Data-driven discovery of small electroactive molecules for energy storage in aqueous redox flow batteries

Qi Zhang[1,2,3,], Abhishek Khetan[1,2,5,*], Elif Sorkun[1,2], Fang Niu[4], Andrea Loss[4], Ilaria Pucher[4], Süleyman Er[1,2,‡]*

[1] DIFFER – Dutch Institute for Fundamental Energy Research, De Zaale 20, 5612 AJ Eindhoven, The Netherlands

[2] CCER – Center for Computational Energy Research, De Zaale 20, 5612 AJ Eindhoven, The Netherlands

[3] Department of Applied Physics, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands.

[4] Green Energy Storage, Via Sommarive 18, 38123 Povo, Italy

[5] Present address: Fuel Science Center, RWTH Aachen University, 52062 Aachen, Germany

[*] These authors contributed equally

[‡] E-mail: s.er@differ.nl

Abstract

Owing to advancements in computing power and improvements in automation of modelling, high-throughput virtual screening (HTVS) has increasingly been used for materials data generation. Here, we applied a HTVS-guided experimental study for the large-scale exploration of quinone-like anolytes for aqueous redox flow batteries (ARFBs). This includes the design of a focused virtual chemical library inspired by small colorant molecules, quantum chemical prediction of redox properties, machine learning prediction of aqueous solubility, automated search for commercial availability on vendor databases, and electrochemical characterization of the most promising compounds. Screening efforts in a chemical space of 3,257 redox pairs led to 205 predicted candidates with higher solubility and lower redox potential than that of the state-of-the-art anthraquinone-2,7-disulfonic acid (AQDS) anolyte used in ARFBs. Through the electrochemical studies on the commercially available compounds, we identified the molecules that show good performance in an ARFB setup. Among them, indigo trisulfonate [Indigo-3(SO$_3$H)] showed higher solubility, capacity retention, and coulombic efficiency than AQDS and its predecessors. The data-driven material design methodology presented here is flexible and applicable for the future exploration of small compounds for electrochemical energy storage.

1 Introduction

Aqueous redox flow batteries (ARFBs) have emerged as one of the most promising technologies for low-cost and long-term storage of renewable energy as their unique design allows for independent scaling of peak power and energy storage.[1,2] Even though there are, mostly metal-based, ARFBs that are commercially available today, the technology still faces obstacles that impede a widespread deployment.[3–6] ARFBs with organic active materials are progressively being investigated because of their rich structural diversity that relate to tunable properties and raw material abundance that relate to production scale-up.[7,8] A promising family of electroactive materials for ARFBs is the quinone-like cyclic dione compounds,[9–13] which possess high electrochemical and chemical tunability[11,14] as well as rapid redox kinetics.[9–11,15] Quinones are found to occur naturally[16] and are also mass produced for application in the chemical and pharmaceutical industries.[17,18]

The battery relevant properties of quinones, including redox potential, solubility, and stability, need improvements for the future commercial use of these compounds.[8,19,20] To this end, a number of recent studies have employed high-throughput virtual screening (HTVS) methods to partially explore the vast chemical space of quinones, which were derived mostly from cyclohexadiene and its fused derivatives, such as benzoquinones (1-ring), naphthoquinones (2-rings), anthraquinones (3-rings),[21–25] or from naturally occurring quinones.[26] The HTVS approach generally begins with creation of a comprehensive virtual library of candidate molecules. This is typically performed by decorating ring positions on quinone-like backbone structures with various functional groups[21–24] and/or by substituting the carbon atoms in the rings with heterocyclic atoms,[25] with an aim to tune the relevant properties of the core molecule. These efforts have not only led to the discovery of new molecules that can potentially perform better than the best experimentally validated quinones,[21,23] but they also have elucidated molecular design principles that are crucial for the further optimizations of the desired properties of the electroactive molecules.[23] It must be noted that reversible electrochemical behavior can also be found in quinone-like molecules in which the carbonyl pairs are attached on cyclopentene

instead of cyclohexadiene, as often found in dyes and cosmetics. For instance, a recent study demonstrated that the compound indigo and its derivatives undergo two distinct reversible redox reactions between three oxidation states.[27] In this regard, the chemical space of cyclopentene with carbonyl pairs has not yet been fully explored.

Herein, we performed a data-driven study on a library of quinone-like molecules, whose backbone structures were derived from colorant compounds that had been approved by the U.S. Food and Drug Administration (FDA)[28] for commercial use. This choice was motivated not only by the excellent battery performance[27] but also by the present availability of the compounds. Inspired by the recurrent motifs of the FDA-approved molecules, a comprehensive virtual chemical space of backbone structures, which primarily includes carbonyl pairs on cyclopentene molecules and their combinatoric variations generated by heterocyclic substitutions, was developed and used for HTVS.
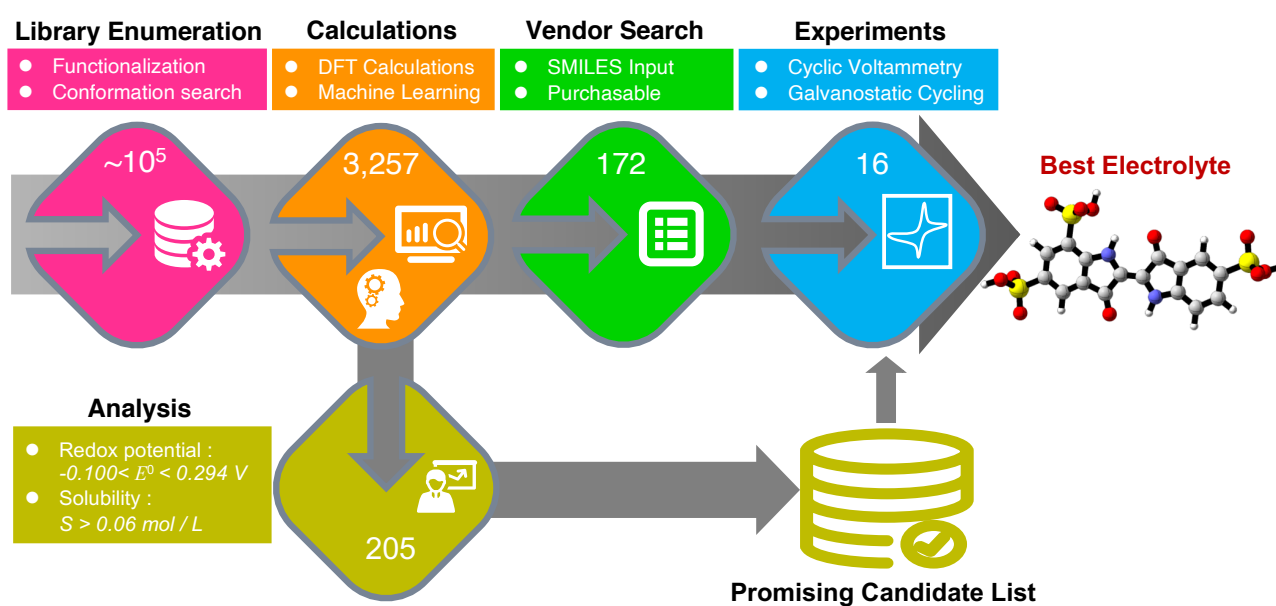

Fig. 1 An overview of the data-driven material discovery workflow that has been employed in the current work for the identification of useful small electroactive molecules for ARFBs.
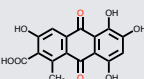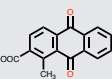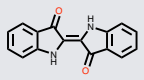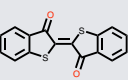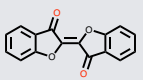
Fig. 1 shows a schematic overview of the data-driven compound discovery workflow that has been used in the current study. The redox potentials of the candidate compounds were predicted by using density functional theory (DFT) powered computational schemes that were introduced in our recent studies.[29,30] Instead of using the aqueous solvation energy as a proxy, the aqueous solubility data of the compounds was created by using a supervised machine learning (ML) model[31], which was trained on a curated aqueous solubility database.[20] All commercially available compounds were identified by using an in-house developed code that performs an automated vendor search on the ZINC database.[32] Some of the molecules thus obtained were further characterized experimentally by using cyclic voltammetry (CV) and galvanostatic cycling methods. Thus, through a combined HTVS-experiment approach, several previously untested dye-based molecules were explored for their use in ARFBs. The molecules that showed the best performance, when compared to their pure backbone structures as well as to the benchmark energy storage compounds from literature, are discussed further.

2 Methods

2.1 Virtual chemical library design

The molecules in the virtual library were derived by functionalizing the backbone structures from a starting set of six FDA-approved compounds that are shown in Fig. 2a (#1 to #6 with grey background). Compound #1 is *N*-methylphthalimide, which is a building block molecule that finds use in plastics and dyes.[33] Compounds #2 and #3 are derived from the colorant carmine[34] molecule and they both are essentially functionalized anthraquinones, thus allowing for the study of the effect of more than one type of functional group on the same backbone structure. Compound #4 is indigo[27] molecule, which is followed by its heterocyclic variants of compounds #5[35] and #6[36].

(a)

1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	

(b)

Library Generation

R = – SO$_3$H
– COOH
– F
– NH$_2$
– OH

6
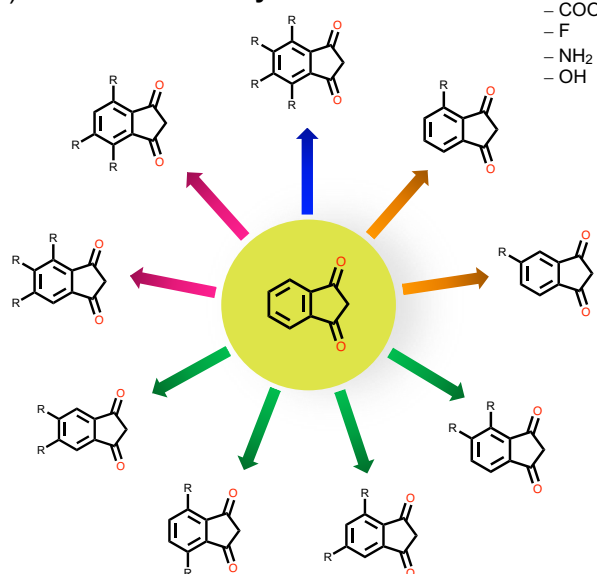
Fig. 2 (a) 29 different backbone molecule structures that were used for library generation and the proceeding HTVS studies, **(b)** the enumeration process of chemical library generation as shown for one of the backbone molecules.

In addition to the colorant molecules, a complementary set of 23 small molecules (#7 to #29) were used as backbone structures. These include carbonyl groups on cyclopentene (#7), cyclohexene (#15) and the 5-membered ring in indane (#22 and #26). The remaining backbone structures were generated by heterocyclic carbon atom substitutions of these structures with N, O, and S atoms. Five functional (R)-groups ($-SO_3H$, $-COOH$, $-NH_2$, $-OH$, and $-F$) were in turn used for the functionalization of backbone structures. An exhaustive enumeration of the functionalized derivatives of the reactant backbone structures was performed by using the Custom R-group Enumeration tool of the Schrödinger Materials Science Suite (SMSS).[37] Fig. 2b shows a simplified representation of the combinatorial virtual library generation process for molecule #22. The product molecules of the redox reactions were next generated according to a two-electron two-proton reaction mechanism.[9] For consistency between the reactants and their respective products, the molecule pairs were matched by using the Reaction-based Enumeration tool in SMSS. The generated SMILES representations of both the reactant and product molecules were converted to three dimensional (3D) geometries.[37] Subsequently, for each candidate molecule in the virtual library, the lowest energy conformer was identified by using the MacroModel Conformation Search tool as implemented in SMSS and the OPLS3e[38,39] force field. Further details of the molecule database creation process is provided in our recent work.[40] The newly built virtual molecule library contains a total of 3,257 O/OH redox pairs for HTVS.

2.2 Physics-based and data-driven modelling

DFT calculations were performed on the lowest energy 3D conformers using the Jaguar program[41] as implemented in SMSS. The PBE exchange-correlation functional[42,43] and LACVP**$^{++}$ basis set with

polarization and diffuse functions[44] were employed to perform gas phase geometry optimizations. A medium electronic grid density was used during structural optimizations. The convergence criteria for energy and density matrix were kept at their default values of 5.0×10^{-5} and 5.0×10^{-6} Hartree, respectively. The default direct inversion scheme was employed in combination with Jaguar's mixed pseudospectral approximation. Lastly, the DFT optimized geometries of the molecules were used as inputs for single point energy calculations, which were performed with fine electronic grid densities and accurate self-consistent field cutoffs. The single point energies were calculated in an implicit aqueous medium based on the Poisson-Boltzmann Solvation Model (PBF).[41,45] Further information on the physics-based calculation methodology can be found in our recent works on the evaluation of various computational chemistry methods for the prediction of redox potentials of small organic-based energy storage compounds.[29,30] Accordingly, the linear regression equations used for predicting the experimental redox potentials of the reduction of C=O groups on quinone-like[29] molecules was $E^{\mathrm{o}} = -0.409[\Delta E_{\mathrm{rxn}}^{\mathrm{DFT}}] - 0.193$ (pH = 0), and of the reduction of in-ring C=N groups on alloxazine-like[30] molecules was $E^{\mathrm{o}} = -0.377[\Delta E_{\mathrm{rxn}}^{\mathrm{DFT}}] - 0.616$ (pH = 7), where on both cases $\Delta E_{\mathrm{rxn}}^{\mathrm{DFT}}$ represents the DFT-calculated reaction energies.

The aqueous solubility data of the reactant molecules was obtained by using AqSolPred v1.0.[31], which is a state-of-the-art consensus ML model that had been trained on the largest open-source measured solubility dataset of chemical compounds.[20] In this context, the approach employed here is in stark contrast to other HTVS efforts that make use of the computed aqueous solvation free energy, $\Delta G_{\mathrm{solv}}^{\mathrm{o}}$, as an approximate descriptor for the solubility of compounds.

2.3 Screening for commercially available molecules

The ZINC database contains approximately 120 million purchasable compounds that have been aggregated from several commercial vendors.[46] Therefore, to screen the readily available molecules from our newly built virtual library, an exact SMILES search was performed in this colossal database.[46]

To automate the search, a client-side app was developed that communicates with the ZINC servers via the ZINC15 application programming interface (API). Our search app runs through sequential and multi-step data exchange steps, as shown in Supplementary Information Fig. S1. Both the ZINC ID that corresponds to the searched SMILES representation of the molecule and the source of the molecule's ZINC page were recorded. Next, the most useful information for our purpose, including a list of vendors and stock status, was extracted and tabulated.
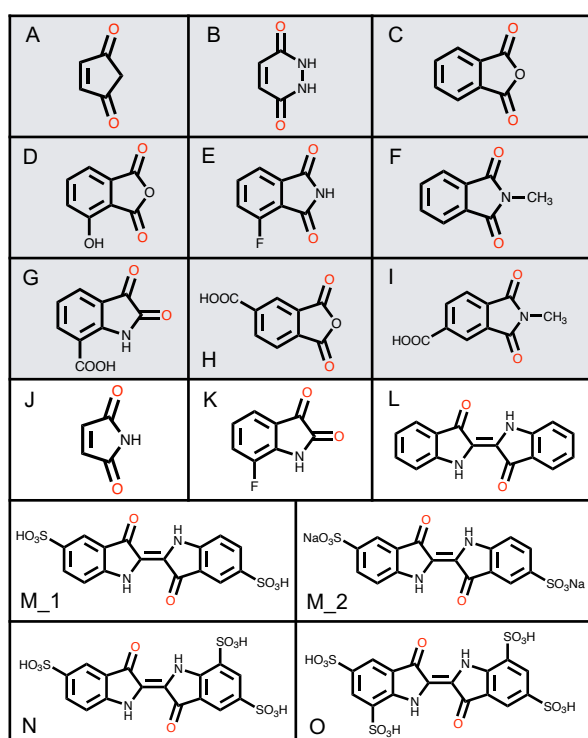
Fig. 3 Commercially available molecules as obtained through a search performed on the ZINC database. Molecules A − I (grey background) were not subjected to further experimentation because of their low aqueous solubility or stability issues.

The search yielded a total of 172 unique molecules from the ZINC database and 16 of them were directly purchasable from different vendors (Fig. 3). All the readily available compounds were acquired for experimental studies. Out of the 16 molecules, nine molecules (Fig. 3: A − I) had either very low solubility or water instability. Therefore, only the remaining seven molecules (J − O) were

subjected to further experimental studies. In addition to these molecules, for comparison, a benchmark, ARFB anolyte molecule anthraquinone-2,7-disulfonic acid (AQDS), was also included in our experiments.

2.4 Experiments

For hydrogen-ion exchange, AQDS disodium salt (Riverside Specialty Chemicals Inc.) and sulphonated forms of indigo sodium/potassium compounds (Sigma-Aldrich) were dissolved in deionized water (DI-H_2O), and then flushed through ion-exchange resin, Amberlite IR 120 H (Sigma-Aldrich), to produce the corresponding acid analogs. The Amberlite resin was firstly activated using 2.5 M H_2SO_4 and then rinsed with DI-H_2O until a neutral pH was obtained. Subsequently, the sodium/potassium salt solution passed through the activated resin three times. Finally, the chemical was collected after drying the passing solution under 110 ˚C for 24 hours.

Solubility tests were carried in a 20 ml vial. A small quantity of molecule powder was weighed in a vial, then DI-H_2O was added in 25 µl/step by using a micropipette. The vials were vigorously shaken then left stationary for an hour to check the undissolved particles at the bottom/walls of the vials. The maximum solubility was reached when the solution presented a homogeneous state without deposits and without precipitation after left standing overnight. It must be noted that some compounds undergo hydrolysis in solution after several days, and such cases are specified in section 3.2. In order to obtain the lowest possible concentrations (e^{-5} M), the solubility tests were carried in a 250 ml bottle.

Cyclic voltammetry (CV) measurements were performed in 3-electrode cell in DI-H_2O, an acidic supporting (buffer) electrolyte, graphite (SGL company; PV15) working electrode (effective area 1 cm^2), platinum wire (> 99.99%) counter electrode and an in-house built Ag/AgCl/3MKCl reference electrode. CVs were conducted at a scan rate of 10 mV/s by using a potentiostat (Gamry, Reference 3000). All CVs were repeated at least three times to ensure reproducibility. The molecule concentration was set at 0.01 M, or at its maximum solubility in cases where solubility was low.

Additionally, in order to make comparisons with literature data, a Britton–Robinson (BR) buffer was used as a supporting electrolyte, which consists of phosphoric acid, boric acid and acetic acid, and titrated to pH = 7 with sodium hydroxide.

Electrochemical cycling experiments were conducted in a custom-made ARFB cell that is schematically shown in Fig. S5. On each side, commercial graphite plates with interdigitated flow channels were in contact with a porous carbon felt electrode (SGL Carbon, 2.5 mm thickness, 4 cm^2 geometric area). An Aquivion® membrane (Solvay, E87-12S, ~175 μm thickness) served as the ion-selective membrane. Electrolytes were pumped through the cell at a fixed flow rate of 275 ml/min using volumetric pumps (Cole Parmer, Masterflex PTFE-Diaphragm Pump Head, 80 to 800 mL/min connected to a Masterflex L/S® Digital Precision Modular Drive with Benchtop Controller, 6 to 600 rpm). Charge/discharge experiments were conducted using a potentiostat (Biologic VMP3 with a 5A booster), under constant galvanostatic conditions. The current was selected in the wake of preliminary voltammetry experiments. All cells were galvanostatically cycled between 1.5 and 0 V and at a constant current of 800 mA. Each half-cycle ended when the cell voltage reached 1.5 V on charge or 0 V on discharge, without a potentiostatic hold.

3　Results and discussion

3.1 Redox potential and solubility map of molecules

For all the molecules in the virtual library, the predicted redox potential (E^o) and aqueous solubility (S) data is shown in Fig. 4.
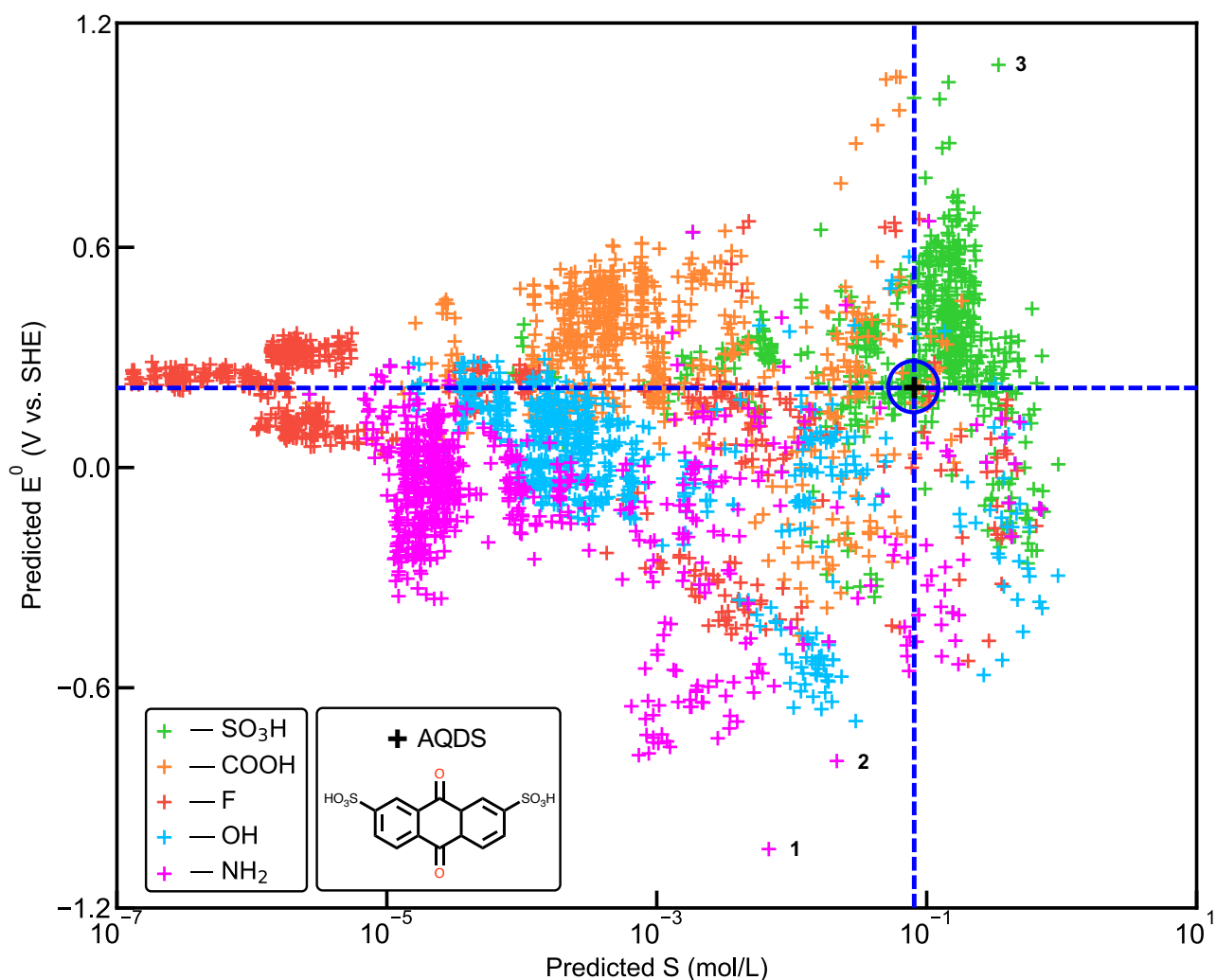
Fig. 4 The distribution of redox potentials and solubility of reactant molecules found in the virtual chemical library. The black-colored cross shows the predicted values for the AQDS molecule, whereas its accompanying blue-colored vertical and horizontal dashed lines are used to guide the reader.

Based on the DFT-calculated redox potential (0.294 V vs. SHE) and the ML-predicted solubility (0.06 mol/L) of AQDS, a first observation is that a total of 2,391 redox couples (73 % of the entire library) have E^o that is lower than AQDS. This indicates that the generated library of small

12

molecules contains several potential candidates for the anode side of ARFBs that have more suitable redox potentials than AQDS. However, it must be noted that molecules with too low redox potential (e.g., E^o < -0.1 V) would likely be unsuitable for application in acidic media due to the undesired H_2 evolution reactions.[21] Secondly, in the group of promising molecules as based on their predicted E^o data, there are 623 candidates (19%) with ML-predicted solubility values that are higher than that of AQDS. Yet again is a large number of candidate molecules with promise for improving the energy density of ARFBs. Based on our screening criteria, with -0.1 < E^o < 0.294 V and S > 0.06 mol/L, 205 quinone-like compounds have been downselected as candidate anolytes for ARFBs. The full list of candidates is provided in Supplementary Information Table S1. Out of the 205 shortlisted molecules, only two molecules that are discussed later in section 3.2, were commercially available. Moreover, assuming that a pair of redox couples with potential difference of more than 0.7 V would make a viable battery,[21] there are as well several candidates that can be employed as catholytes in ARFBs (Fig. 4). However, the wide window of predicted redox potentials, which is approximately from -1.1 to 1.1 V, necessitates a detailed examining of the outliers. For instance, the most obvious outliers are marked with the numbers from 1 to 3 in Fig. 4 and their optimized structures are shown in Fig. S2. Outliers 1 and 2 were derived from backbone structure #21 (Fig. 2) by incorporating the –NH$_2$ functional group, whereas outlier 3 was derived from backbone structure #15 (Fig. 2) by incorporating the –SO$_3$H functional group. It is likely that the geometry optimization of these molecules leads to a sub-optimal minimum because of the strong intramolecular steric interactions. To quantify the effect of functional groups on the overall energy density of the backbone electrolyte material, we propose a metric for energy density of the backbone structure, C_b, that is proportional to the solubility and cell voltage. Under the assumption of an ideal catholyte with redox potential of 1.23 V vs. SHE, the cell voltage can be defined in terms of the calculated anolyte redox potential as $V = (1.23 - E^o)$. Thus,

$$C_b = SV = S(1.23 - E^o) \tag{1}$$

To surpass the C_b of backbone structures, the solubility of the functionalized molecules should be as high as possible and meanwhile the redox potential should be as close as possible to 0 V. The advantage of using C_b as a metric is that it allows trade-offs between redox potential and solubility when estimating the energy density of the candidate materials. Additionally, such trade-offs between redox potential and solubility may be rationalized based on the structural features of compounds as discussed in next section, and used for the material screening purposes.

Fig. 5 The predicted redox potential-solubility maps of quinone-like compounds of different sizes. The functional groups that tune the E^o and S are highlighted using different colors. The dashed black circles show the capacity values of backbone molecules. The dashed black lines represent the cut-off criterion as based on the predicted capacity of parent molecules.

The predicted redox potential and solubility data of the backbone structures and their functionalized derivatives are shown in Figs. 5, 6, and 7. In each sub-figure, the C_b values for the respective backbone structures are provided in addition to a locus that satisfies Eq. 1. Accordingly,

this implies that the molecules positioned on the bottom-right side of this curve are predicted to offer higher energy density when compared to their respective backbone structures. As shown in Fig. 5, consistent trends are observed upon analyzing the effect of functional groups within the FDA-approved group of molecules (#1 to #6 from Fig. 2). While the electron-donating $-NH_2$ and $-OH$ functional groups tend to decrease the redox potential, the electron-withdrawing $-F$, $-COOH$ and $-SO_3H$ groups have an opposite effect. Furthermore, with the exceptions of molecule #1 and group $-F$, nearly all combinations of backbones and functional groups result in increased solubility. Thus, there are several hundred virtual molecules that have higher energy density than C_b. Here, backbone structure #4 represents only one of the redox states of indigo, which otherwise exists in three states: fully oxidized (#4 - 2H), partially reduced (#4), and fully reduced (#4 + 2H). The stepwise reduction processes are assumed to undergo two redox reactions, each involving a two-electron two-proton mechanism, on the C=O and the in-ring C=N bonds at low and high redox potentials, respectively. Therefore, this molecule can in principle be used as a bifunctional electrolyte in a symmetrical ARFB,[27] and it will be discussed later in this work. For indigo, the incorporation of S (#5) or O (#6) heteroatoms to the backbone of the molecule decreases the C_b. Nevertheless, the addition of functional groups on the same molecules with heteroatoms (#5 and #6) has a similar overall effect to that of the indigo compound (#4).
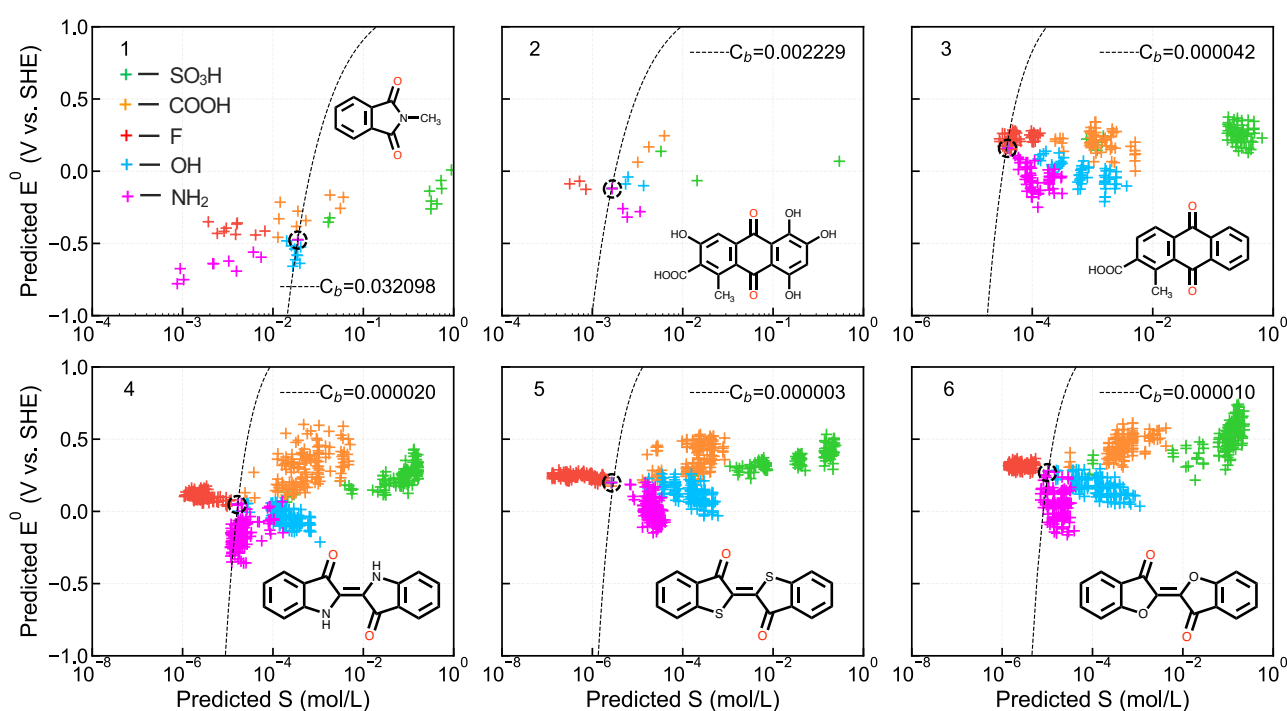
Fig. 6 The predicted redox potential-solubility maps of quinone-like compounds of small size. The functional groups that tune the E^o and S are highlighted using different colors. The dashed black circles show the capacity values of backbone molecules. The dashed black lines represent the cut-off criterion as based on the predicted capacity of parent molecules.

As shown in Fig. 6, when considering carbonyl groups and heteroatom substitutions on the one-ring cyclopentene (#7 to #14) and cyclohexene (#15 to #21) derivatives, the trends concerning the effect of functional groups on redox potentials remain largely unchanged. However, functionalization has mostly a detrimental effect on solubility, even though functionalization with –OH for some of the molecules leads to a higher energy density than their respective backbone structures (Fig. 6). Importantly, the C_b of cyclopentene-derived backbones are higher than the ones with cyclohexene-derived backbones. In addition to this, benzoquinones that are derived from cyclohexenes have been investigated before,[22] whereas the cyclopentene derivatives have not yet been covered well in literature. For one-ringed backbone structures shown in Fig. 6, the effect of heteroatoms on the theoretically achievable energy density is significant, however, it is not systematic. For instance, for backbone structure #7, both N (#8) and O (#10) substitutions lead to higher energy densities, whereas S substitution (#9) has an adverse effect. For backbone structure #11, all heterocyclic substitutions (#12 – #14) lead to lower energy densities. And lastly, for backbone structure #15, all heterocyclic substitutions lead to higher energy densities. For the same compound, the single atom substitutions (#16 – #18) are significantly more effective than the two atom substitutions (#19 – #21) in increasing the energy density.
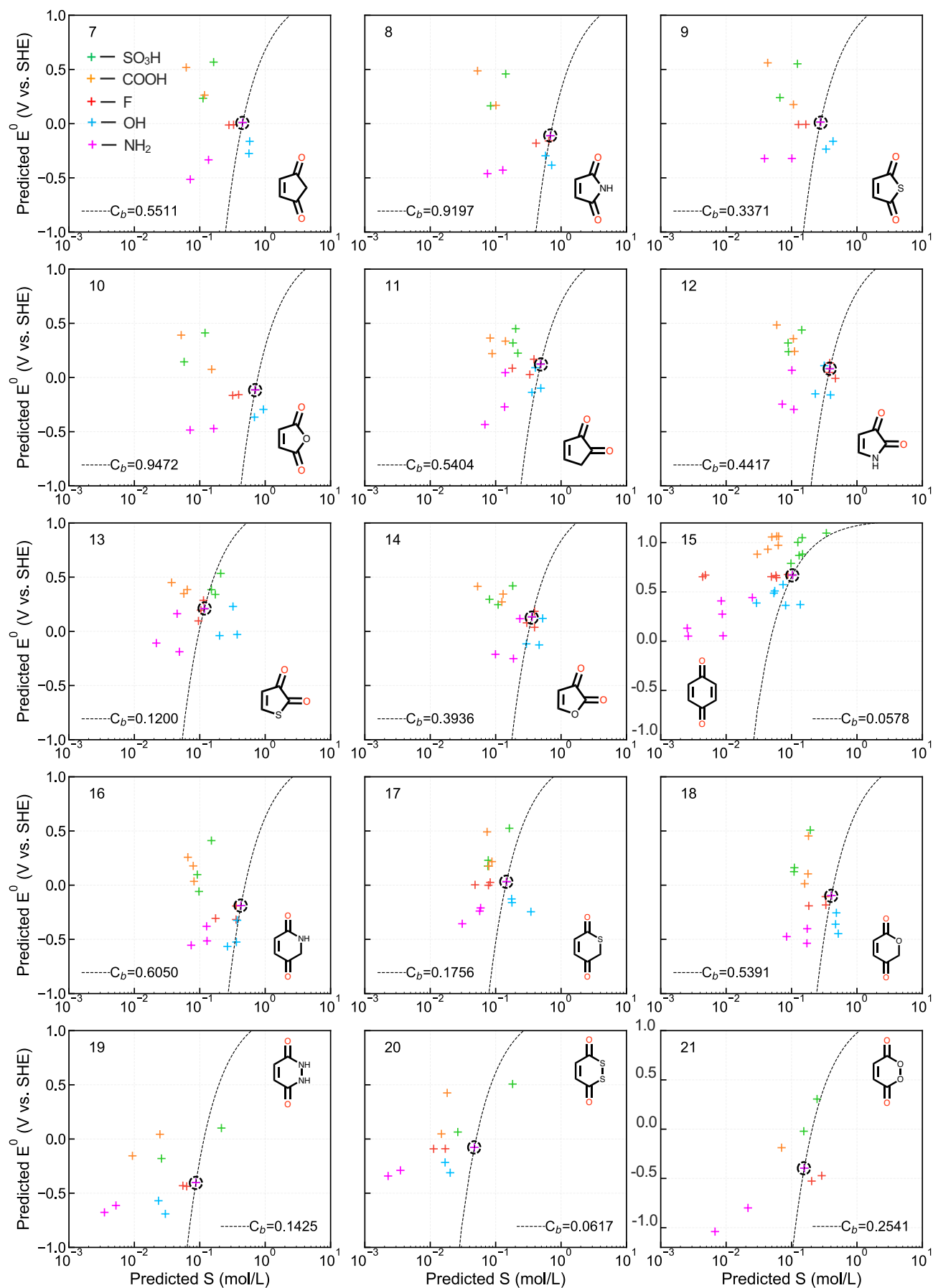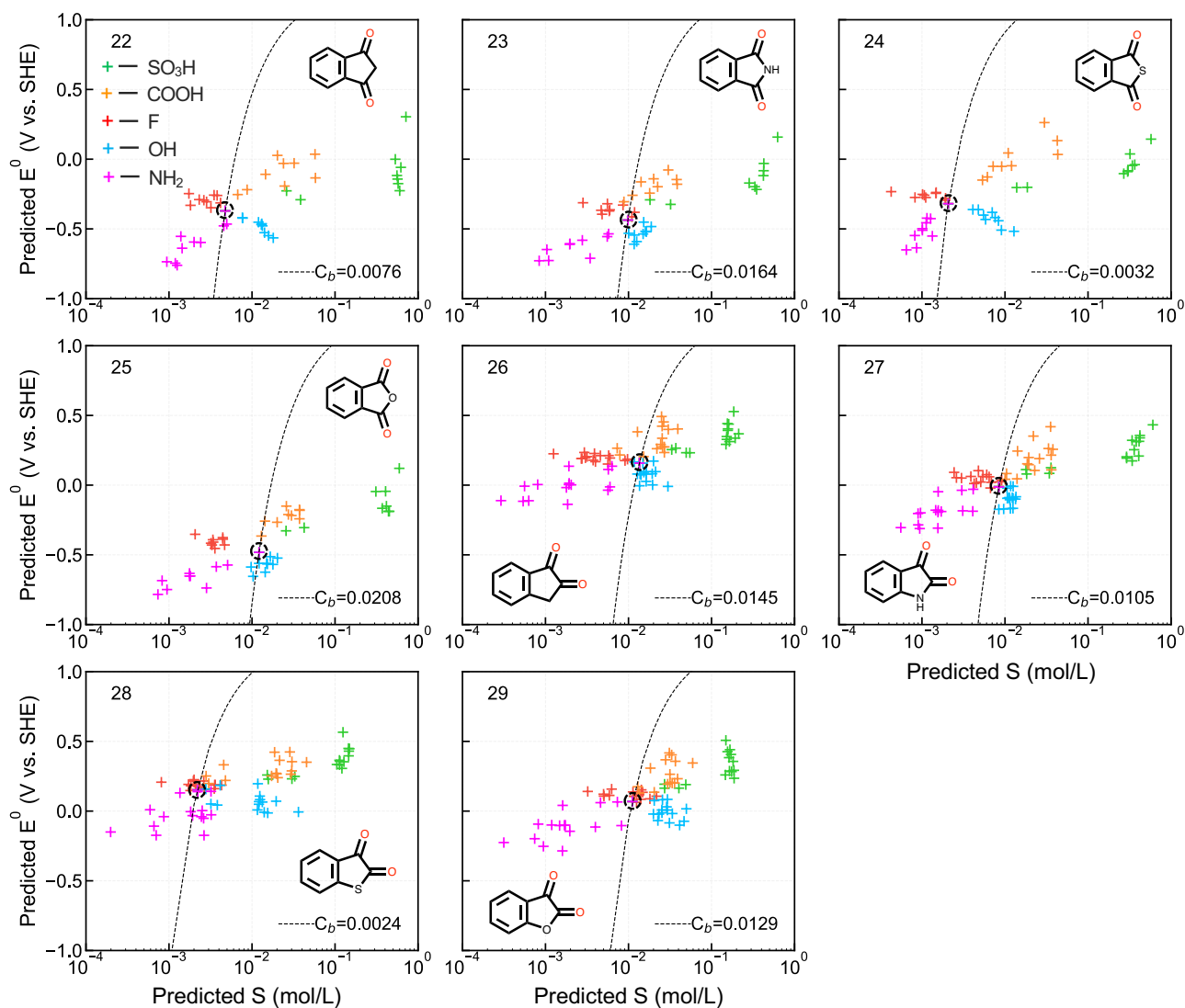
Fig. 7 The predicted redox potential-solubility maps of quinone-like compounds of moderate size. The functional groups that tune the E^o and S are highlighted using different colors. The dashed black circles show the capacity values of backbone molecules. The dashed black lines represent the cut-off criterion as based on the predicted capacity of parent molecules.

As shown in Fig. 7, when considering carbonyl groups and heteroatom substitutions on the two-ringed indane derivatives, the effects are quite different for redox potential and solubility. While the effect of functional groups on redox potential is similar as before (#1 to #21), their effect on solubility is more akin to the FDA list of molecules considered here. The functional groups –SO₃H, –COOH, and –OH increase the solubility, whereas –F and –NH₂ have a detrimental effect. As a result

of improved solubility, there are several indane derivatives that have a higher energy density than that of their respective backbone structures. For the two-ringed backbone structures, the effect of heteroatoms on the energy density is quite similar to their one-ringed cyclopentene counterparts. For backbone structure #22, both N (#23) and O (#25) substitutions result in higher energy densities than C_b, whereas S substitution (#24) lowers energy density. For backbone structure #26, all heterocyclic substitutions (#27 − #29) result in lower energy density than C_b.

Based on these observations, it can be argued that the overall effect of functional groups on energy density of the electroactive molecule is strongly coupled to the chemical motif of the backbone structures. For molecules with two or more rings, the solubilizing effect of –SO₃H is found to be strong enough to overcome the losses in energy density because of increased redox potential, and thus, the reduced cell voltage. For the same group of molecules, functionalizing with –OH is a promising strategy as it both leads to compounds with increased solubility and decreased redox potential, and thus, the increased cell voltage. In practical ARFB applications, molecules with compact size usually would suffer more, as compared to bulky ones, from membrane crossover issues.[47] On the contrary, molecules that are comparable in size to AQDS, or larger, are prone to suffer from low solubility and sluggish kinetics. In this context, quinone-like molecules of moderate size are highly promising for further exploration as electrolyte materials for ARFBs.

3.2 Experimental characterization

Out of the 16 commercially available molecules that were obtained from the vendor search on ZINC, only seven molecules were soluble and stable enough for electrochemical analysis (Fig. 3). Further information on solubility and water (in)stability of the commercially available molecules is provided in Table S2. It should be noted that out of the 16 commercially available molecules, only the molecules A and O are in the 205 shortlisted molecules with better predicted performance than AQDS. Among the seven compounds, J (derived from backbone #8) and K (from backbone #27) did not show any

clear or reproducible peaks in the CV experiments (Fig. S3). The compound J, also known as maleimide, was not stable in solution and formed gel-like deposits in DI-H_2O (pH = 5.4), H_2SO_4 (pH = 0.7), and NaOH (pH = 13) solutions. The amide bonds in maleimide have the tendency to be hydrolyzed under acidic and alkaline aqueous conditions.[48–50] The amount of hydrolysis product, *N*-alkylmaleamic acid, remained stable in solution after weeks. The compound K, also known as 7-Fluorisatin, showed a change in color from colorless to brown within four days of storage in DI-H_2O, which indicates a slow reaction that possibly is related to water instability. For simplicity of discussion the other compounds, henceforth named as Indigo-2(SO_3H) [M_1], Indigo-2(SO_3Na) [M_2], Indigo-3(SO_3H) [N], and Indigo-4(SO_3H) [O], are all –SO_3H functionalized derivatives of Indigo [L] (same as #4). Compounds M_1 and M_2 are nearly identical but sourced from two different salts as were provided by the vendors. M_1 is the protonated form of M_2 after hydrogen ion-exchange.

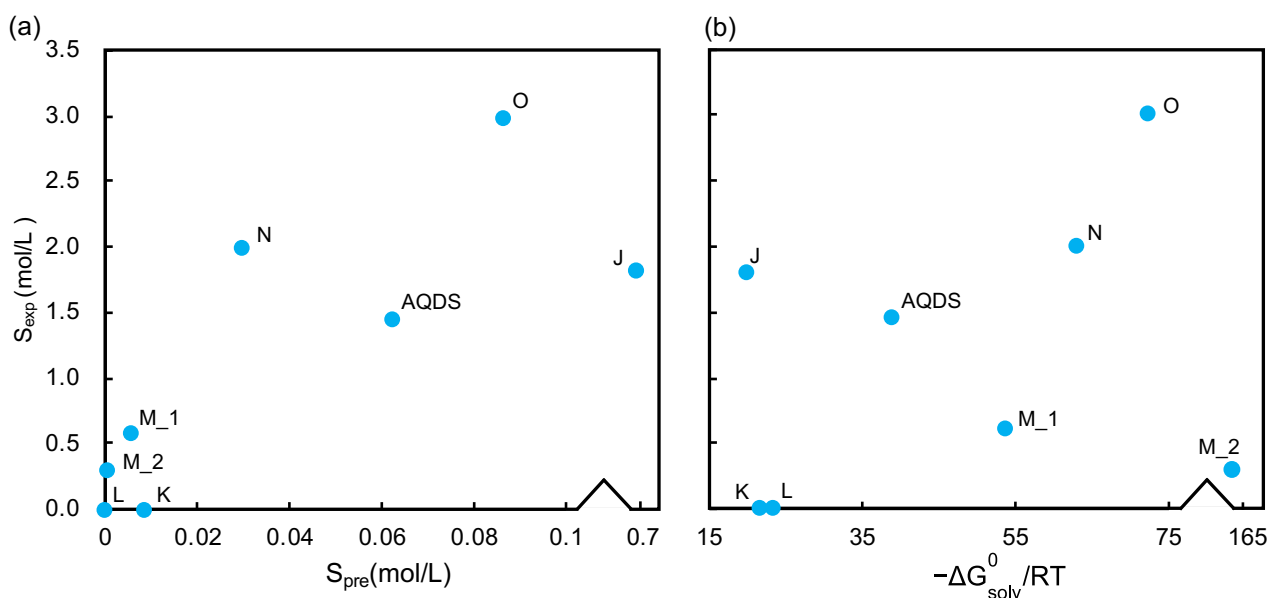**Fig. 8** Comparison between solubility measurements and the predicted data as obtained by **(a)** the AqSolPred ML model, and **(b)** DFT-calculated ΔG^o_{solv}.

We first look at the solubility of the seven compounds in Fig. 8, in which the ML-predicted solubility (Fig. 8a) and the commonly used metric ΔG^o_{solv} (Fig. 8b) were compared to the measured

values. It can be observed that the AqSolPred ML model underestimates solubility, however, it is able to predict the trend within the same order of magnitude. Especially when looking at compounds L, M_1, M_2, N, and O, which are the functional derivatives of indigo (#4), the model-measurement trend is monotonic. This behavior is driven by solubilization due to increasing number of $-SO_3H$ groups. The experimentally measured solubilities of L and K are too low in magnitude and they are in a region where typically significant uncertainties in measurements can occur. On the other hand, the calculated ΔG_{solv}^o does not show any particular trend with respect to the measured solubility.

It must be noted that the solubility is related to ΔG_{solv}^o according to the exact thermodynamic relation $S = p^o \exp\left(-(\Delta G_{sub}^* + \Delta G_{solv}^o)/RT\right)/RT$, where ΔG_{sub}^* is the sublimation free energy. Usually, ΔG_{sub}^* is extremely difficult to predict for organic compounds[51] because of the lack of knowledge about their crystal structures. Besides the fact that the missing value of ΔG_{sub}^* makes the comparison difficult, it can be observed that the exponential proportionality $S \propto \exp\left(-\Delta G_{solv}^o/RT\right)$ doesn't hold for the compounds considered in this study. Based on these observations, in the current work, prediction of solubility by using the AqSolPred ML model is a more accurate approach than by using ΔG_{solv}^o. However, it is difficult to make broader generalizations due to insufficient number of data points and well-known uncertainties in solubility measurements.[52]

The electrochemical performance of candidate compounds was compared with that of the well-studied AQDS molecule. The maximum solubility of AQDS in DI-H_2O was measured as 1.46 M. Although it is a photosensitive molecule, it's stable in acidic solution and when stored in a dark environment the solution does not change color over months. As shown in Fig. S4, electrochemical tests of 0.01 M AQDS dissolved in acidic media with different pH values were evaluated. The linear fit between redox potential and pH data yields a y-axis intercept at 0.227 V which is in reasonable agreement with the DFT-calculated redox potential value of 0.294 V vs. SHE at pH = 0, and in good agreement with the literature data.[9]
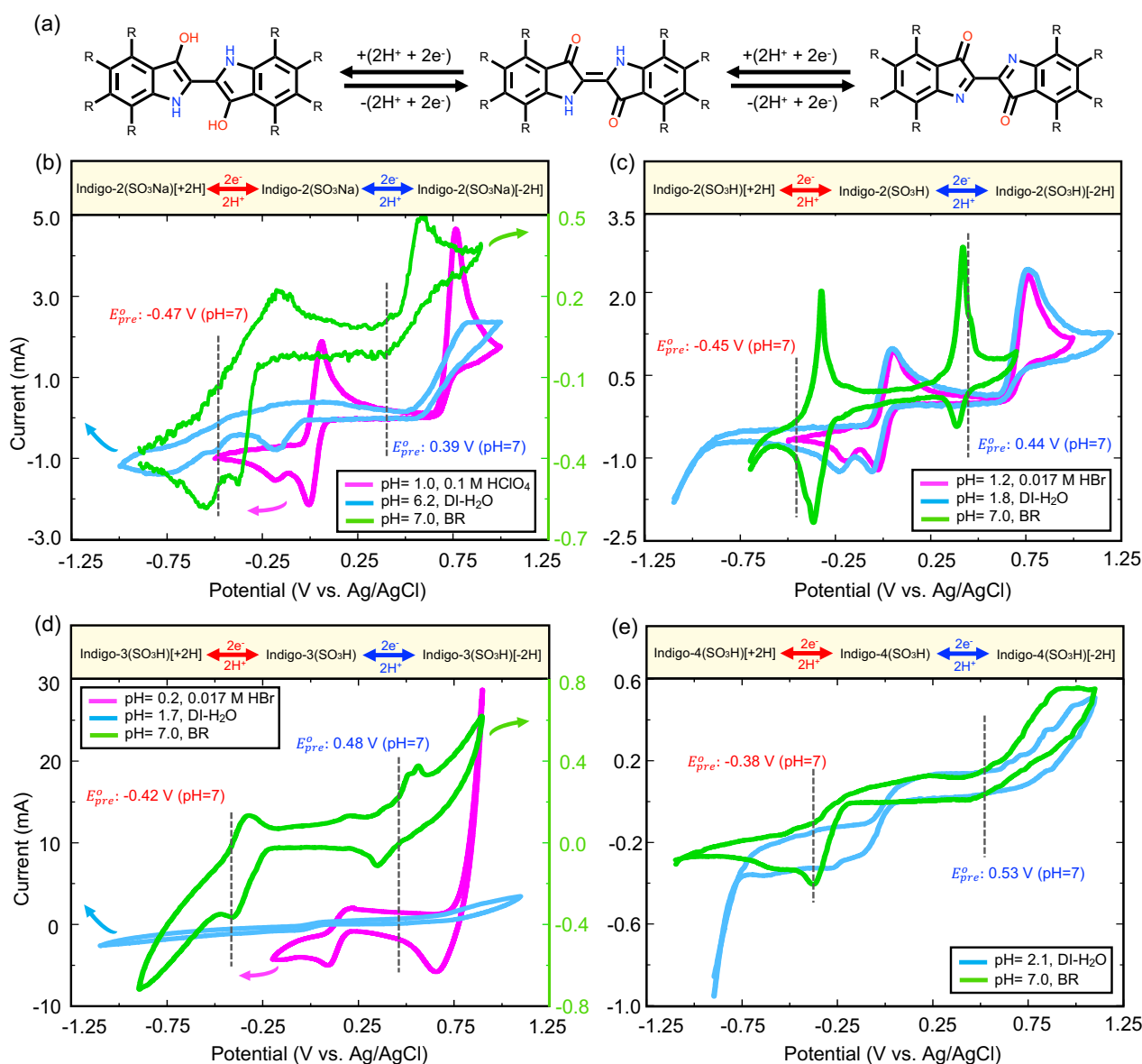
Fig. 9 (a) Two-step reaction scheme for Indigo molecules with three possible oxidation states. Three-electrode cell cyclic voltammetry of **(b)** Indigo-2(SO₃Na), **(c)** Indigo-2(SO₃H), **(d)** Indigo-3(SO₃H), and **(e)** Indigo-4(SO₃H) in various electrolyte solutions. The measured values were in reference to Ag/AgCl/3M KCl electrode. The vertical black lines and the associated data shown in red and blue colored text are the DFT-predicted redox potentials.

Next, we evaluate the electrochemical performance of indigo derivatives that showed high solubility and water stability, namely, Indigo-2(SO₃H)[M_1], Indigo-2(SO₃Na)[M_2], Indigo-

3(SO$_3$H)[N], and Indigo-4(SO$_3$H)[O]. The presence of both donor (nitrogen) and acceptor (carbonyl) groups in the same conjugated organic skeleton leads to three redox states, as shown schematically in the top panel in Fig. 9a. The results of CV experiments on the four compounds are shown in Fig. 9b to 9e. The reduction/oxidation peak centers were determined by reading the highest/lowest current, and the redox potential values were taken as the average of the corresponding reduction and oxidation potential values.

The CV scans for protonated Indigo-2(SO$_3$Na) with two –SO$_3$H functional groups were conducted in 0.1 M HClO$_4$ (pH = 1.0), DI-H$_2$O (pH = 6.2), and BR buffer (pH = 7.0) supporting electrolytes (Fig. 9b). The maximum concentration of Indigo-2(SO$_3$Na) in DI-H$_2$O was 0.21 – 0.3 M. However, the stability of the compound was difficult to determine due to the dark navy color of the solution. In the 0.1 M HClO$_4$ acidic electrolyte, two oxidation peaks were observed at 0.06 and 0.76 V vs. Ag/AgCl in the forward scan, while in the reverse scan there was a dominant reduction peak at -0.01 V that was accompanied by a smaller peak at -0.19 V. These results are in good agreement with the earlier measurements performed in the same electrolyte.[27] In DI-H$_2$O, the CV showed only one oxidation peak at 0.83 V, and two reduction peaks at 0.19 and -0.50 V. The CVs reported by Carretero-González et al.[27] showed an oxidation peak located at 0 V and a reduction peak at -0.4 V within the range from -0.4 to 0.9 V. The differences in number of peaks and their positions are likely because of the different potential ranges chosen for scanning as well as the different electrolyte concentrations. In the BR buffer, the CV showed two oxidation peaks, at -0.16 and 0.59 V in the forward scan, and two reduction peaks at -0.56 and -0.38 V in the reverse scan. While the difference between the pair of peaks (~0.75 V) is similar to the value reported in the previous study,[27] their locations were different. It's likely that the discrepancy is due to the use of a regular 3-electrode 10 mL cell in our study as compared to the cavity microelectrode in their study. The difference in electrode set-up may lead to different cell resistances, which is related with chemical concentration and effective electrode area, etc. The DFT-

predicted redox potentials for the two-steps at pH = 7 for Indigo-2(SO$_3$Na) [-0.47 and 0.39 V vs. Ag/AgCl] agree with our experimentally measured values.

Indigo-2(SO$_3$H) showed a maximum concentration of 0.5 – 0.6 M in DI-H$_2$O, which is similar to the earlier value of 0.583 M,[27] 0.37 M in BR buffer, and 0.75 – 0.79 M in 0.1M HClO$_4$, which is also similar to the earlier value of 0.761 M.[27] Indigo-2(SO$_3$H) did not exhibit reversible electrochemistry as a bifunctional anolyte at pH = 1.2 in 0.017 M HBr. The forward scan showed two oxidation peaks at 0.06 and 0.77 V vs. Ag/AgCl, therefore having a separation of 0.71 V, whereas in the reverse scan only a single reduction peak was observed at -0.04V. These results are in good agreement with the earlier reported measurements in a similar acidic electrolyte.[27] In DI-H$_2$O the molecule showed two oxidation peaks at 0.06 and 0.76 V in the forward scan. In the reverse scan, two reduction peaks were recorded at -0.07 and -0.23 V. In the BR buffer, the forward scan showed two major peaks at -0.32 and 0.42 V, while the reverse scan showed peaks at -0.36 and 0.39 V, in good agreement with the reported measurements.[27] Yet again, the peak positions differed with the reported measurements, although the peak separation was found to be quite similar (~0.75 V).[27] In addition, at pH = 7 Indigo-2(SO$_3$H) exhibits electrochemistry which is distinctively more reversible than Indigo-2(SO$_3$Na). The DFT-predicted redox potentials for the two-steps at pH = 7 for Indigo-2(SO$_3$H) were -0.45 and 0.44 V (vs. Ag/AgCl), which agreed in the range of ~0.1 V of the experimentally observed values of -0.34 and 0.41 V, respectively.

Indigo-3(SO$_3$H) with three –SO$_3$H groups was prepared with a maximum concentration of 2 M in DI-H$_2$O. The electrochemical measurements were performed in 0.017 M HBr (pH = 0.2), DI-H$_2$O (pH = 1.7), and BR buffer (pH = 7) supporting electrolyte. In the acidic HBr electrolyte, the CV showed one oxidation peak at 0.18 V vs. Ag/AgCl and its corresponding reduction peak at 0.10 V. An additional reduction peak at 0.66 V is attributed to HBr itself. In DI-H$_2$O, the CV showed only one oxidation peak at 0.07 V and its corresponding reduction peak at 0.04 V. In the neutral BR buffer, the CVs showed two redox couples: an oxidation peak at -0.33 V and its corresponding reduction peak at

-0.41 V and an oxidation peak at 0.56 V and its corresponding reduction peak at 0.35 V. The separation between the two redox couples was 0.83 V, which is slightly higher than in the case of Indigo-2(SO_3H) (0.77 V), indicating that Indigo-3(SO_3H) is potentially a better bifunctional electrolyte than Indigo-2(SO_3H). For Indigo-3(SO_3H), the DFT-predicted redox potentials for the two-steps at pH = 7 were -0.42 and 0.48 V (vs. Ag/AgCl), which agree well with the experimentally measured values of -0.37 and 0.46 V, respectively.

Lastly, the measured solubility of Indigo-4(SO_3H) in DI-H_2O was approximately 3 M. The electrochemical measurements were performed in DI-H_2O (pH = 2.1) and BR buffer (pH = 7). The CV in DI-H_2O presented one featureless oxidation peak at 0.29 V vs. Ag/AgCl and the corresponding reduction peak at -0.27 V. In the neutral BR buffer, the CV showed a redox couple with oxidation shoulder peak at 0.24 V and a more defined reduction peak at -0.38 V. In addition, another oxidation peak was observed at 0.9 V, which couldn't be assigned to a known electrochemical reaction. The DFT-predicted E^o for the reduction of the carbonyl group was -0.38 V, which agrees well with the observed reduction peak. However, it is ascertained from our experiments that this redox couple doesn't show a reversible electrochemistry. Therefore, the electrochemical behavior of Indigo-4(SO_3H) shows that this compound is neither suitable as a bifunctional electrolyte nor as an anolyte. A likely reason for this instability can be due to structural distortions, which result from a large number of electron-withdrawing groups positioned on the compound that relate to strong intramolecular hydrogen bonds.

In order to better evaluate the performance of the most promising candidate molecule that was identified in the current study, we performed galvanostatic cycling tests and compared the results to the data of AQDS molecule that was experimented under the same conditions, that are at a rate of 200 mAcm^{-2} and 20 °C. The anolyte contained 0.3 M of either Indigo-3(SO_3H) or AQDS and 1.5 M HBr (48%, Acros Organics). On the opposite side, a solution of 1.2 M HBr (62%, Israel Chemicals Ltd), 0.1 M MEP (99.99% purity, Israel Chemicals Ltd) and 0.2 M H_2SO_4 (96%, Acros Organics) were

employed as catholyte, with the Br$^-$/Br$_2$ couple generated by the redox reaction. The catholyte concentration was set as twice of the anolyte concentration that would be required for the cell reactions. Thus, the observed cell capacity fading was attributed to the anolyte active species only. Each electrolyte tank contained 40 ml of electrolyte. All electrochemical impedance spectroscopy (EIS) tests showed similar cell resistances of 0.2-0.3 Ω.
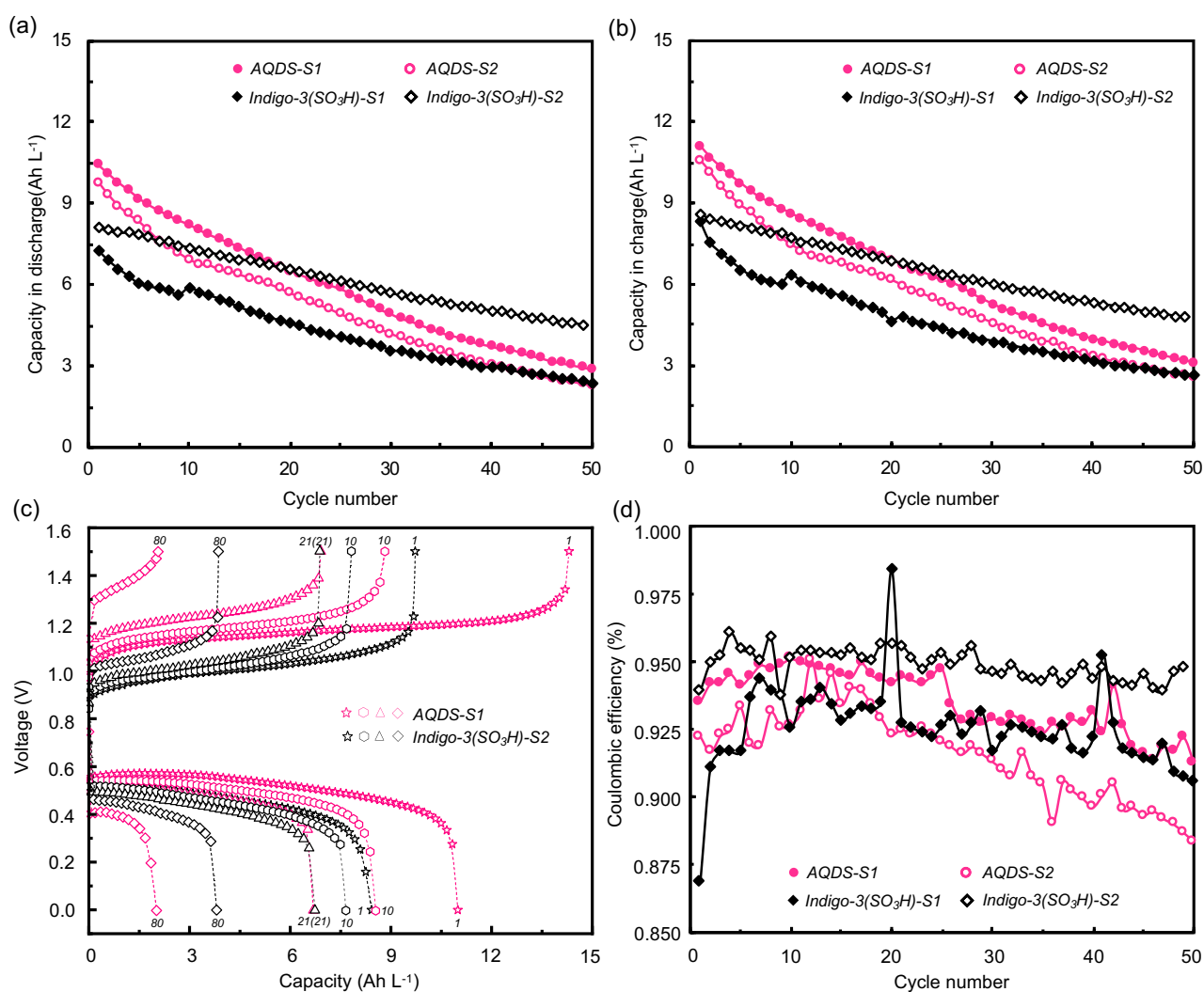

Figure 10. Electrochemical performance of Indigo-3(SO$_3$H) and AQDS: **(a)** discharge capacity, **(b)** charge capacity, **(c)** galvanostatic cycling, and **(d)** coulombic efficiency function of the cycle number.

During the cycling experiments, as shown in plots for discharge (Fig. 10a) and charge (Fig. 10b) capacity as a function of cycle number, it is observed that the sample AQDS-S1 delivered a high initial discharge capacity of 10.4 Ah L^{-1}. However, after 50 cycles, it undergoes a severe capacity fade from 10.4 to 2.9 Ah L^{-1} although the average coulombic efficiency remained above 91.3% (Fig. 10d). In contrast, the sample Indigo-3(SO$_3$H)-S2 delivered a lower initial discharge capacity of 8.1 Ah L^{-1} and showed a slower capacity fade to 4.6 Ah L^{-1} after 49 cycles, while the average coulombic efficiency was above 94.9% (Fig. 10d). After 21 cycles, the sample Indigo-3(SO$_3$H)-S2 showed higher capacity than the AQDS samples in the discharge process. After 80 cycles, the capacity of Indigo-3(SO$_3$H)-S2 was nearly twice as that of AQDS-S1 (Fig. 10c).

4   Conclusions

In this work, we performed a data-driven study on a focused a chemical space of small molecules, which included carbonyl pairs on cyclopentene molecules and their combinatoric variations generated by heterocyclic substitutions. We predicted the candidate materials' redox potentials using a DFT-empowered regression model and aqueous solubilities using a consensus ML model. We identified 205 quinone-like compounds as candidate anolytes for ARFBs that are worthy of experimental validation, which would potentially lead to improvements in battery capacity when compared to a state-of-the-art electroactive molecule, AQDS. In addition, structure-property relationships of the candidate molecules, including the effects of heteroatoms and chemical functional groups, that pertain to tuning the redox potential and aqueous solubility of the compounds were discussed. While the effects of heteroatoms and functional groups on the battery-relevant properties were evident, they were not always systematic, but instead, were peculiar to the different chemical motifs.

To guide our experimental efforts, we performed an automated vendor search on the ZINC database for the entire virtual library of reactant molecules and procured a total of 16 distinct molecules by direct order. Among them, seven molecules were further characterized by using CV and

galvanostatic measurements. As an important result of our data-driven approach, we identified Indigo-$3(SO_3H)$ compound for energy storage in ARFBs. The electrochemical performance as well as the high solubility of Indigo-$3(SO_3H)$ show that the compound is competitive in performance to state-of-the-art organic energy storage compounds. In addition, Indigo-$3(SO_3H)$ can be employed in symmetric flow battery designs, thus offering reduced irreversible capacity losses that are due to active material crossover through the separating membranes. Moreover, Indigo-$3(SO_3H)$ is a relatively large molecule, therefore it is also expected to exhibit a minimal membrane crossover, which in turn would benefit to the battery lifetime. On the down side, unlike the organic anthraquinone species that can be synthesized from inexpensive commodity chemicals,[53] Indigo-$3(SO_3K)$ salt is a relatively more expensive material with a low solubility (< 0.05M in DI water), thereby facing the challenges related to purification and ion-exchange processes. Therefore, the optimization of synthesis routines will be a rewarding step forward for the large-scale production and ample use of this compound in future flow batteries.

Conflicts of interest

There are no conflicts to declare.

Author contributions

S.E. conceived and directed this work. Q.Z. performed the calculations and acquired the computational data. Q.Z. and A.K. interpreted the computational data. E.S. developed the necessary codes and parsed the calculation outputs. F. N., A.L., and I.P. performed the experiments. All authors contributed to the writing of the manuscript.

References

1   Y. Ding, C. Zhang, L. Zhang, Y. Zhou and G. Yu, *Chem. Soc. Rev.*, 2018, **47**, 69–103.

2   D. G. Kwabi, Y. Ji and M. J. Aziz, *Chem. Rev.*, 2020, **120**, 6467–6489.

3   G. Kear, A. A. Shah and F. C. Walsh, *Int. J. Energy Res.*, 2012, **36**, 1105–1120.

4   J. Noack, N. Roznyatovskaya, T. Herr and P. Fischer, *Angew. Chemie - Int. Ed.*, 2015, **54**, 9776–9809.

5   J. Winsberg, T. Hagemann, T. Janoschka, M. D. Hager and U. S. Schubert, *Angew. Chemie - Int. Ed.*, 2016, **55**, 2–28.

6   J. Luo, B. Hu, M. Hu, Y. Zhao and T. L. Liu, *ACS Energy Lett.*, 2019, **4**, 2220–2240.

7   E. J. Son, J. H. Kim, K. Kim and C. B. Park, *J. Mater. Chem. A*, 2016, **4**, 11179–11202.

8   K. Wedege, E. Dražević, D. Konya and A. Bentien, *Sci. Rep.*, 2016, **6**, 1–13.

9   B. Huskinson, M. P. Marshak, C. Suh, S. Er, M. R. Gerhardt, C. J. Galvin, X. Chen, A. Aspuru-Guzik, R. G. Gordon and M. J. Aziz, *Nature*, 2014, **505**, 195–198.

10   K. Lin, Q. Chen, M. R. Gerhardt, L. Tong, S. B. Kim, L. Eisenach, A. W. Valle, D. Hardee, R. G. Gordon, M. J. Aziz and M. P. Marshak, *Science.*, 2015, **349**, 1529–1532.

11   M. R. Gerhardt, L. Tong, R. Gómez-Bombarelli, Q. Chen, M. P. Marshak, C. J. Galvin, A. Aspuru-Guzik, R. G. Gordon and M. J. Aziz, *Adv. Energy Mater.*, 2017, **7**, 1601488.

12   Y. Liu, S. Lu, S. Chen, H. Wang, J. Zhang and Y. Xiang, *ACS Appl. Energy Mater.*, 2019, **2**, 2469–2474.

13   S. Jin, Y. Jing, D. G. Kwabi, Y. Ji, L. Tong, D. De Porcellinis, M.-A. Goulet, D. A. Pollack, R. G. Gordon and M. J. Aziz, *ACS Energy Lett.*, 2019, **4**, 1342–1348.

14   W. Lee, G. Park and Y. Kwon, *Chem. Eng. J.*, 2020, **386**, 123985.

15   C. Wang, Z. Yang, Y. Wang, P. Zhao, W. Yan, G. Zhu, L. Ma, B. Yu, L. Wang, G. Li, J. Liu and Z. Jin, *ACS Energy Lett.*, 2018, **3**, 2404–2409.

16   L. Li, Q. Zhao, Z. Luo, Y. Lu, H. Ma, J. Hu, Y. Li, J. Chen, L. Liu, W. Huang, Z. Luo, L. Liu, Y. Lu, Y. Li, L. Li, J. Hu, H. Ma and J. Chen, *Sci. Adv.*, 2018, **4**, eaao1761.

17   N. El-Najjar, H. Gali-Muhtasib, R. A. Ketola, P. Vuorela, A. Urtti and H. Vuorela, *Phytochem. Rev.*, 2011, **10**, 353–370.

18   I. Klopčič and M. S. Dolenc, *Chem. Res. Toxicol.*, 2019, **32**, 1–34.

19   B. Yang, S. R. Narayanan, L. Hoober-Burkhardt, F. Wang and G. K. Surya Prakash, *J. Electrochem. Soc.*, 2014, **161**, A1371–A1380.

20   M. C. Sorkun, A. Khetan and S. Er, *Sci. data*, 2019, **6**, 143.

21   S. Er, C. Suh, M. P. Marshak and A. Aspuru-Guzik, *Chem. Sci.*, 2015, **6**, 885–893.

22   M. T. Huynh, C. W. Anson, A. C. Cavell, S. S. Stahl and S. Hammes-Schiffer, *J. Am. Chem. Soc.*, 2016, **138**, 15903–15910.

23   D. P. Tabor, R. Gómez-Bombarelli, L. Tong, R. G. Gordon, M. J. Aziz and A. Aspuru-Guzik, *J. Mater. Chem. A*, 2019, **7**, 12833–12841.

24   L. Cheng, R. S. Assary, X. Qu, A. Jain, S. P. Ong, N. N. Rajput, K. Persson and L. A. Curtiss, *J. Phys. Chem. Lett.*, 2015, **6**, 283–291.

25   R. P. Fornari, M. Mesta, J. Hjelm, T. Vegge and P. De Silva, *ACS Mater. Lett.*, 2020, **2**, 239–246.

26   S. B. Kristensen, T. van Mourik, T. B. Pedersen, J. L. Sørensen and J. Muff, *Sci. Rep.*, 2020, **10**, 13571.

27   J. Carretero-González, E. Castillo-Martínez and M. Armand, *Energy Environ. Sci.*, 2016, **9**, 3521–3530.

28   G. A. Chappell, J. K. Britt and S. J. Borghoff, *Food Chem. Toxicol.*, 2020, **140**, 111310.

29   Q. Zhang, A. Khetan and S. Er, *Sci. Rep.*, 2020, **10**, 22149.

30   Q. Zhang, A. Khetan and S. Er, *Sci. Rep.*, 2021, **11**, 4089.

31   M. C. Sorkun, J. M. V. A. Koelman and S. Er, *iScience*, 2021, **24**, 101961.

32   J. J. Irwin and B. K. Shoichet, *J. Chem. Inf. Model.*, 2005, **45**, 177–182.

33   D. A. Kirby, *Microgram J.*, 2011, **8**, 36–38.

34   P. R. N. Carvalho and J. C. H. Collins, *Chromatographia*, 1997, **45**, 63–66.

35   A. Schmillen and R. Legler, *Luminescence of Organic Substances*, 1967.

36   H. Langhals and B. Wagner, U.S. Pat. 6,156,914[P], 2000-12–05.

37   Schrödinger, LLC, New York, NY. Schrödinger materials science suite (2019).

38   E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel and R. A. Friesner, *J. Chem. Theory Comput.*, 2016, **12**, 281–296.

39   K. Roos, C. Wu, W. Damm, M. Reboul, J. M. Stevenson, C. Lu, M. K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel, R. A. Friesner and E. D. Harder, *J. Chem. Theory Comput.*, 2019, **15**, 1863–1874.

40   E. Sorkun, Q. Zhang, A. Khetan, M. C. Sorkun and S. Er, https://doi.org/10.26434/chemrxiv.14398067.v1.

41   A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang and R. A. Friesner, *Int. J. Quantum Chem.*, 2013, **113**, 2110–2142.

42   M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew and K. A. Lyssenko, *Science.*, 2017, **355**, 49–52.

43   J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **78**, 3865–3868.

44   P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 299–310.

45   D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, B. Honig, M. Ringnald and W. A. Goddard, *J. Am. Chem. Soc.*, 1994, **116**, 11875–11882.

46   T. Sterling and J. J. Irwin, *J. Chem. Inf. Model.*, 2015, **55**, 2324–2337.

47   J. P. Janet, S. Ramesh, C. Duan and H. J. Kulik, *ACS Cent. Sci.*, 2020, **6**, 513–524.

48   S. Matsui and H. Aida, *J. Chem. Soc. Perkin Trans. 2*, 1978, 1277–1280.

49   M. Anear and G. Yagil, *J. Am. Chem. Soc.*, 1962, **84**, 1790–1796.

50   J. M. J. M. Ravasco, H. Faustino, A. Trindade and P. M. P. Gois, *Chem. - A Eur. J.*, 2019, **25**, 43–59.

51   R. E. Skyner, J. L. McDonagh, C. R. Groom, T. Van Mourik and J. B. O. Mitchell, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6174–6191.

52   A. Llinàs, R. C. Glen and J. M. Goodman, *J. Chem. Inf. Model.*, 2008, **48**, 1289–1303.

53   Clarence, *J. Am. Chem. Soc.*, 1915, **37**, 2178–2181.