THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

Bio-inspired Electroactive Organic Molecules for Aqueous Redox Flow Batteries: 1. Thiophenoquinones

Sergio D. Pineda Flores, Geoffrey C. Martin-Noble, Richard L. Phillips, and Joshua Schrier

J. Phys. Chem. C, **Just Accepted Manuscript** • DOI: 10.1021/acs.jpcc.5b05346 • Publication Date (Web): 03 Sep 2015

Bio-inspired Electroactive Organic Molecules for Aqueous Redox Flow Batteries: 1. Thiophenoquinones

Sergio D. Pineda Flores, Geoffrey C. Martin-Noble, Richard L. Phillips, and Joshua Schrier*

Department of Chemistry, Haverford College, Haverford, Pennsylvania 19041 USA

E-mail: jschrier@haverford.edu

Phone: +1 (610) 896-1388.

*To whom correspondence should be addressed

Abstract

Redox flow batteries (RFB) utilizing water-soluble organic redox couples are a new strategy for low-cost, eco-friendly, and durable stationary electrical energy storage. Previous studies have focused on benzoquinones, napthoquinones, and anthraquinones as the electroactive species. Here, we explore a new class of molecules—thiophenoquinones—specifically focusing on the caldariellaquinone-, sulfolobusquinone- and benzodithiophenoquinone-like frameworks that are used for metabolic processes in thermophilic aerobic *Sulfolobus* archaebacteria. We demonstrated that B3LYP/6-311+G(d,p) thermochemical calculations (using the SMD solvation model) reproduce experimental reduction potentials to within $\pm$0.04V. We then studied the effect of amine, hydroxyl, methyl, fluoro, phosphonic acid, sulfonic acid, carboxylic acid, and nitro functional group modifications on the reduction potential and Gibbs energy of solvation in water (using density functional theory) and aqueous solubility (using cheminformatics). Next we enumerated all of the 10,611 possible combinations of functional group substitutions on these frameworks, and identified 1056 potential molecules with solubilities exceeding 2 mol/L; of these, 36 molecules have reduction potentials below 0.25V and 15 molecules above 0.95V (versus the standard hydrogen electrode (SHE)). The combination of high solubility and wide voltage range makes these molecules promising candidates for high performance aqueous RFB applications. Finally, using our dataset of *ab initio* reduction potentials, we developed a cheminformatics model that predicts *ab initio* reduction potentials to within $\pm$0.09V based solely on molecular connectivity. We found that a model trained with as few as 200 examples generates rank-ordered predictions allowed us to identify the highest performance candidates with half the number of *ab initio* calculations. This offers a strategy for improving the tractability of future computational searches for high performance RFB molecules.

Keywords: organic redox flow battery (ORBAT), aqueous redox flow battery (RFB), quinone, battery, electrochemistry, thiophenoquinone, caldariellaquinone, sulfolobusquinone, benzodithiophenoquinone, machine learning, data mining, cheminformatics

1 Introduction

Low-cost electrical energy storage is crucial for widespread adoption of renewable energy, but current electrochemical (battery) technologies are too expensive.[1,2] Electrochemical cells store and extract electrical energy via oxidation/reduction reactions of electroactive components. In redox flow batteries (RFBs), the electroactive species are stored external to the electrodes, which offers a number of safety, cost and engineering advantages.[2–7] RFBs have typically used metal ions (e.g., Vanadium) or metal-containing coordination compounds as the electroactive species, but these are toxic and expensive.[1,7] A new strategy is to use organic molecules, specifically quinones. This was first demonstrated for non-aqueous solvent RFBs using anthraquinone derivatives by Wang et al.,[8] and more recently for aqueous RFBs using anthraquinone derivatives by Huskinson et al.[9] and anthraquinone and benzoquinone derivatives by Yang et al.[10] While non-aqueous electrolytes theoretically have higher energy density (due to greater operating voltages), aqueous electrolytes avoid many of the engineering challenges faced by non-aqueous RFBs and thus are more likely to be successfully commercialized in the near future.[11]

Two physicochemical limitations hinder aqueous organic RFB technology.[11] First, the half-cell voltages for individual quinones do not span a sufficiently large range to maximize the theoretical energy density. Water electrolysis places a limit of 1.23 V on the full-cell voltage, although it may be possible to exceed this by taking advantage of the slow kinetics of water electrolysis. Until now, the highest full-cell potential for all-quinone aqueous RFBs is only 0.94 V,[10] although a recent computational study has identified quinone derivatives that might increase this.[12] Second, engineering cost analysis suggests that the solubility of the organic charge carrier should exceed 2 mol/L,[11] but current sulfonic-acid functionalized quinones have only attained 1 mol/L solubilities.[9,10] Several recent papers have discussed the use of large-scale computational screening to identify molecules that improve upon both of these limitations by operating at a wider range of half-cell voltages and having higher solubility.[12,13] Here we use a similar computational strategy to explore a new class of organic electroactive molecules for aqueous RFB applications.

Inspired by Aristotle's dictum that "if one way is better than another, that is the way of Na-

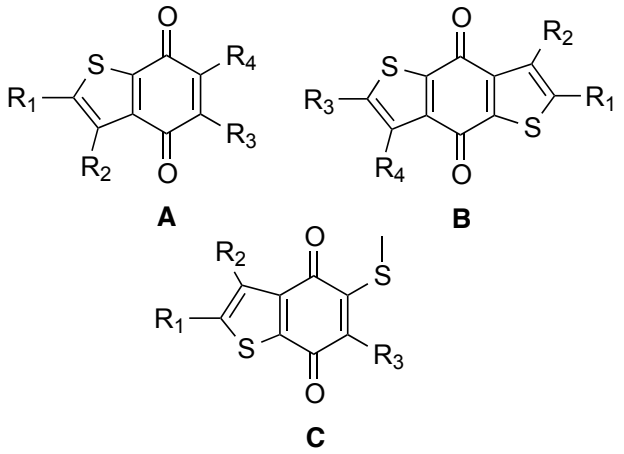

Chart 1: Thiophenoquinone frameworks used in this work. Sulfolobusquinone is **A** with R_3 = (3,7,11,15,19,23-hexamethyltetracosyl), R_4 = CH_3; Caldariellaquinone is **C** with R_3 =(3,7,11,15,19,23-hexamethyltetracosyl); Benzodithiophenoquinone is **B** with R_3 = (6,10,14,18,22-pentamethyltricosan-2-yl).

ture",[14] we observe that benzoquinones and napthoquinones are used almost ubiquitiously for metabolic charge-transport in living organisms. (In turn, previous computational and experimental studies of organic aqueous RFBs have only considered benzo-, naptho-, and anthraquinones.[9,10,12]) However, there are two known exceptions:[15] (1) Thermophilic aerobic *Sulfolobus* archaebacteria that use thiophenoquinones (caldariellaquinone, sulfolobusquinone and benzodithiophenoquinone, shown in Chart 1);[16] and (2) Methanogenic *Methanosarcinales* archaebacteria that use methanophenazines.[17] In this paper, we will discuss the former, and present results on the latter in a future paper. The study of thiophenoquinones for electrochemical energy storage is not completely unprecedented: Hernández-Burgo et al. recently studied thiophenoquinones for lithium battery cathodes.[18] The use of thiophenoquinones by thermophilic organisms suggests that these molecules have enhanced stability at high temperatures. Operating at higher temperatures would both simplify RFB environmental management (reducing the cost) and increase the solubility of the electroactive molecules (increasing the energy density). Moreover, a wide variety of thiophenoquinone derivatives have been synthesized.[19]

Here we describe a combination of *ab initio* quantum chemistry and cheminformatics calculations to assess the electrochemistry and water solubility of $> 10^4$ thiophenoquinone derivatives.

Of the molecules with solubilities predicted to be > 2 mol/L, we identify 36 molecules with reduction potentials < 0.25 V and 15 molecules > 0.95 V as promising new leads for aqueous RFB applications. Using this dataset, we construct a cheminformatics model for predicting the voltage of thiophenoquinone derivatives. Surprisingly, we find that structural information alone results in more accurate predictions than *ab initio* molecular orbital energies or Hückel theory descriptors. We conclude by discussing the implications for future large-scale electrochemical screening projects.

2 Computational Methods

The thiophenoquinone reduction was treated as a single-step two-electron two-proton process, as in previous work.[9,10,12] Starting from a SMILES representation of the oxidized and reduced species, a lowest energy conformer search was performed with the Dreiding force field to obtain initial geometries, using the ChemAxon calculator plugins, Marvin 14.8.25.0, 2014.[20] Although intramolecular hydrogen bonding yields a variety of distinct conformers,[9,21] we found that the use of only a single lowest energy conformer per species was sufficient to obtain quantitative predictions of the reduction potential (see Figure 1 and *vide infra*). All electronic structure calculations were performed using the B3LYP/6-311+G(d,p) model chemistry, as implemented in Gaussian 09.[22] Previous computational studies of quinone electrochemistry found that more computationally expensive methods (MP2, MP4, G4MP2, MP2, CCSD, CCSD(T)) yielded results no more reliable than B3LYP.[23,24]

To compute the reduction potential of a species, Q, we simulated the thermodynamic cycle

$$\begin{array}{ccccc} Q(g) & + & H_2(g) & \xrightarrow{\Delta_r G^\circ_{gas}} & QH_2(g) \\ \downarrow \Delta G_{solv}(Q) & & \downarrow \Delta G_{solv}(H_2) & & \downarrow \Delta G_{solv}(QH_2) \\ Q(aq) & + & H_2(aq) & \xrightarrow{\Delta_r G^\circ_{aq}} & QH_2(aq) \end{array} \quad (1)$$

The gas phase geometries of each species were optimized; these gas phase geometries were kept

fixed for the subsequent solution-phase calculations. The Gibbs energy of each gas-phase species was computed under conditions of T = 298.15 K and $P^\circ = 1$ atm. $\Delta_r G^\circ_{gas}$ was computed by subtracting the Gibbs energy of each of the gas phase reactant species from the product species. The Gibbs energy change of solvation, ΔG°_{solv} of each species has two contributions. The first contribution is the effect of creating a molecular cavity within the solvent and the effect of the solvent medium on the electronic and vibrational energy levels of the solute. The former is calculated by subtracting the gas phase Gibbs energy from the aqueous phase Gibbs energy, the latter computed using (fixed) gas phase geometries and the SMD implicit solvent model with water as the solvent.[25,26] The second contribution is due to the concentration change from the standard state of the gas phase (an ideal gas at $P^\circ = 1$ atm) to the standard state of the solution phase (an ideal solution at 1 mol/L concentration). This contributes an additional $RTln(RT/P^\circ V^\circ)$, for each species (where R is the gas constant). The optimized geometries of all species, and example input file for each step of the electronic structure calculation process are included in the Supporting Information. The reduction potential was calculated by the Nernst equation, $E^\circ_{calc} = -\Delta_r G^\circ_{aq}/nF$, where $n = 2$ is the number of electrons, and F is the Faraday constant. Our computed E° is relative to the standard hydrogen electrode (SHE). As discussed below, the *ab initio* E° quantitatively agrees with experimental potentials for a variety of related quinones and thiophenoquinones, and an empirical correction of $E^\circ_{corr} = 0.03892\text{V} + 0.97751E^\circ_{calc}$ (red line in Figure 1) further improves the quality of the predictions.

Aqueous solubility calculations were performed using the ChemAxon calculator plugins (at a pH of zero),[20] that implement the structure-property relations described in Refs.[27] and.[28] This method is trained to reproduce the aqueous solubility at 20-25°C.

Linear regression models to predict calculated half-cell potentials of derivatives **A** were created in Python 2.7 and Numpy 1.8. The first model ("*ab initio* LUMO") used the *ab initio* LUMO energy of the oxidized gas phase species, from the DFT calculations described above. The second model ("Hückel") used semiempirical Hückel theory calculations of the LUMO energy, the charge density at each atom in the base framework, and the nucleophilic localization energy of each of

the two oxygens on the framework. Semiempirical Hückel calculations were performed using the ChemAxon calculator plugins (at a pH of 1).[20] The third model ("Structure") used one feature for each functional group at each substitution site, calculated using substructure matching in RDKit 2013.09.1.[29] All models were trained using descriptors that are differences of each compound from the base structure. The first two models were trained with an additional constant parameter, whereas the third had its constant parameter fixed to the half-cell potential of the base structure. A table of the descriptors used in each case, and a Python script implementing the parameterization and property predictions are available in the Supporting Information.

3 Results and Discussion

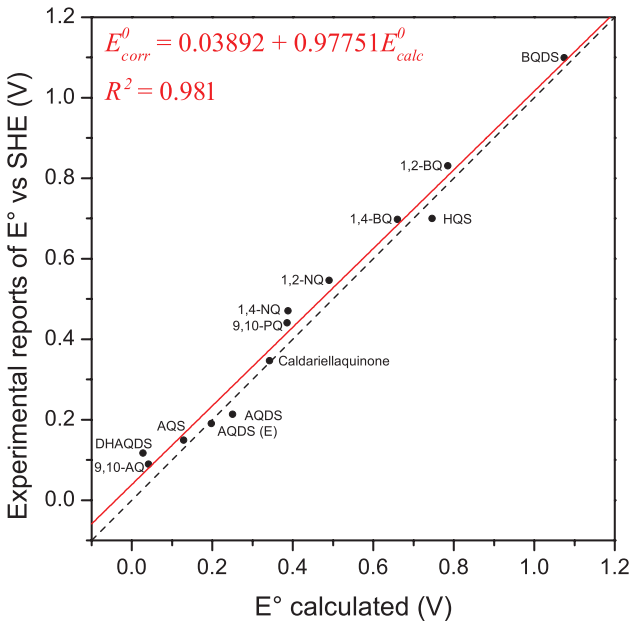

Figure 1: Comparison of calculated half-cell reduction potentials (vs. SHE) to experimental measurements of benzoquinones (BQ), napthoquinones (NQ), anthraquinones (AQ), phenanthreqinones (PQ), sulfonated versions of these (suffix S), and caldariellaquinone. The black dashed line is the bisectrix; the red solid line indicates a least-squares fit to the data. A complete list of abbreviations, potentials, and references for the experiments is contained in Table S1 in the Supporting Information.

How accurate are the calculated half-cell potentials? Figure 1 shows a comparison of our calculated values to experimental redox potentials taken from the literature, for twelve quinones with

a variety of molecular masses, shapes, number of rings, and redox potentials. (To our knowledge, caldariellaquinone is the only thiophenoquinone measurement reported in the literature.) The bisectrix (dashed black line) indicates hypothetical perfect agreement between calculated and experimental results. The calculations agree well with experiment, with mean absolute deviations (MAD) and root-mean-square errors (RMSE) of 0.0435 V and 0.0499 V, respectively. The agreement can be further improved by an empirical linear correction (red line), which yields MAD and RMSE values of 0.0342 V and 0.0397 V, respectively. Both the slope and R^2 values for this fit are close to one, confirming the validity of the underlying DFT calculations. We note that our computed errors compare favorably with the reported experimental errors of 0.002-0.016 V.[9,10,16,24] The empirically corrected values of E° are used in the discussion below; however both the corrected and uncorrected values are available in the data tables contained in the Supporting Information.

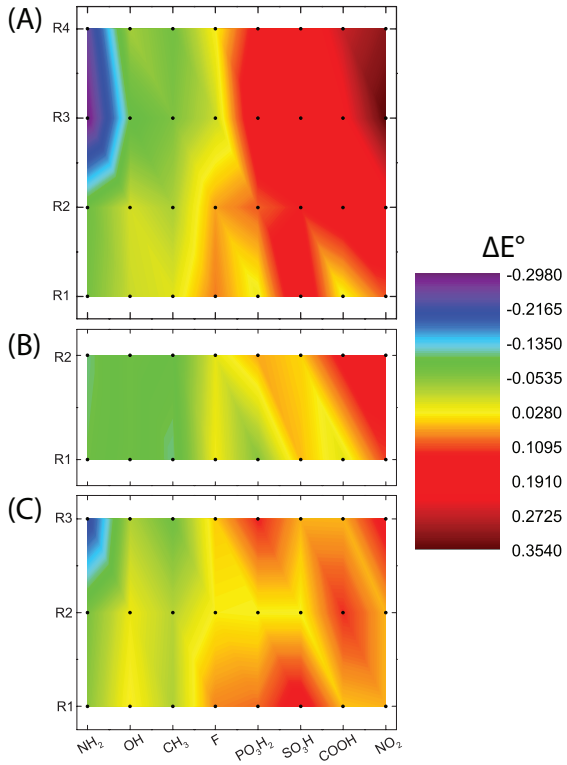

Figure 2: Redox potential change, ΔE°, due to single substitutions of functional groups at each of the symmetry-unique positions. Insets (A), (B), (C) correspond to the framework labels in Chart 1.

How does incorporating functional groups change the redox potential of the various thiophe-

noquinone frameworks? Previous computational studies have explored a variety of functional group substitutions for "tuning" the redox potential.[9,10,12,18,30] In general, one expects Electron Withdrawing Groups (EWG) to decrease the energy of the Lowest Unoccupied Molecular Orbital (LUMO) and Electron Donating Groups (EDG) to increase it. In turn, a lower LUMO energy yields a higher redox potential, because it is energetically more favorable to add an electron into this lower-energy state.[31] To verify this intuition, we examined the change in the redox potential, ΔE°, upon adding a single amine, hydroxyl, methyl, fluoro, phosphonic acid, sulfonic acid, carboxylic acid, or nitro functional group (ordered from most electron donating to most electron withdrawing) on the three thiophenoquinone frameworks shown in Chart 1. Previous computational studies on quinones indicated that these functional groups cause the largest changes in E°.[12] The results, shown in Figure 2A-C, allow us to identify three general trends: (i) EDGs tend to reduce E° and EWGs tend to increase E° as expected. This relationship holds best for EDGs, and deviations are more common for EWGs; (ii) Adding functional groups to the sites on the quinone-ring (R_3 and R_4 on **A**; R_3 on **C**) results in the largest change in redox potential, $|\Delta E^\circ|$. These are the sites where the LUMO has greatest magnitude (see Fig. S1 in the Supporting Information). Adding functional groups to the thiophene-ring sites (R_1 and R_2) causes a smaller $|\Delta E^\circ|$ than addition to the quinone-ring. Considering only these thiophene-ring sites, for **A** and **C**, the largest $|\Delta E^\circ|$ occurs at the site closest to the sulfur atom (R_1), which is consistent with the LUMO magnitude. In the case of **B**, the largest $|\Delta E^\circ|$ occurs at the site farthest from the sulfur atom (R_2), despite the fact that the LUMO density is lowest there. (iii) Exceptions to the general EWG and EDG trends only occur on thiophene-ring sites (R_1 and R_2), and these often result in changes in the sign of ΔE°. For example, (a) adding an OH functional group to the thiophene-ring sites on **A** and **C** results in an opposite sign of ΔE° compared to the quinone-ring sites; (b) adding a PO_3H_2 group on **B** (where there are only thiophene-ring sites), lowers E° when placed near the sulfur (R_1) atom, but raises E° when placed at the other site (R_2).

How does incorporating multiple substitutions of each functional group change the redox potential? Figure 3a shows the computed E° for the 251 possible mono-, di-, tri-, and tetra- sub-

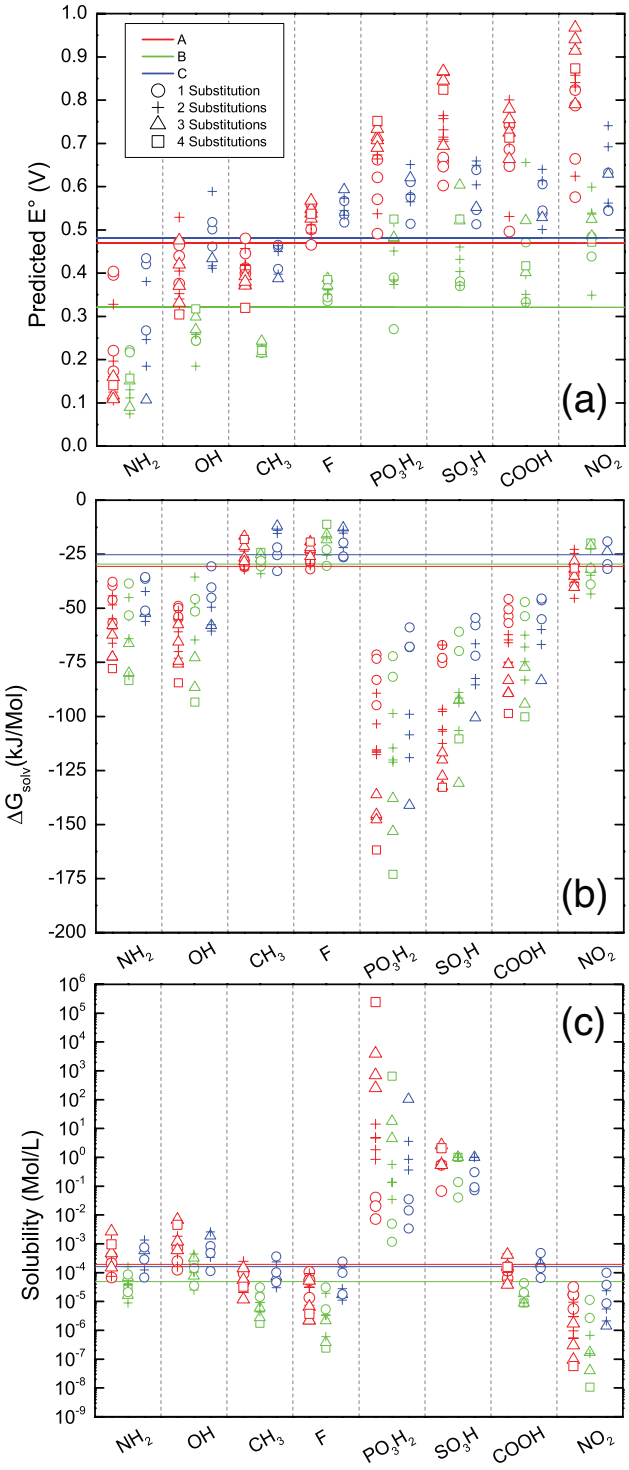

Figure 3: Effects of functional group substitutions on the (a) redox potentials; (b) Gibbs energy of solvation; and (c) aqueous solubility constant of thiophenoquinone frameworks. The horizontal lines indicate the potentials of the unsubstituted thiophenoquinone molecules; symbols indicate potentials for all of the possible single-, double-, triple-, and quadruple- substitutions for each of the functional group types; colors indicate the different thiophenoquinone frameworks.

stitutions of each of the previously considered functional groups.. The results confirm the general trend discussed above—EDGs tend to reduce E° and EWGs tend to increase E°. However, a closer examination shows several exceptions to this general trend. For example, the hydroxyl and methyl EDGs counter-intuitively *increase* E° for some single (circle) and double (cross) substitutions of **C** (blue) and **A** (red). Exceptions also occurs for EWGs: as discussed above, **B** with a single (circle) phosphonic acid functional group has a counter-intuitive *decrease* in E°. Our results also contradict the assumption made by Er et al.[12] that intermediately substituted molecules have E° between the single and full substitutions. Rather, our calculations show that E° changes non-monotonically with increasing number of substitutions, and the most extreme values of E° are typically not the fully substituted molecules. **A** and **B** both have four potential substitution sites and **C** has three potential substitution sites. Consequently, if the assumption were correct, the quadruple-substitutions of **A** and **B** (red and green squares) or triple-substitutions of **C** (blue triangles) should be the most extreme values in Figure 3a. But this is only true for 7 of the 24 columns—in all other cases, molecules with fewer substitutions have the most extreme E°. In fact, the fully hydroxyl-substituted **B** has the *smallest* change in E° from the unsubstituted parent. More frequently (10 of 24 cases) double substitutions have the most extreme value of E°. Surprisingly, dihydroxyl substituted **C** compounds have both the largest possible positive *and* negative changes in E° depending on the substitution sites. There is more variability in E° for di-substituted compounds, simply because more unique cases are possible. However, there is a pattern: In general, we find that $|\Delta E^\circ|$ is smallest when the two functional groups are both on the same ring, and largest when they are on opposite rings. As shown in Table 1, only five exceptions occur (marked with an asterisk): methyl, fluoro, phosphonic acid and nitro di-substitutions on **A**, and amine di-substitutions on **B**. While our present results are limited to thiophenoquinones, we suspect this is also true for napthoquinones and anthraquinones, suggesting the ability to further expand the range of E° beyond the limits described by Er et al.[12] by considering intermediate numbers of substitutions.

What is the effect of functional groups on solvation free energy, ΔG_{solv} (Figure 3b)? In general,

Table 1: R-group positions of di-substituted compounds with the largest $|\Delta E^\circ|$.

Functional Group	**A**	**B**	**C**
NH_2	(1,4)	(1,2)*	(1,3)
OH	(1,4)	(2,4)	(1,3)
CH_3	(3,4)*	(2,4)	(2,3)
F	(1,2)*	(1,4)	(1,3)
PO_3H_2	(3,4)*	(2,4)	(1,3)
SO_3H	(1,3)	(1,4)	(1,3)
COOH	(2,4)	(2,4)	(2,3)
NO_2	(3,4)*	(2,4)	(2,3)

* Asterisks indicates functional groups that are on the same ring.

methyl, fluoro, and nitro groups do not change ΔG_{solv} very much; in fact these can cause small increases *and* decreases in ΔG_{solv} for all frameworks. The other functional groups make solvation more spontaneous, and in particular, phosphonic, sulfonic, and carboxylic acids make ΔG_{solv} the most negative. Unlike the behavior of E° (*vide supra*), increasing the number of functional groups of a given type tends to lower ΔG_{solv} (though the behavior is not strictly monotonic). As a result, because **A** and **B** have four possible sites, they tend to have lower ΔG_{solv} than **C**, which has only three possible sites. Relative to the unsubstituted frameworks, single substitutions have the smallest change in ΔG_{solv} and maximal substitutions have the largest change in ΔG_{solv}, consistent with the previous results of Er *et al.*[12] However, some exceptions occur: (i) some dihydroxyl and diamine substituted **C** molecules have lower ΔG_{solv} than more highly substituted forms; (ii) tri-sulfonic acid **B** has a lower ΔG_{solv} than the tetra-sulfonic acid **B**; (iii) dinitro-**A** compounds show maximal increases *and* decreases in ΔG_{solv} but tetranitro-**A** has only a slightly lower ΔG_{solv}.

Does the *ab initio* ΔG_{solv} (Figure 3b) agree with the cheminformatics-predicted aqueous solubility, S (Figure 3c)? In principle, there is no reason to expect the former (which measures the spontaneity of solvating a molecule originating in the gas phase) to agree with the latter (which measures the spontaneity of solvating a molecule originating in a pure crystal). While both processes will tend to be favored by increasing numbers of hydrogen bonding groups (since we are

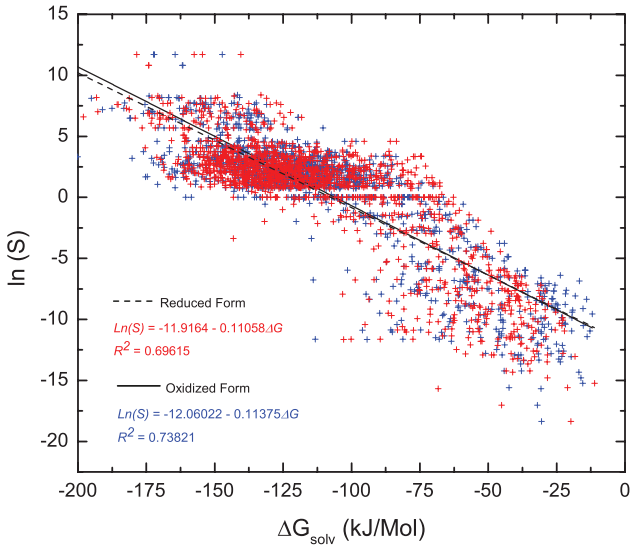

Figure 4: Correlation between the cheminformatics-predicted aqueous solubility constant and the DFT-predicted Gibbs free energy of solvation for all 1,056 molecules considered in this study. Lines indicate least-squares fits to the data.

considering solubility in water), the latter will also depend on intermolecular interactions and packing geometries in the crystal. Despite the fact that S governs the performance of actual RFB devices, previous computational screening surveys only considered ΔG_{solv}. This is understandable since it can be obtained from *ab initio* calculations, whereas *ab initio* molecular crystal thermodynamic properties are prohibitively challenging due to the importance of non-covalent interactions that are only captured with high-level correlation methods.[32] In practice though, we have found ΔG_{solv} and $\ln(S)$ are correlated, as shown in Figure 4. Each point represents a substituted thiophenoquinone; red and blue indicate the reduced and oxidized forms of each molecule, and the lines show a least-squares regression between the variables. Since S is an equilibrium constant, one expects a functional form of the type $\Delta G \propto -RT\,lnS$, and Figure 4 supports a modest correlation following this relationship. Discrepancies between the effects of functional group substitutions on ΔG_{solv} and S are seen by comparing Figure 3b and c. While amine, hydroxyl, and carboxylic acid substitutions cause large decreases in ΔG_{solv} for all frameworks, there are not concomitant increases in S. There are even some cases where despite lower ΔG_{solv} a smaller S is predicted.

Can we use this knowledge to find novel, high-performance thiophenoquinones for aqueous RFB applications? Performing *ab initio* calculations for all of the possible 10,611 possible func-

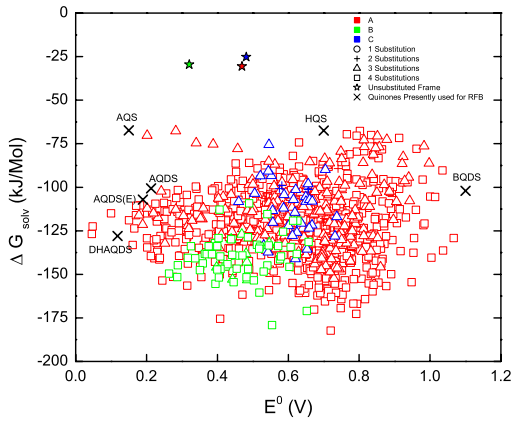

Figure 5: Graphical summary of reduction potential and Gibbs energy of solvation for the 1056 substituted thiophenoquinones predicted to have aqueous solubilities >2 mol/L.

tional group substitutions on **A, B, C** is impractical—this is nearly an order of magnitude larger than any previous study[12,13] of RFB materials—so instead we used the rapid cheminformatics-based calculation of S to perform an initial elimination of many possible candidate molecules before subjecting the remaining compounds to further *ab initio* calculations. Using this as a first stage in our screening process improves upon previous computational exploration strategies in two ways: (i) it takes only a fraction of a second for each candidate, because no quantum mechanical calculations are required; (ii) it optimizes the relevant experimental figure of merit, S. Even if the reader may be skeptical of the precise value obtained by empirical structure-property relationships, the correlation between S and the *ab initio* ΔG_{solv}, discussed above, should allay any fears. From the initial 10,611 possible molecules, retaining only high solubility ($S > 2$ mol/L) molecules leaves 1056 different candidates, which we then subjected to *ab initio* calculations. The Supporting Information contains a machine-readable datafile of E° and ΔG_{solv}—indexed by SMILES strings—along with the optimized geometries of all of these molecules. We summarize these results in graphical form in Figure 5 and Figure 6a-c. First, to achieve $S > 2$ mol/L one typically needs at least three (and often four) functional group substitutions, as evidenced by the preponderance of triangles and squares in Figure 5. No high-S molecules had only a single substitution, and only 21 are di-substituted. Consequently, 98% of the high-S candidates have three or

more substituents attached. The need for highly-substituted frameworks may complicate synthesis, but we note that DHAQDS has four substitutions, so it is not unreasonable. All of the 1056 high solubility candidates contain either a phosphonic or sulfonic acid group, and most contain both—98% of high-*S* candidates have at least one PO_3H_2 and 76% have at least one SO_3H. All 21 di-substituted high-*S* compounds have at least one PO_3H_2 group. This agrees with the increases in solubility associated with phosphonic and sulfonic acid substitutions reported by Er *et al.*,[12] and in our results discussed above. As shown in Figure 6c, all of the thiophenoquinone derivatives have larger values of *S* than the previously reported quinone derivatives (which have *S* of 1.0-1.7 mol/L). Second, Figure 5 shows that the range of reduction potentials is greatly increased compared to the unsubstituted (star) thiophenoquinones. However, as shown in Figure 6a, only a few molecules expand the voltage range beyond previous quinone RFB experiments (labeled vertical lines). Only three **A** derivatives have E° below DHAQDS, and no molecules have E° above BQDS. Third, ΔG_{solv} for these molecules is significantly lower (more spontaneous) than the unsubstituted thiophenoquinones—unsurprising given our *S*-based selection criterion. Moreover, 418 of these candidates have lower ΔG_{solv} (Figure 6b) than DHAQDS. We did not find any molecules with both E° and ΔG_{solv} lower than DHAQDS. However, if one is willing to accept a slight reduction in voltage range compared to DHAQDS and BQDS, there are many thiophenoquinones where ΔG_{solv} is substantially lower.

Within these high-*S* candidates, we identified 36 molecules with $E^\circ < 0.25$ V and 15 molecules with $E^\circ > 0.95$ V. Given the ± 0.05 V MAD of our *ab initio* predictions, this should include all molecules within the < 0.2 V and > 1.0 V feasibility criteria used by Er *et al.*[12] Tables S2 and S3 in the Supporting Information show the functional group substitutions and ΔG_{solv} for both the oxidized and reduced forms of the thiophenoquinones.

To summarize the high E° candidates shown in Table S2: (i) All are derivatives of **A** (as seen in Figure 6a); (ii) 13/15 molecules are fully substituted and 2/15 are tri-substituted; (iii) Most molecules exclusively contain EWGs, but there are 4/15 that contain amine and hydroxyl EDGs; (iv) No molecules contain a methyl group; (v) R_1 and R_2 have a variety of functional groups types;

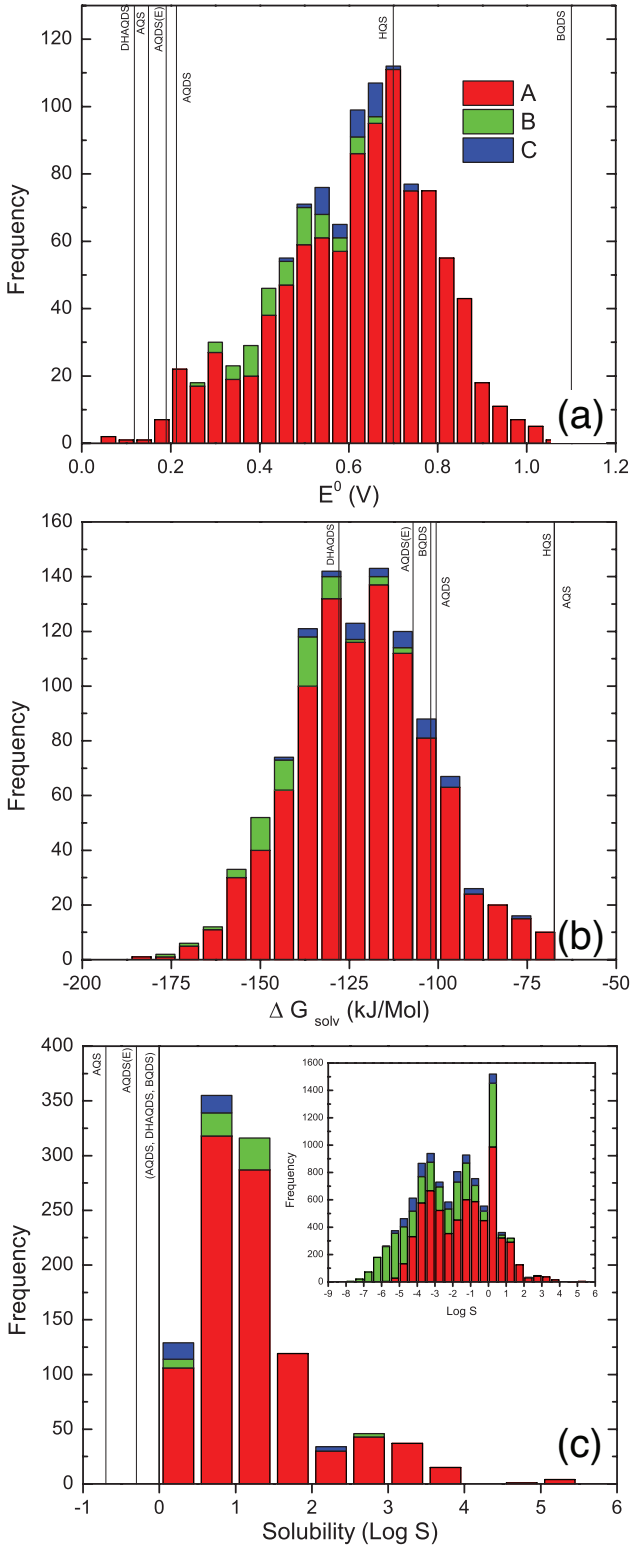

Figure 6: Distributions of (a) reduction potential, E°; (b) Gibbs energy of solvation, ΔG_{solv}; (c) logarithm (base-10) of aqueous solubility for the candidate thiophenoquinones predicted to have solubility $>$ 2 mol/L, $\log S$ (inset shows the values for all of the 10,611 possible substitutions). Vertical lines indicate values for quinones previously studied for use in aqueous RFBs.

(vi) 10/15 molecules have a phosphonic acid on R_3, 4 have sulfonic acid, and 1 has carboxylic acid; (vii) 12/15 molecules have a sulfonic acid on R_4, 1 has phosphonic acid, 1 has fluoro, and 1 has nitro; (viii) 10/15 molecules have a lower ΔG_{solv} than BQDS.

To summarize the low E° candidates shown in Table S3: (i) All (36/ 36) are **A**-derivatives (this is also apparent in Figure 6a); (ii) Again, most (29/36) are fully substituted, and the remaining 7 are tri-substituted; (iii) Predominantly EDGs are present: all molecules contain at least one amine, and there are only 2 carboxylic acid substituted molecules and no nitro substituted examples; (iv) Unlike the high E° case, 8/ 36 molecules contain a methyl group; (v) While 34/36 contain at least one phosphonic acid and 22/36 contain at least one sulfonic acid, this is not significantly different from the global distribution; (vi) Unlike the high E° case, the placement of phosphonic and sulfonic acid groups on the R_3 and R_4 sites shows no systematic trend; (vii) Of the 32/36 examples having exactly two phosphonic and/or sulfonic acid groups, only 2/32 have both of those groups on the same ring (exclusively on R_1 and R_2), all of the rest have the two groups on opposite rings; (viii) 35/36 have a lower ΔG_{solv} than AQDS.

How stable are these molecules to repeated reduction and oxidation cycles? No previous computational study of organic redox flow materials has attempted a direct determination of the reaction pathways and transition state barriers for possible degradation products, as this is a computationally challenging problem. However, Cheng et al. suggested the use of the maximum bond-length change during reduction as a proxy for molecular stability.[13] The well-known bond-order/length/strength relationship suggests that bonds that become significantly longer during the redox process will have the largest reduction in bond order, and thus have the greatest loss in bond-strength and be most susceptible to bond-breaking. The acceptable limits can be determined by considering the changes to the Lewis structures of the molecules during the redox reaction. During the reduction process, the C=O bond on the quinone is converted to a C-O single bond; bond-length tables indicate a difference of 0.2 Å for these two types of bonds. Similarly, the bonds between carbon atoms in the quinone ring are single bonds or double bonds, which become aromatic C-C bonds following reduction; bond length tables suggest an expected change of 0.14 Å. Changes

greater than this suggest weakening of the bonds beyond an acceptable limit, and thus the increased likelihood of degradation during reduction reactions. Tables S2 and S3 in the Supporting Information report the maximum bond length changes for C=O and non-C=O bonds that occur during the reduction process, calculated using the gas-phase optimized geometries. To summarize the results: For the high E° candidate shown in Table S2, the C=O bond length changes were between 0.13-0.15 Å, and the maximum change in a different bond was between 0.11-0.13Å, occurring exclusively on the C-C bond in the quinone ring. Similarly, for the low E° candidates shown in Table S3, the maximum C=O bond length changes were 0.12-0.15 Å, and the C-C bonds in the quinone ring changed by a maximum of 0.09-0.14 Å. Because none of these exceed the thresholds discussed above, we conclude that all of our proposed molecules should be chemically stable over repeated charge-discharge cycles.

Table 2: RMSE and MAD for the predictions of the three models on the training set of 639 compounds and the test set of 319 compounds

	RMSE Training Set	RMSE Test Set	MAD Training Set	MAD Test Set
LUMO	0.142	0.137	0.113	0.109
Hückel	0.109	0.113	0.087	0.092
Structure	0.086	0.088	0.067	0.069

Is it possible to further reduce the number of *ab initio* calculations by developing a cheminformatics model for the reduction potentials? We investigated three linear regression models for reproducing the calculated E° of **A** derivatives. The first model ("*ab initio* LUMO") used the *ab initio* LUMO energy of the oxidized gas-phase compound. Previous work has identified linear correlations between the LUMO of this species and the redox potentials of quinones[12,23] and quinoxalines.[13,30] While this is the simplest model, it is also the most computationally demanding because computing the LUMO energy descriptor requires performing a complete *ab initio* geometry optimization. Moreover, we found that this correlation was not as strong for thiophenoquinone derivatives as the correlations predicted in those earlier works (vide infra). The second model ("Hückel") used semiempirical Hückel theory calculations to obtain: LUMO energy, charge den-

sity at each atom in the base framework, and nucleophilic localization energy of each of the two oxygens on the framework. Like the *ab initio* LUMO model, its descriptors have clear physical meaning, but it has the advantage of the computational ease of performing Hückel calculations to obtain these descriptors. In contrast, the third model ("Structure") relies solely on a group-additivity hypothesis, in which each type of functional group at a particular location makes an additive contribution to E°. Completely neglecting quantum chemistry may seem unlikely to be successful, but hypotheses like these are widely used in quantitative structure activity relationship (QSAR) studies, Hammett free energy relationships, etc. While this is the most complicated model, requiring parameters for each substituent type and location, it is also the most expedient to evaluate, since the descriptors depend only on the atomic connectivity of the candidate model. We trained our models on a random subset of the **A** derivatives on which we performed *ab initio* calculations and reserved one-third (319) of the compounds as a test set. Fitted parameters for each model are available in the supporting information. The structure model significantly outperforms the other two models, despite entirely neglecting quantum chemistry. It reaches a MAD of 0.069 V and an RMSE of 0.088 V (Table 2) on the test set. We investigated regularization for all three models, but there was no significant improvement in the test set error and so we deemed it unnecessary. We also examined hybrid models that combine descriptors of the three models above (Structure + *ab initio* LUMO, Structure + Hückel, Structure + *ab initio* LUMO + Hückel), but found that the MAD and RMSE were only slightly lower (< 0.004 V) than the Structure model shown here, and therefore the extra complication was unwarranted. (See Figure S3 in the Supporting Information.)

How many compounds are needed to properly train the model? Figure 7 shows learning curves of MAD and RMSE for each model as a function of training set size, while keeping the test set constant. It is clear that the models are not overfit as the test and training set errors converge to the same value. All three models stabilize and converge to their final values relatively quickly, with the biggest improvements realized after adding the first 100-200 compounds to the training set. The slightly larger error on the training set (than the test set) in the *ab initio* LUMO model is due to the relatively large errors of the model (see Figure 8) and a fortuitous choice of the training set. We

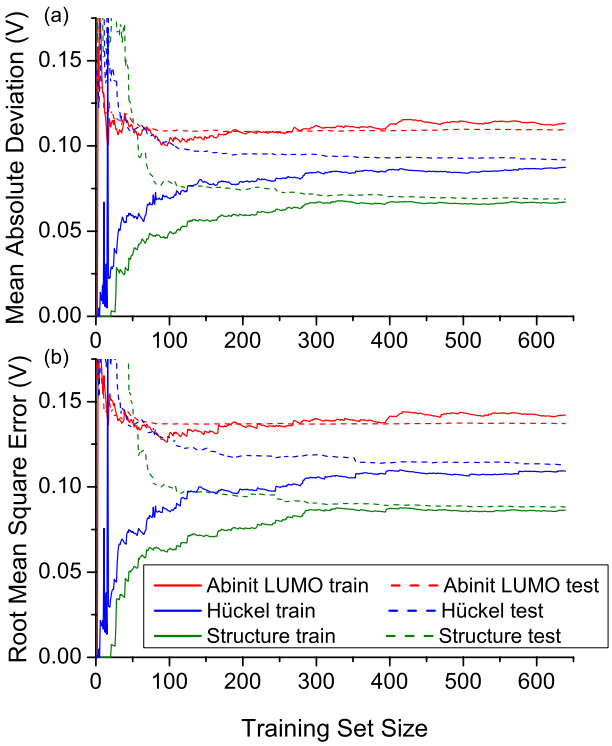

Figure 7: The variation in (a) MAD and (b) RMSE for each regression model with respect to the number of compounds included in the test set. Solid and dashed lines indicate the errors on the training and test sets, respectively.

verified this by creating ten different random training sets, and observed that the average learning curve had the expected behavior.

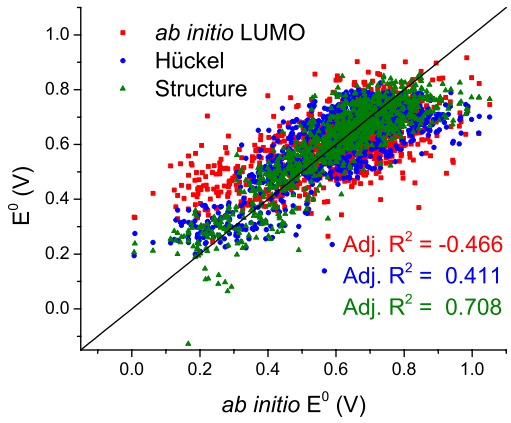

Figure 8: Predicted half cell potentials, E°, from the *ab initio* LUMO (red squares), Hückel (blue circles), and Structure (green triangles) models versus the calculated *ab initio* E° for both the test and training set. The black line is the bisectrix.

Figure 8 shows the predicted reduction potentials for each model versus the calculated *ab initio* potentials. The Structure model is visibly superior (adjusted-$R^2 = 0.708$), with points relatively close to the bisectrix, although there are several significant outliers. We note that the *ab initio* LUMO model has a negative adjusted-R^2 value (-0.466), indicating that the model has larger errors than predicting the mean value for all compounds. The Hückel model (adjusted-$R^2 = 0.411$) does have some explanatory power, but is still significantly worse than the Structure model. Examining the plot of the residuals (difference between predicted and *ab initio* E°) of the structure model versus the calculated *ab initio* E° in Figure 9, we see that the model has a tendency to underestimate the extreme values, especially at high potentials. As shown in Figure 6, most compounds have intermediate voltages, which biases the model towards the mean. This is unfortunate because our goal is to find compounds with extreme E° for RFB applications. However, to use these models to narrow the search space, the rank correlation is a more valuable metric for model performance. Table 3 shows the Spearman rank correlation coefficient for each of the models compared to the calculated *ab initio* E°. Again, the Structure model significantly outperforms the other two models.

How effective is the model at narrowing the search space? We tested the ability of our model to

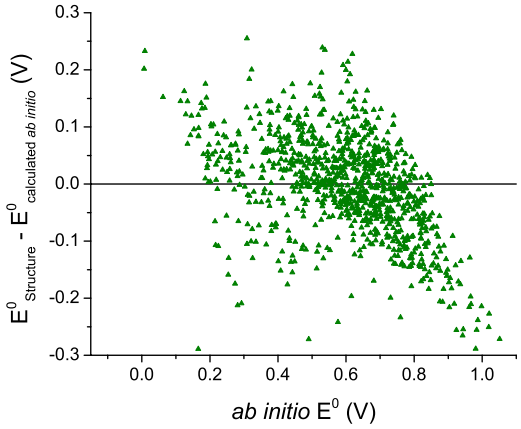

Figure 9: Residuals for the Structure model prediction.

Table 3: Spearman rank correlation coefficients for the test set of 319 compounds using a model trained on the full training set of 639 compounds or on the reduced training set of 200 compounds.

Model	Train 639	Train 200
ab initio LUMO	0.637	0.637
Hückel	0.701	0.698
Structure	0.860	0.828

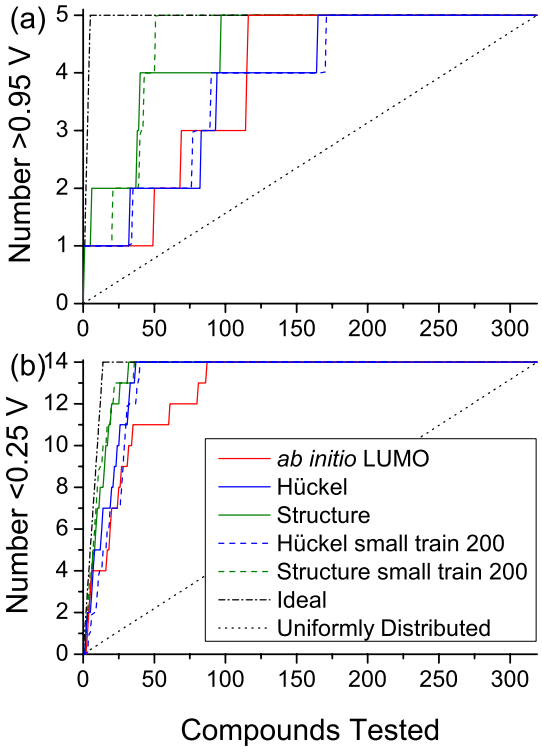

Figure 10: Simulated screening experiment illustrating use of the models for identification of the (a) high and (b) low E° compounds in the test set of 319 compounds. Solid lines indicate models trained on the full training set and dashed lines indicate models trained on only 200 compounds. The reduced training set results are not shown for the *ab initio* LUMO model because the ranking simply corresponds to a reverse ranking by *ab inito* LUMO regardless of the training set size. The black dashed and dotted lines indicate perfect performance from the model and expected performance from simple random selection, respectively.

find "high" ($E^\circ > 0.95$ V) and "low" ($E^\circ < 0.25$ V) compounds using a simulated screening experiment, which is widely used to assess property prediction in identifying bioactive compounds.[33] Figure 10 shows how many of these "active" examples in the test set are identified as such by each of the three models as a function of the number of compounds returned as "active" by the model. The solid lines indicate how many of the high (Figure 10a) and low (Figure 10b) E° compounds in the test set are included in the compounds flagged as active by the models. Once again, the Structure model is the most successful, with all the extreme E° compounds identified earlier than in the other two models. All three models are significantly better at identifying low-potential compounds than high-potential compounds, probably because the low-potential compounds are more numerous and therefore better represented in the training set. One way to quantify these results is by the enrichment factor, the ratio of the number of "active" compounds retrieved when using the model to the number expected when choosing compounds at random. For example, with 50 compounds tested, we measure enrichment factors for the *ab initio* LUMO, Hückel, and Structure models, respectively, of 2.6, 2.6, and 5.1 for the high-potential compounds, and 5.0, 6.4, and 6.4 for the low-potential compounds.

But since training these models required four *ab inito* calculations for each of > 600 compounds, this enrichment does not correspond to a large reduction in computational resource requirement. Figure 10 also shows the results from using a reduced training set of only 200 compounds (dashed lines). Somewhat surprisingly, using a reduced training set does not significantly worsen the models' accuracy, indicating that not many compounds are needed in order to train a model that can effectively reduce the search space. The enrichment factors at 50 compounds are all the same except for the Structure model on the high-potential compounds (Figure 10a), which (fortuitously) increased to 6.4. Thus while Table 2 shows that decreasing the size of the training set increases various error measures, it has little impact on the ability of the model to identify extreme-potential compounds. This is also reflected in the minimal reduction in Spearman correlation coefficient as shown in Table 3. To ensure that these results were not unique to our choice of training and test sets, we performed an additional simulated screening experiment using a different

randomized training set of 200 compounds and the remaining 758 compounds as the test set. The results using this different training set were nearly identical to those described above, and indicate that this method can find all of the high-performance candidates with half the *ab inito* calculations needed for a full enumeration. (See the Supporting Information for details.)

4 Conclusion

We surveyed thiophenoquinone derivatives for aqueous RFB applications, elucidating the effects of functional group placement on reduction potential and solvation properties. In the process, we undermined assumptions made in previous papers about the behavior of reduction potentials with increasing number of functional group substitutions. From a technological standpoint, we identified 51 candidate molecules that meet or exceed existing operating voltages and greatly exceed existing solubilities, thereby increasing the power density (which is the product of the operating voltage and the solubility). As in previous computational studies, we only examined the thermodynamic properties, but we hope this will motivate experimental efforts to synthesize these molecules and characterize the kinetics of the protonation process. More generally, we demonstrated two strategies to accelerate the computational screening process by removing non-viable molecules: (i) use of existing aqueous solubility prediction codes; and (ii) construction of a cheminformatics model to predict reduction potentials based on a limited set of *ab initio* results. Based on a simulated screening experiment, we found that a model trained on 200 examples can identify leading candidates at a fraction of the computational cost needed for a full *ab initio* characterization of every compound. This strategy is generally applicable to future computational search projects for electrochemical materials.

Acknowledgement

We thank Prof. Jonathan Wilson for directing us to Ref.,[15] and Julian Taylor for a careful reading of the manuscript. JS acknowledges the Henry Dreyfus Teacher-Scholar Award program. This

work used the resources of the National Energy Research Scientific Computing Center (NERSC) which supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Supporting Information Available

Table of experimental quinone redox potentials, plots of LUMO, tables of high-performance candidates, and regression model parameters. Machine-readable comma-separated-value (CSV) datafile of computed solvation and electrochemical results indexed by simplified molecular-input line-entry system (SMILES) representation. Optimized geometries in XYZ format. Sample Gaussian 09 input files for each step of the calculation. Python script implementing the cheminformatics models and their parameterization. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

The authors declare no competing financial interests.

References

(1) Dunn, B.; Kamath, H.; Tarascon, J.-M. Electrical Energy Storage for the Grid: A Battery of Choices. *Science* **2011**, *334*, 928–935.

(2) Yang, Z.; Zhang, J.; Kintner-Meyer, M. C. W.; Lu, X.; Choi, D.; Lemmon, J. P.; Liu, J. Electrochemical Energy Storage for Green Grid. *Chem. Rev.* **2011**, *111*, 3577–3613.

(3) Skyllas-Kazacos, M.; Chakrabarti, M. H.; Hajimolana, S. A.; Mjalli, F. S.; Saleem, M. Progress in Flow Battery Research and Development. *J. Electrochem. Soc.* **2011**, *158*, R55.

(4) Weber, A. Z.; Mench, M. M.; Meyers, J. P.; Ross, P. N.; Gostick, J. T.; Liu, Q. Redox Flow Batteries: A Review. *J. Appl. Electrochem.* **2011**, *41*, 1137–1164.

(5) Weber, A. Z.; Mench, M. M.; Meyers, J. P.; Ross, P. N.; Gostick, J. T.; Liu, Q. Erratum to: Redox Flow Batteries: A Review. *J. Appl. Electrochem* **2011**, *41*, 1165–1166.

(6) Huang, Q.; Wang, Q. Next-Generation High-Energy-Density Redox Flow Batteries. *ChemPlusChem* **2015**, *80*, 312–322.

(7) Ding, C.; Zhang, H.; Li, X.; Liu, T.; Xing, F. Vanadium Flow Battery for Energy Storage: Prospects and Challenges. *J. Phys. Chem. Lett.* **2013**, *4*, 1281–1294.

(8) Wang, W.; Xu, W.; Cosimbescu, L.; Choi, D.; Li, L.; Yang, Z. Anthraquinone with Tailored Structure for a Nonaqueous Metal–Organic Redox Flow Battery. *Chem. Commun.* **2012**, *48*, 6669.

(9) Huskinson, B.; Marshak, M. P.; Suh, C.; Er, S.; Gerhardt, M. R.; Galvin, C. J.; Chen, X.; Aspuru-Guzik, A.; Gordon, R. G.; Aziz, M. J. A Metal-free Organic–Inorganic Aqueous Flow Battery. *Nature* **2014**, *505*, 195–198.

(10) Yang, B.; Hoober-Burkhardt, L.; Wang, F.; Prakash, G. K. S.; Narayanan, S. R. An Inexpensive Aqueous Flow Battery for Large-Scale Electrical Energy Storage Based on Water-Soluble Organic Redox Couples. *J. Electrochem. Soc.* **2014**, *161*, A1371–A1380.

(11) Darling, R. M.; Gallagher, K. G.; Kowalski, J. A.; Ha, S.; Brushett, F. R. Pathways to Low-cost Electrochemical Energy Storage: A Comparison of Aqueous and Nonaqueous Flow Batteries. *Energy Environ. Sci.* **2014**, *7*, 3459–3477.

(12) Er, S.; Suh, C.; Marshak, M. P.; Aspuru-Guzik, A. Computational Design of Molecules for an All-quinone Redox Flow Battery. *Chem. Sci.* **2015**, *6*, 885–893.

(13) Cheng, L.; Assary, R. S.; Qu, X.; Jain, A.; Ong, S. P.; Rajput, N. N.; Persson, K.; Curtiss, L. A. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 283–291.

(14) Aristotle, Progression of Animals. In *Complete works of Aristotle, Volume 1: The Revised Oxford Translation*; Barnes, J., Ed.; Princeton University Press, 2014; p 1097.

(15) Nowicka, B.; Kruk, J. Occurrence Biosynthesis and Function of Isoprenoid Quinones. *Biochim. Biophys. Acta, Bioenergetics* **2010**, *1797*, 1587–1605.

(16) Schäfer, G. Bioenergetics of the archaebacterium Sulfolobus. *Biochim. Biophys. Acta, Bioenergetics* **1996**, *1277*, 163–200.

(17) Beifuss, U.; Tietze, M. Methanophenazine and Other Natural Biologically Active Phenazines. In *Natural Products Synthesis II*; Mulzer, J., Ed.; Springer Berlin Heidelberg, 2005; Vol. 244, pp 77–113.

(18) Hernández-Burgos, K.; Burkhardt, S. E.; Rodríguez-Calero, G. G.; Hennig, R. G.; Abruña, H. D. Theoretical Studies of Carbonyl-Based Organic Molecules for Energy Storage Applications: The Heteroatom and Substituent Effect. *J. Phys. Chem. C* **2014**, *118*, 6046–6051.

(19) Chenard, B. L.; McConnell, J. R.; Swenton, J. S. An Efficient Route to Benzo[b]thiophenequinones Quinone Bisketals, and Quinone Monoketals via Anodic

Oxidation of Methoxylated Benzo[b]thiophenes. Studies Directed at the Synthesis of Functionalized Benzo[b]thiophene-4,7-quinones. *J. Org. Chem.* **1983**, *48*, 4312–4317.

(20) *Marvin 14.8.25.0, 2014, ChemAxon (http://www.chemaxon.com)*, 2014. `http://www.chemaxon.com` (accessed September 3, 2015).

(21) Qu, R.; Liu, H.; Feng, M.; Yang, X.; Wang, Z. Investigation on Intramolecular Hydrogen Bond and Some Thermodynamic Properties of Polyhydroxylated Anthraquinones. *J. Chem. Eng. Data* **2012**, *57*, 2442–2455.

(22) Frisch, M. J. et al. *Gaussian 09 Revision C.01*, 2010, Gaussian, Inc., Wallingford CT, 2010.

(23) Bachman, J. E.; Curtiss, L. A.; Assary, R. S. Investigation of the Redox Chemistry of Anthraquinone Derivatives Using Density Functional Theory. *J. Phys. Chem. A* **2014**, *118*, 8852–8860.

(24) Wass, J. R. T. J.; Ahlberg, E.; Panas, I.; Schiffrin, D. J. Quantum Chemical Modeling of the Reduction of Quinones. *J. Phys. Chem. A* **2006**, *110*, 2005–2020.

(25) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(26) Marenich, A. V.; Ho, J.; Coote, M. L.; Cramer, C. J.; Truhlar, D. G. Computational Electrochemistry: Prediction of Liquid-phase Reduction Potentials. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15068.

(27) Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. J. ADME Evaluation in Drug Discovery. Part 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach. *J. Chem. Inform. Comput. Sci.* **2004**, *44*, 266–275.

(28) Shoghi, E.; Fuguet, E.; Bosch, E.; Ràfols, C. Solubility–pH Profiles of some Acidic Basic and Amphoteric Drugs. *Eur. J. Pharm. Sci.* **2013**, *48*, 291–300.

(29) *RDKit: Open-source Cheminformatics (http://www.rdkit.org)*. `http://www.rdkit.org` (accessed September 3, 2015) .

(30) Assary, R. S.; Brushett, F. R.; Curtiss, L. A. Reduction Potential Predictions of some Aromatic Nitrogen-containing Molecules. *RSC Adv.* **2014**, *4*, 57442–57451.

(31) Streitwieser, A. *Molecular Orbital Theory for Organic Chemists*; Wiley: New York, 1961.

(32) Beran, G. J. O. Eine neue Ära in der quantenchemischen Berechnung der Gitterenergie von Molekülkristallen. *Angew. Chem.* **2014**, *127*, 406–408.

(33) Leach, A. R.; Gillet, V. J. *An Introduction to Cheminformatics*; Springer: Dordrecht, 2007 .

Graphical Table of Contents