.csd.cc231wkm).

