Electronic Supplementary Material (ESI) for Digital Discovery.


**Supplementary Information**

# *In silico* Discovery of a New Class of Anolyte Redoxmers for Non-aqueous Redox Flow Batteries

Akash Jain,[1,2] Ilya A. Shkrob,[1,3] Hieu A. Doan,[1,2], Lily A. Robertson, [1,3] Lu Zhang, [1,3] and Rajeev S. Assary[1,2*]

[1]Joint Center for Energy Storage Research (JCESR), Argonne National Laboratory, Lemont, IL 60439, U.S.A

[2]Materials Science Division, Argonne National Laboratory, Lemont, IL 60439, U.S.A

[3]Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, U.S.A

* Corresponding Author

**Corresponding Author**

Rajeev S. Assary, assary@anl.gov , Phone: 630-252-3536

We provide the SRMGA code and data of different molecular libraries (SMILES, complexity score, and redox potentials in CSV files) on GitHub at https://github.com/akashjn/MolGenerator. The active learning code can be found on GitHub at https://github.com/akashjn/Machine_Learning_Chemistry/blob/main/BTZ_1500_mols/Active_Learning_for_1500_BTZmols.ipynb

Table of Contents

## S1. Supporting Tables.

**Table S1: Molecular descriptors calculated using the RDKit package.[2-6]**

| feature # | descriptor | feature# | descriptor |
|---|---|---|---|
| 0 | sssr | 25 | number_HBD |
| 1 | clogp | 26 | number_hetero_atoms |
| 2 | mr | 27 | number_hetero_cycles |
| 3 | mw | 28 | number_rings |
| 4 | tpsa | 29 | number_rotatable_bonds |
| 5 | chi0n | 30 | number_saturated_rings |
| 6 | chi1n | 31 | number_heavy_atoms |
| 7 | chi2n | 32 | number_nh_oh |
| 8 | chi3n | 33 | number_n_o |
| 9 | chi4n | 34 | number_valence_electrons |
| 10 | chi0v | 35 | max_partial_charge |
| 11 | chi1v | 36 | min_partial_charge |
| 12 | chi2v | 37 | fr_methoxy |
| 13 | chi3v | 38 | fr_ether |
| 14 | chi4v | 39 | fr_NH1 |
| 15 | fracsp3 | 40 | fr_NH0 |
| 16 | hall_kier_alpha | 41 | fr_aniline |
| 17 | kappa1 | 42 | fr_unbrch_alkane |
| 18 | kappa2 | 43 | fr_bicyclic |
| 19 | kappa3 | 44 | fr_aryl_methyl |
| 20 | labuteasa | 45 | fr_Ndealkylation1 |
| 21 | number_aliphatic_rings | 46 | fr_ester |
| 22 | number_atom_stereocenters | 47 | fr_C_O_noCOO |
| 23 | number_bridgehead_atoms | 48 | fr_C_O |
| 24 | number_HBA | | |

**Table S2: SMILES casts for scaffolds in Figure 3.**

| Scaffold | SMILES |
|---|---|
| $B$ | [R1]C1=C([R2])C([R3])=C([R4])C2=NSN=C21 |
| $S_1$ | [R1]C1=C2C(N(C[R4])C([R3])N2C[R2])=C([R5])C3=NSN=C31 |
| $S_2$ | [R1]C1=C(N(C[R2])CCN2C[R3])C2=C([R4])C3=NSN=C31 |
| $S_3$ | [R1]C1=C(OC([R2])C([R3])O2)C2=C([R4])C3=NSN=C31 |
| $S_4$ | [R1]C1=C(C2CCC3C2)C3=C([R2])C4=NSN=C41 |
| $S_5$ | [R1]C1=C(C2C([R2])C([R3])C3CC2)C3=C([R4])C4=NSN=C41 |
| $S_6$ | [R1]C(C([R2])O1)OC2=C1C([R3])=C([R4])C3=NSN=C32 |

The symbols $R_1$ to $R_5$ refer to the sites identified in Figure 3. Tables 1, and 2 (in the main text) are used to decorate these sites with groups listed in this table.

## S2. Supporting Figures.

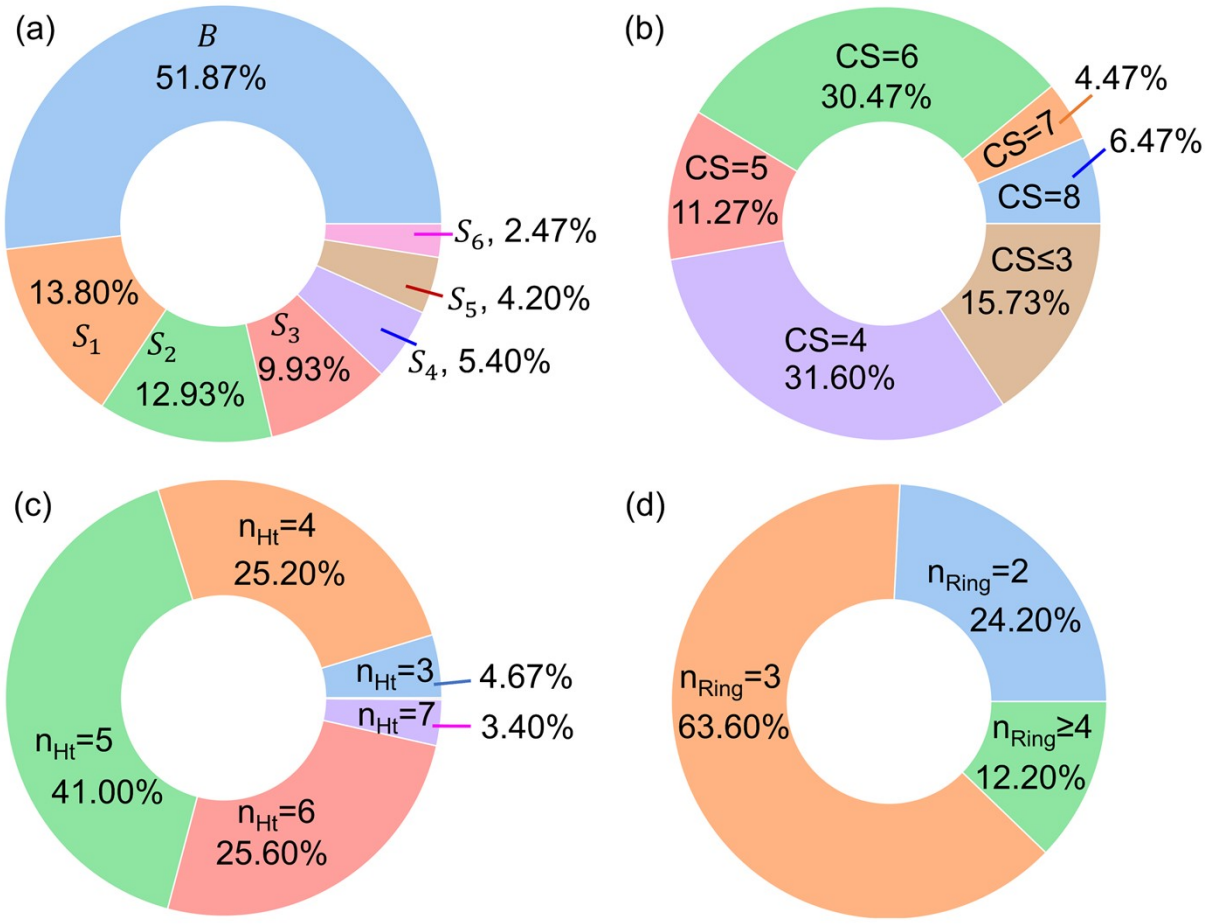


**Figure S1.** The bird's eye view of the set of 1,500 computed molecules drawn from a larger set of 35,500 derivatives (see Figures 4a-d for this larger set). The ring diagrams show the fractions of (a) the scaffolds defined in Figure 3, (b) the complexity scores (CS), (c) the number of heteroatoms ($n_{Ht}$), and (d) the total number of rings, including aromatic and aliphatic rings ($n_{Ring}$). See also Figure S3.

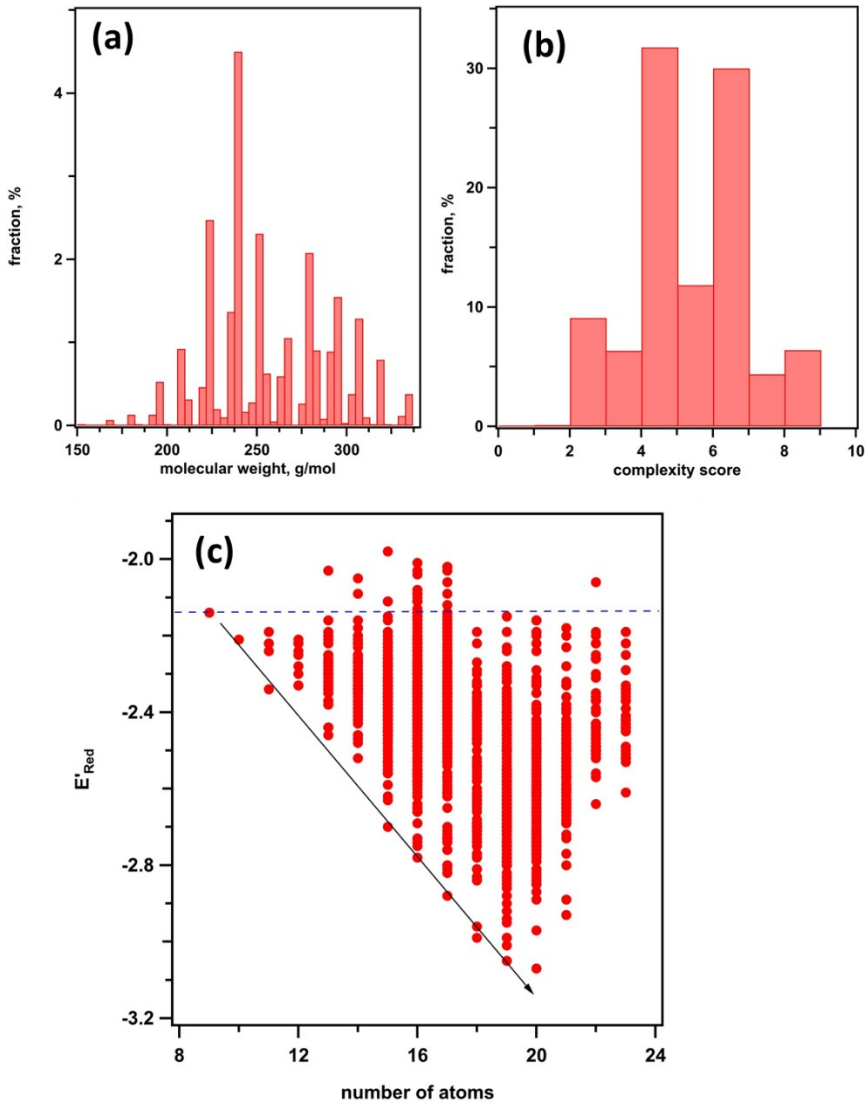


**Figure S2.** The frequency histograms for (a) molecular weights and (b) complexity scores in a set of 1,500 DFT-computed BTZ derivatives drawn randomly from a larger library of 35,5000 derivatives. (c) The plot of $1e^-$ redox potential ($E'_{Red}$) vs. the number of atoms in a molecule, with each compound represented with a filled circle. $E'_{Red}$ does not include the vibrational corrections and it is given in the units of V vs. $Ag/Ag^+$ in acetonitrile. The horizontal dashed line in panel c shows the redox potential for the parent BTZ molecule. The black line indicates molecules with the lowest redox potentials for a given number of atoms. As seen from this plot, there is a tendency for such molecules to have lower redox potentials as their complexity increases. The molecules densely cover the triangular space defined by these two lines. The incomplete filling of this region for larger molecules is caused by our intentional under sampling of complex molecules. This can be addressed by using a higher $CS_{max}$ cutoff or a lower exponential factor $\beta$ in Eq. 5.

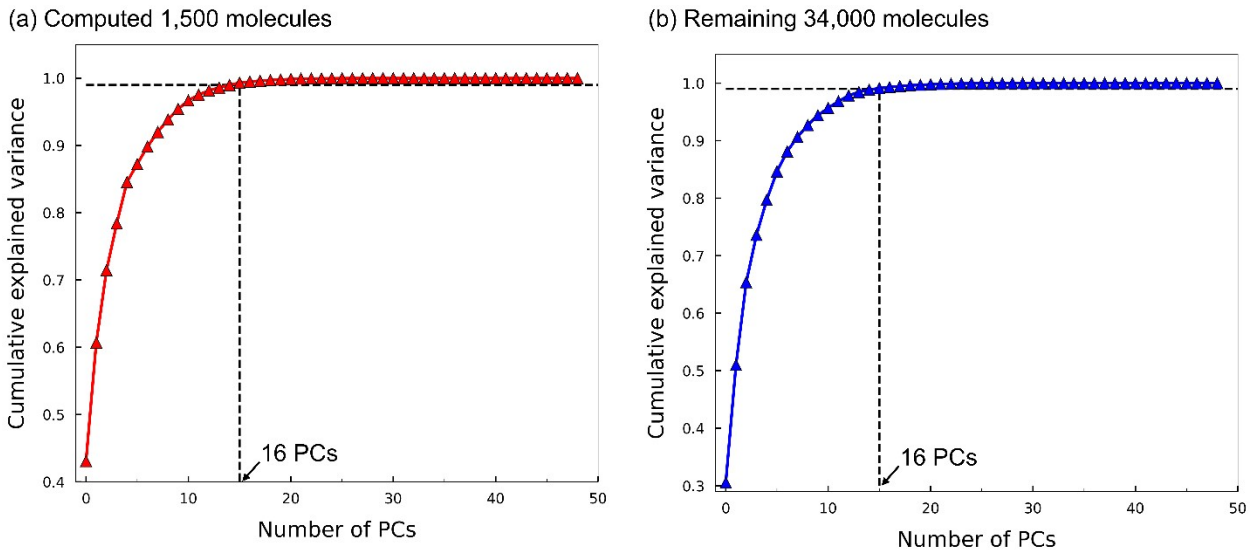


**Figure S3.** The principal component analysis of 49 RDKit generated molecular descriptors[2-6] for (a) computed 1,500 molecules (the red triangles and red lines) and (b) the remaining 34,000 molecules (the blue triangles and blue lines). See Table S1 for the names of these molecular descriptors. Sixteen principal components (PCs) explain at least 99% variance in the data in (a) and (b) (see the vertical and horizontal dashed lines in the plot). 16 PCs were used as features for AL, see Figure 2.

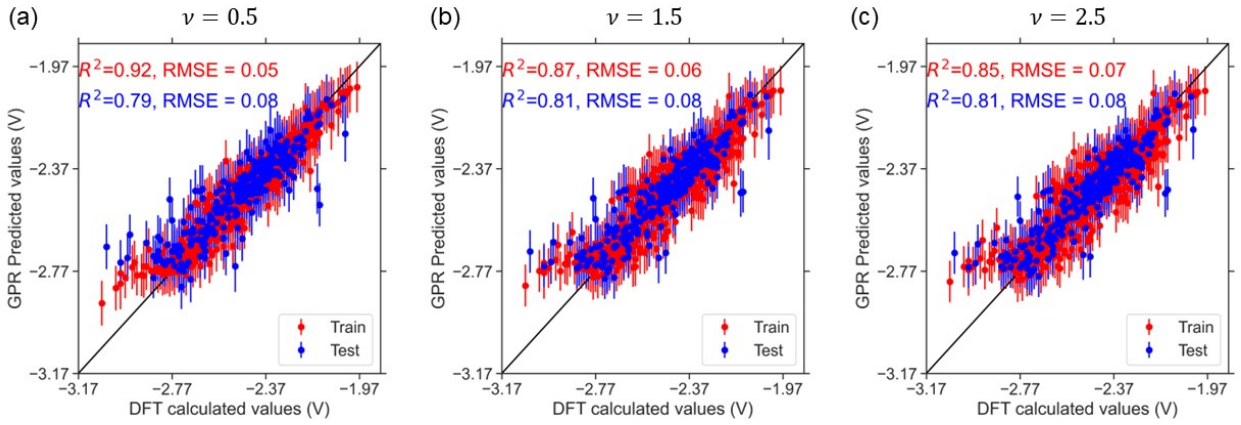


**Figure S4.** Parity plots of GPR-predicted and DFT-computed reduction potentials ($E_{Red}^{'}$) for different values of $\nu = 0.5$, 1.5 and 2.5 in the Matern kernel. Dataset of randomly selected 1500 is split into train (80%) and test (20%) datasets that are shown in red and blue colors, respectively. The vertical bar at each datapoint is standard deviation in the GPR predicted $E_{Red}^{'}$ value. We selected the default $\nu$=1.5 because the performance ($R^2$ and RMSE) of GPR model doesn't significantly change with $\nu$.

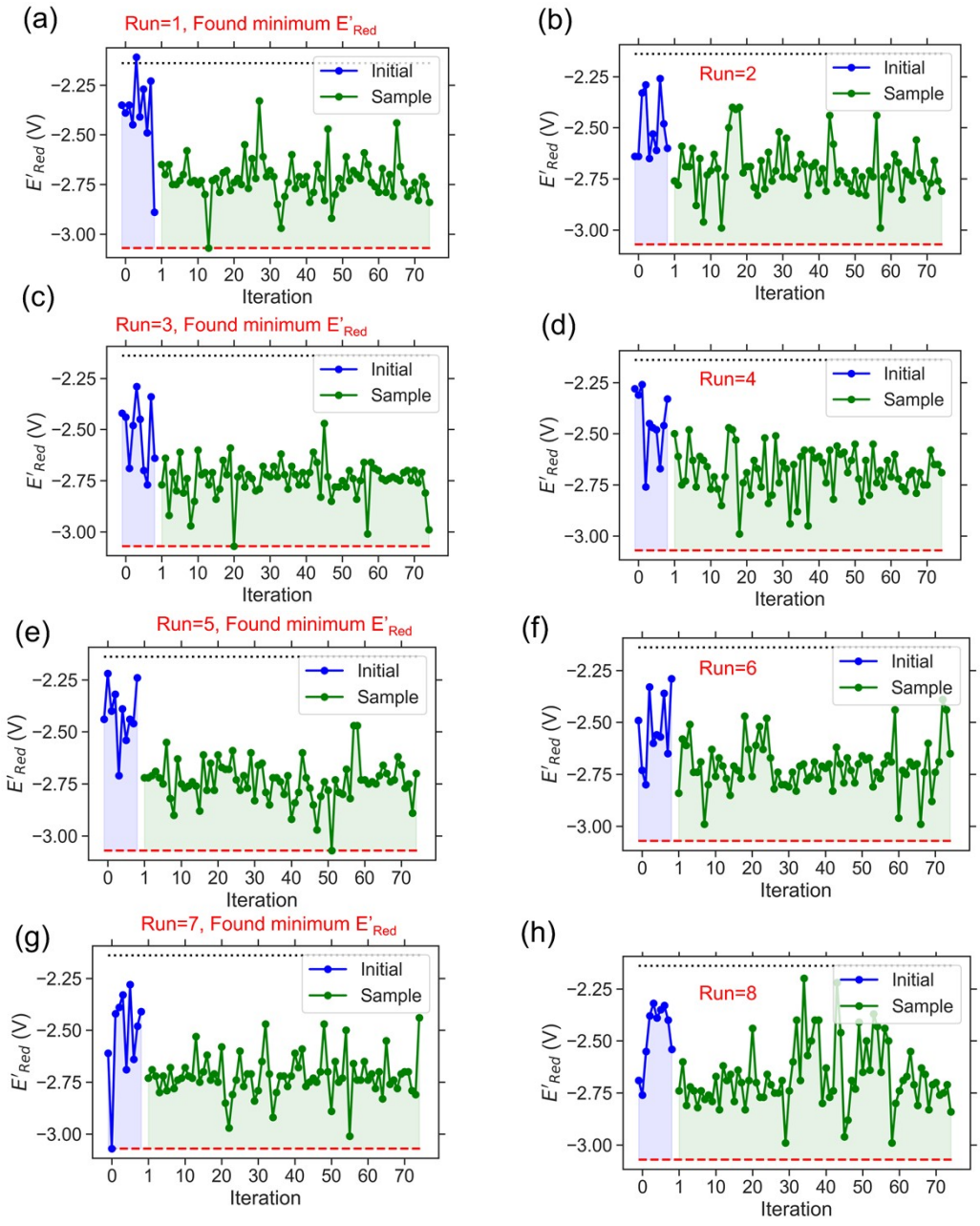


**Figure S5 (part 1 of 3).** Shown in blue is the redox potential $E'_{Red}$ for BTZ molecules in the initial set for GPR training which was drawn randomly from a library of 1,500 computed molecules. Shown in green are the molecules selected by the AL algorithm from the remaining species in this library. The plots show the progression of $E'_{Red}$ with the AL iterations. The algorithm has been repeated 20 times with different initial sets. The black and red dashed horizontal lines correspond to the parent BTZ molecule and the global $E'_{Red}$ minimum, respectively. Part 1 shows run trials 1 to 8. While the global minimum at –3.07 V was not reached in each run, the molecules with $E'_{Red}$ < –3 V were accessed in each run.

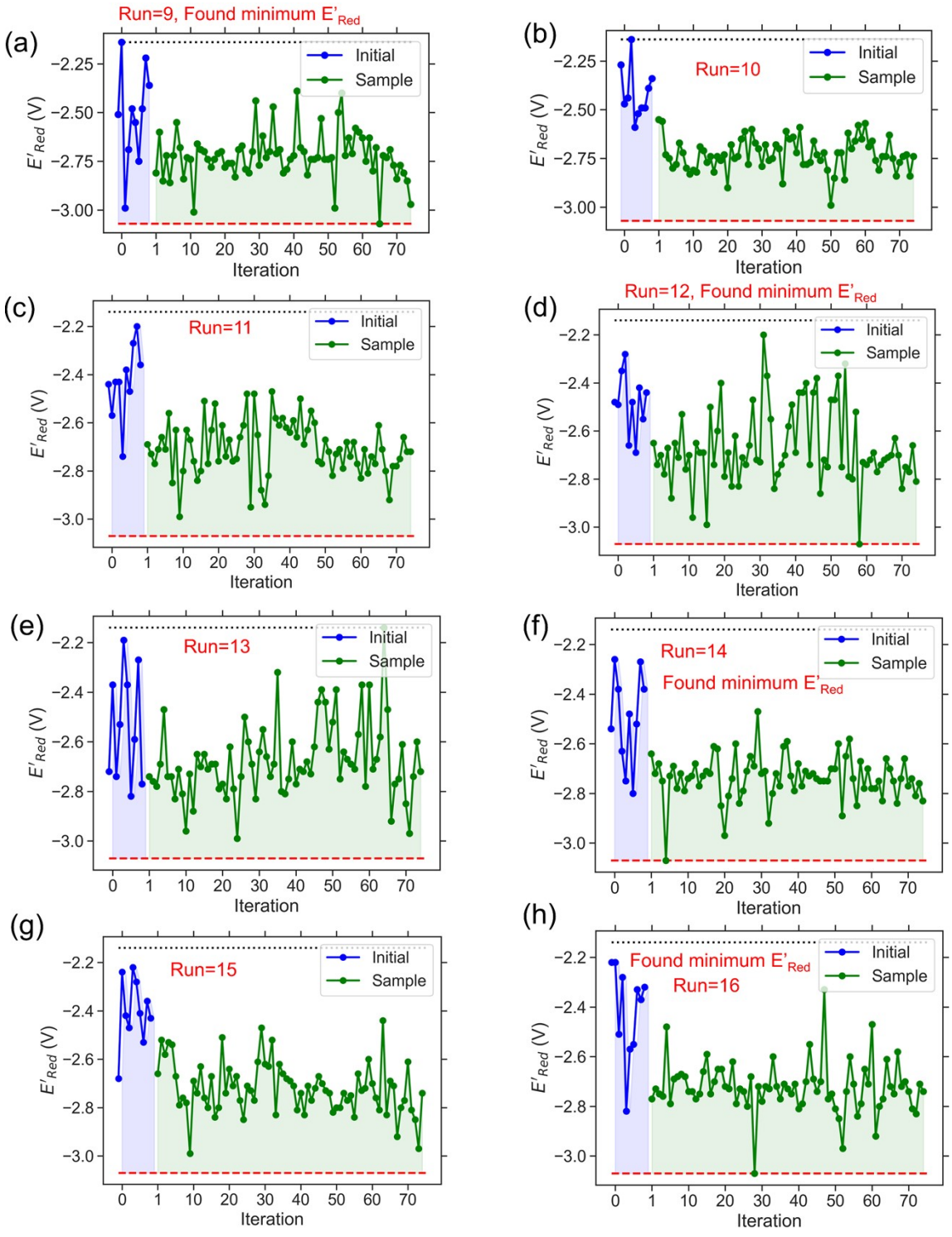


**Figure S5(part 2 of 3).** Figure S5 continued (trial runs 9 to 16).

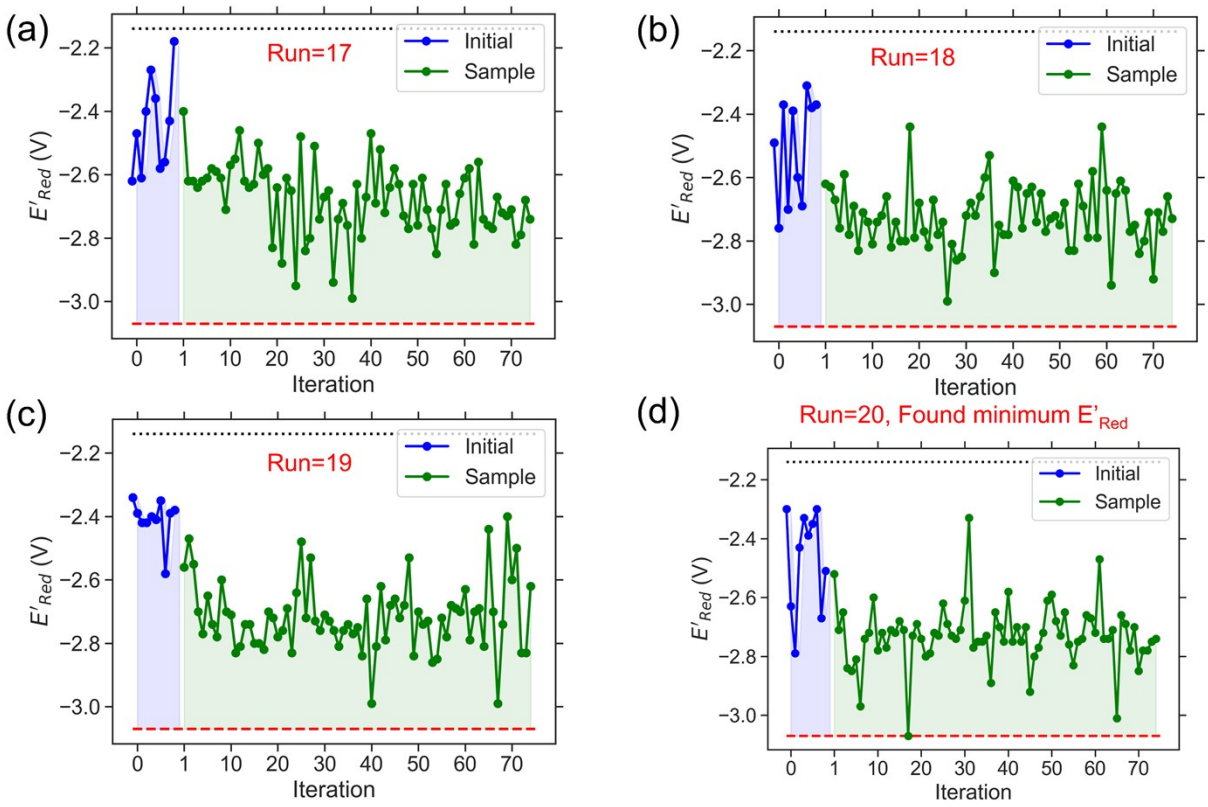


**Figure S5 (part 3 of 3).** Figure S5, continued (trial runs 17 to 20).

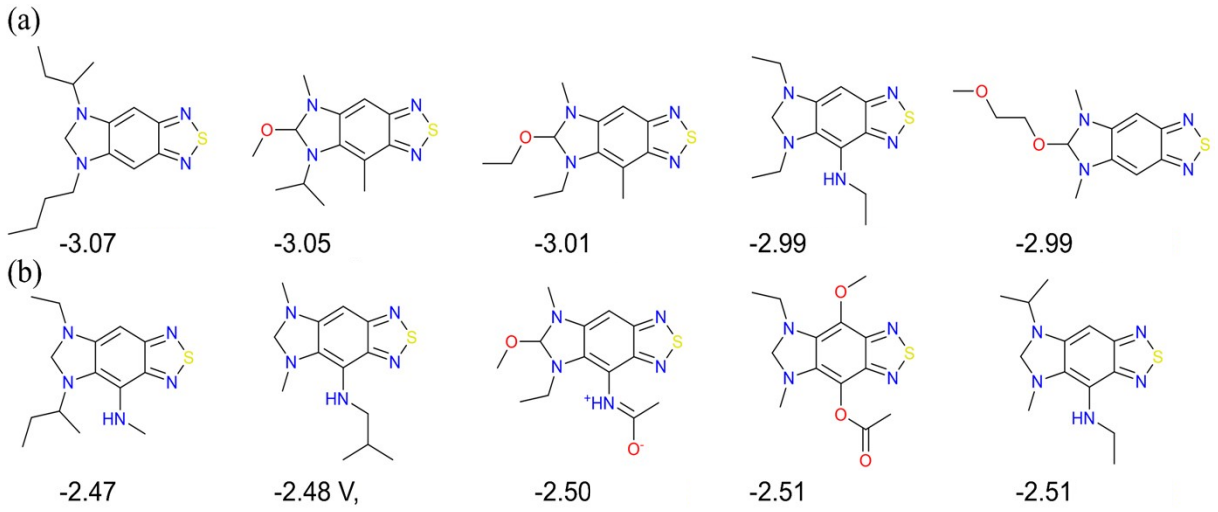


**Figure S6.** Five $S_1$ scaffold derivatives having (a) the most negative and (b) the most positive redox potentials among the computed molecule set. The redox potentials are indicated in the plot. The redox potentials $E'_{\mathrm{Red}}$ in the plot are given in volts vs. $Ag/Ag^{+}$ in acetonitrile. The amide groups were introduced in the zwitterionic form to obtain the correct stereochemistry.

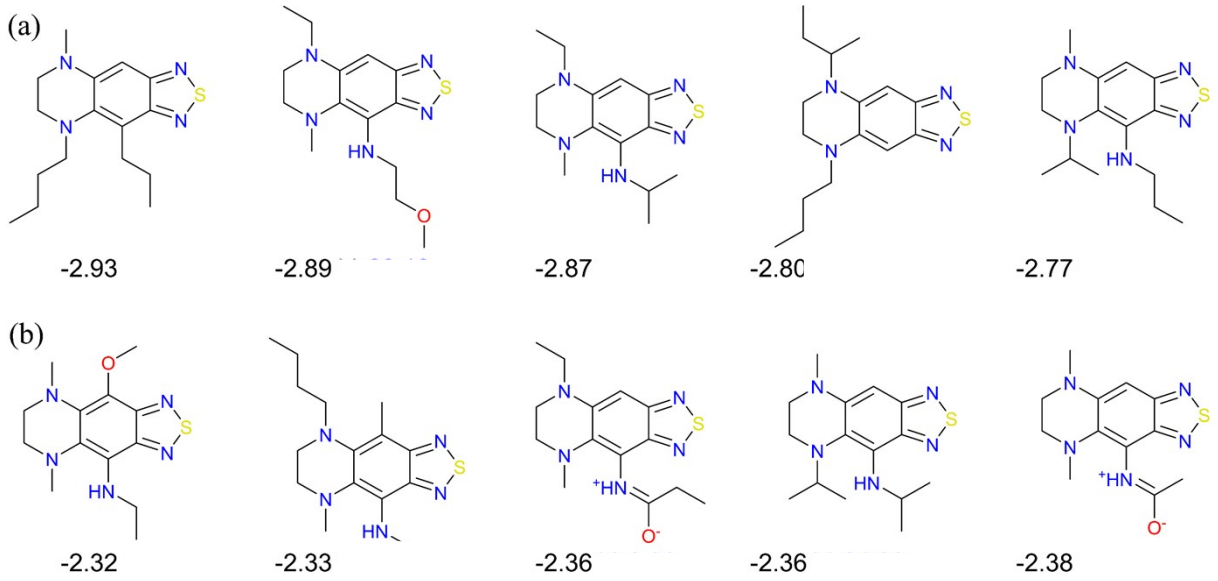


**Figure S7.** Like Figure S5, for the $S_2$ scaffold.

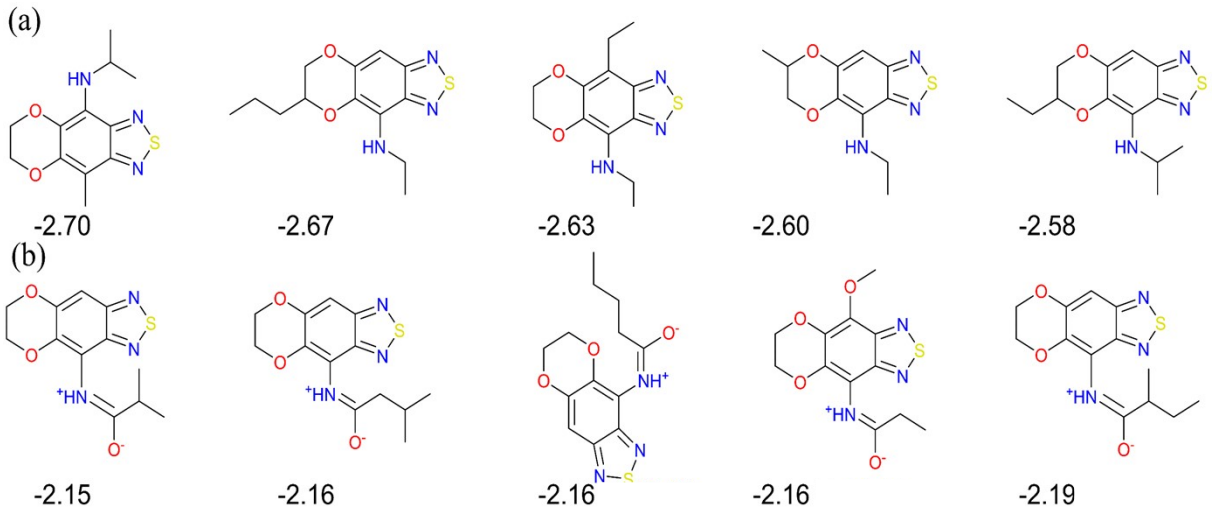


**Figure S8.** Like Figure S5, for the $S_3$ scaffold.

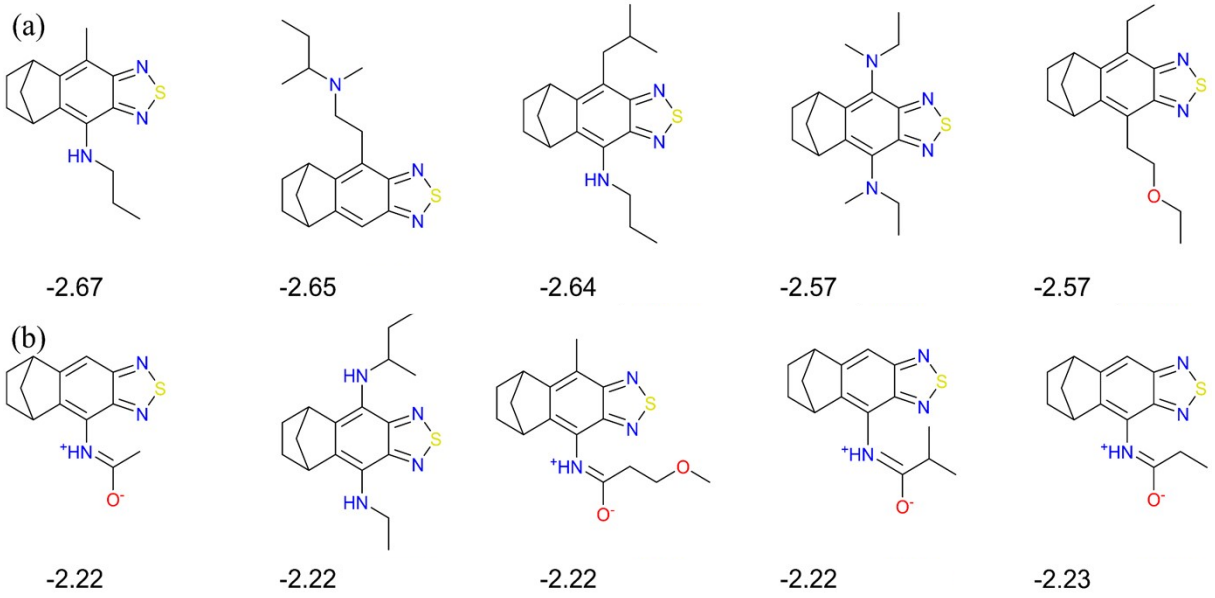


**Figure S9.** Like Figure S5, for the $S_4$ scaffold.

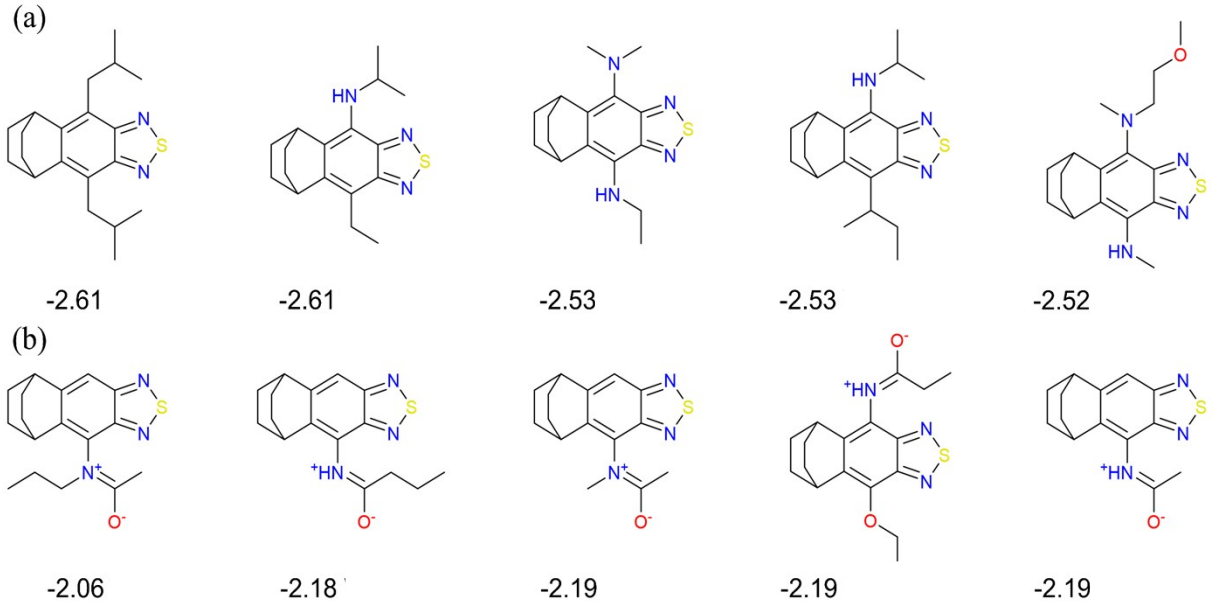


**Figure S10.** Like Figure S5, for the $^{S_5}$ scaffold.

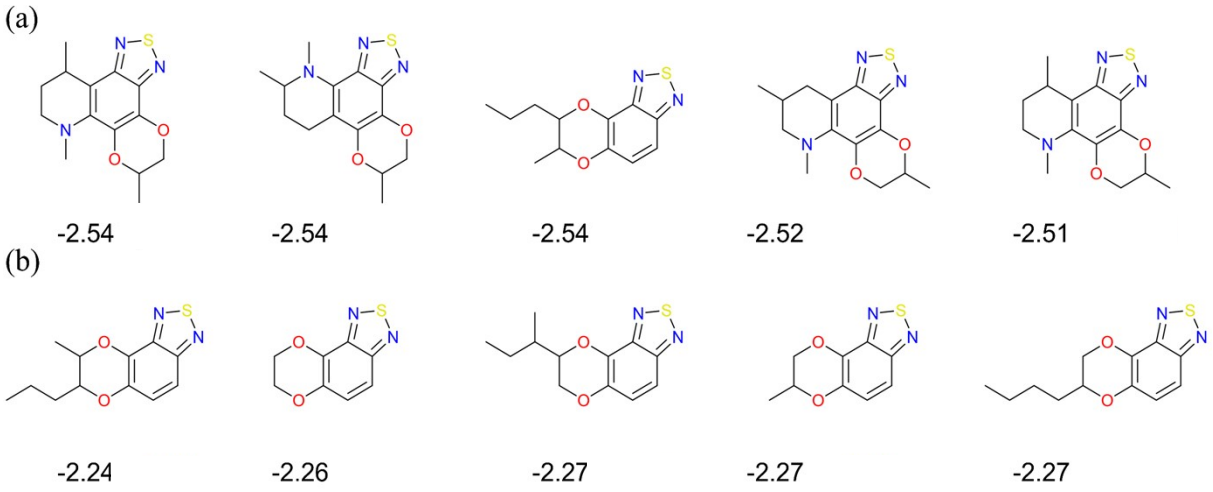


**Figure S11.** Like Figure S5, for the $^{S_6}$ scaffold.

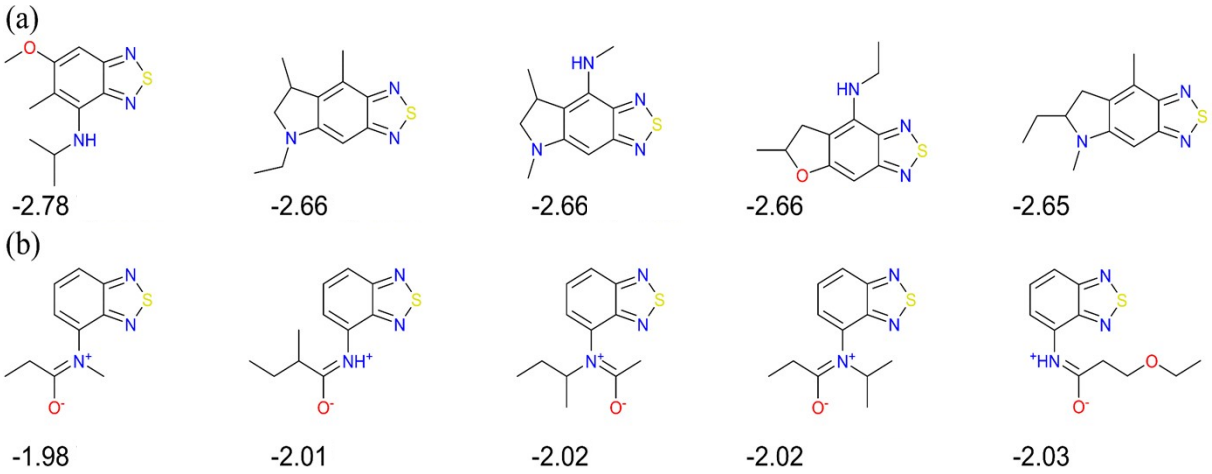


**Figure S12.** Like Figure S5, for the generic (B) scaffold.

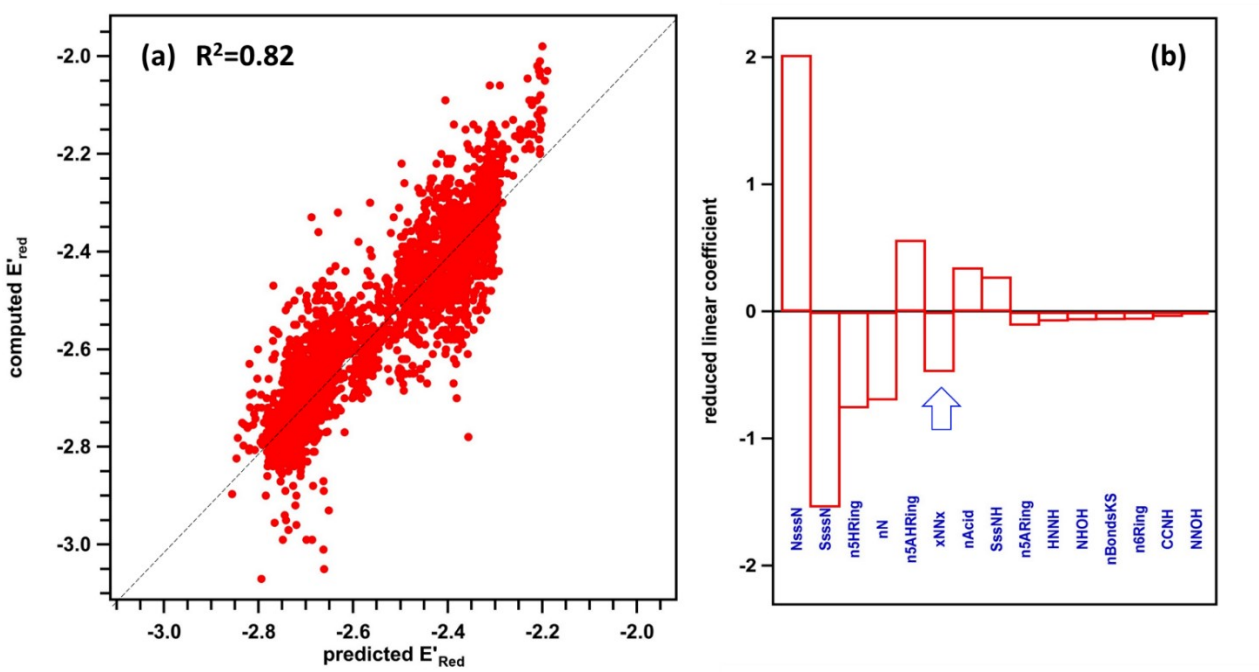


**Figure S13.** Multivariate linear regression (predicted $E'_{Red}$) of the DFT-computed redox potential ($E'_{Red}$) using a set of one-dimensional molecular descriptors from the Mordred package.[1] This standard set was complemented with 4-atom positional binary descriptors (constructed as explained in the text) and the categorical descriptor xNNx indicating the presence of 5,6-diamino groups in a BTZ molecule. (a) The parity plot between the computed and predicted redox potentials using the 15-descriptor set that globally minimizes the deviation between the predicted and computed values. (b) Reduced linear correlation coefficients for these descriptors (all variables are in the standard form, i.e., they are corrected by their averages and normalized by the standard deviations). Larger linear coefficients in this plot correspond to a stronger effect on the redox potential. Among all descriptors added to the standard set, the categorical xNNx descriptor (indicated with the blue arrow in panel b) has the largest effect. The plot indicates that the 5,6-diamino substitution is the most efficient way to reduce the redox potential of a BTZ molecule. The NHxx positional descriptors are also selected (with smaller weights) to add sensitivity to the 4-amino group planarization due to the formation of internal N-H bonds with one of the ring nitrogen atoms.

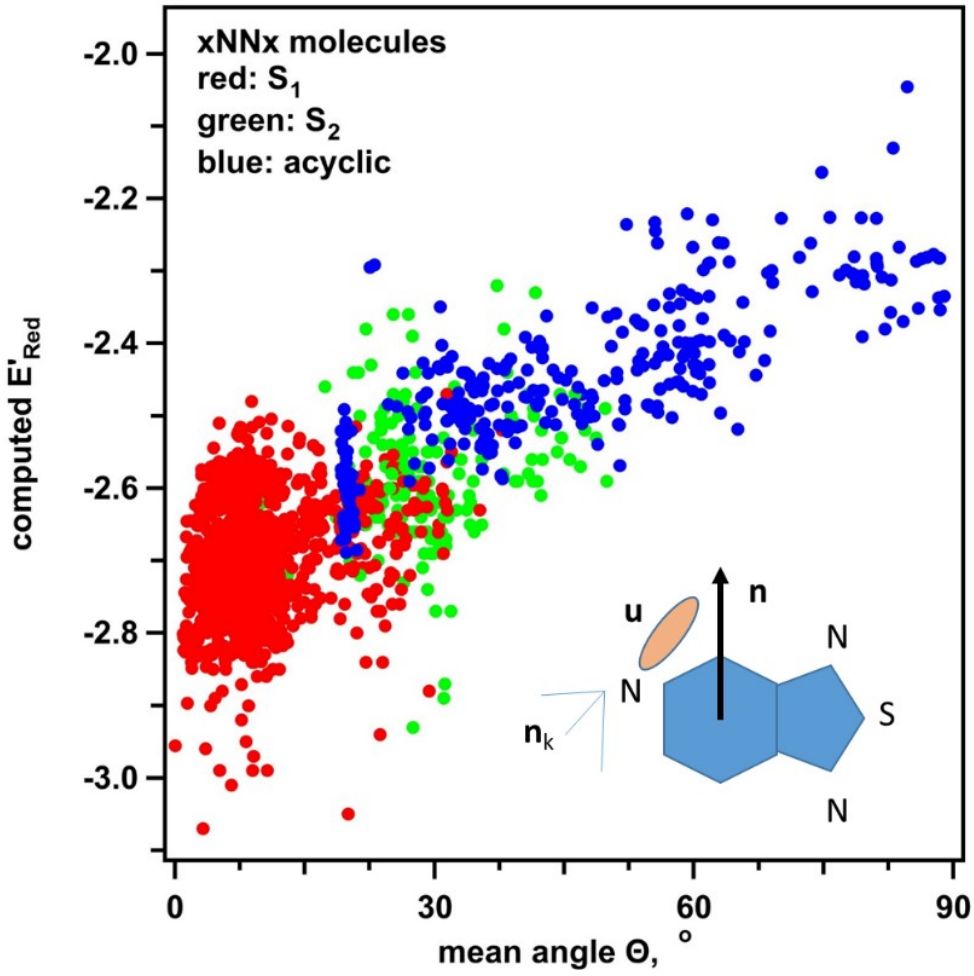


**Figure S14.** The redox potentials for xNNx molecules plotted vs. the mean angle Θ between the nitrogen N 2p lone-pair director **u** and the normal **n** to the plane of the benzene ring in a BTZ molecule (this angle is averaged over both nitrogen atoms). The $S_1$ scaffold molecules are in the red, the $S_2$ scaffold molecules are in the green, and the acyclic molecules are in the blue. The director **u** is defined as the lowest eigenvector of tensor $T_{ab} = \sum_k n_{ka} n_{kb}$ constructed from the unit length bond vectors $\boldsymbol{n}_\mathbf{k}$ pointing from the nitrogen atom. In acyclic molecules, absent steric hindrance between the amino groups, **u** tends to be near the plane of the benzene ring, whereas in the cyclic molecules, it aligns with the plane normal. There is a clear correlation between the redox potential and the orientation of the lone-pair N 2p orbitals. See also Figure S15 below.

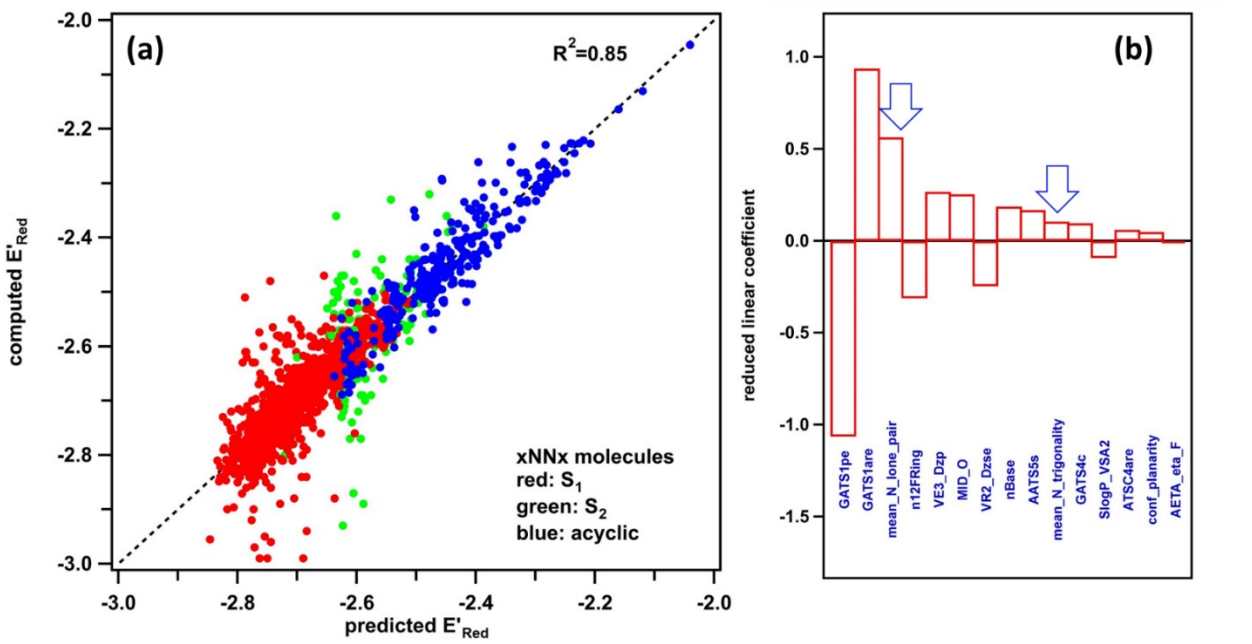


**Figure S15.** (a) The parity plot between the computed and predicted redox potentials for xNNx molecules (the same color scheme is used as in Figure S14). (b) Like Figure S13(b). The arrows indicate (i) the mean angle Θ introduced in the inset of Figure S14 and (ii) the mean trigonality for the N atoms (defined as the deficit from 180° for the improper dihedral angle centered on the nitrogen atom). The planarization of the amino groups is another manifestation of conjugation between the N 2p orbitals and the pi-system of the benzene ring. Other descriptors shown in the plot are the standard 2-dimensional descriptors from the Mordred package.[1] The mean angle Θ is one of the three descriptors most strongly defining the redox potential for the xNNx molecules.

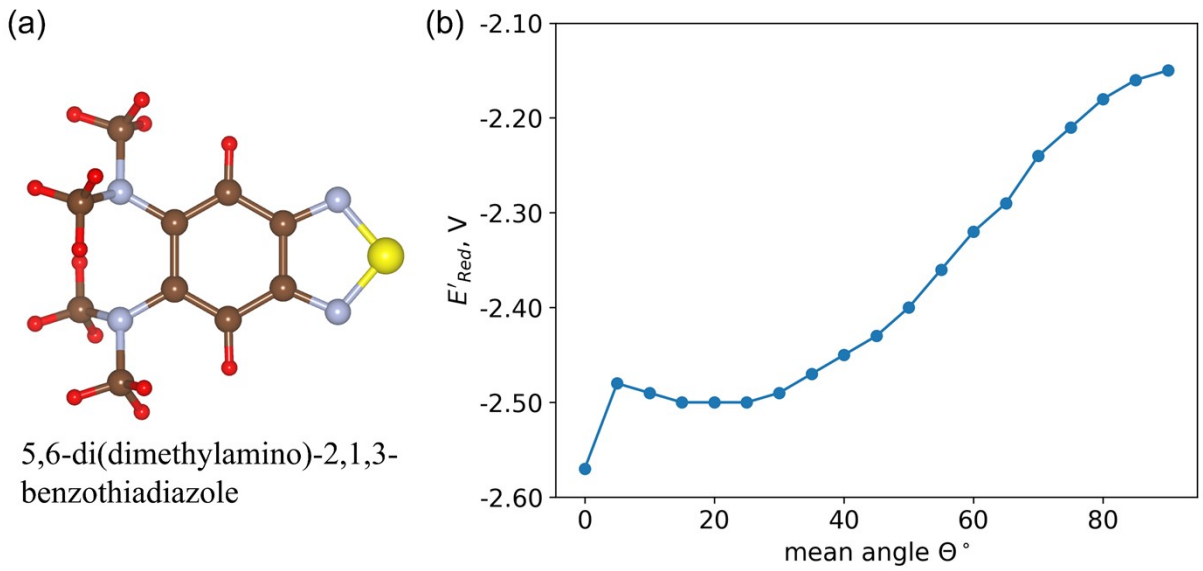


**Figure S16.** (a) The lowest-energy structure of axisymmetric 5,6-di(dimethylamino)-2,1,3-benzothiadiazole. Red, yellow, dark grey, and brown spheres represent the H, S, N, and C atoms, respectively. (b) The reduction potential $E'_{Red}$ plotted as a function of the mean angle Θ between the lone pair orbital of the nitrogen atom and the normal to the plane of the benzene ring (see the caption to Figure S14). To obtain $E'_{Red}$ at various angles Θ in this acyclic molecule, the fixed angles Θ were incremented from 0° to 90° in both states of charge while preserving the axial symmetry. The smaller Θ corresponds to the *N* 2p orbitals normal to the plane of the ring. Panel b shows that Θ and $E'_{Red}$ are positively correlated, that is, the smaller the angle Θ, the greater is the extent of pi-conjugation through the molecule and the lower is the redox potential.

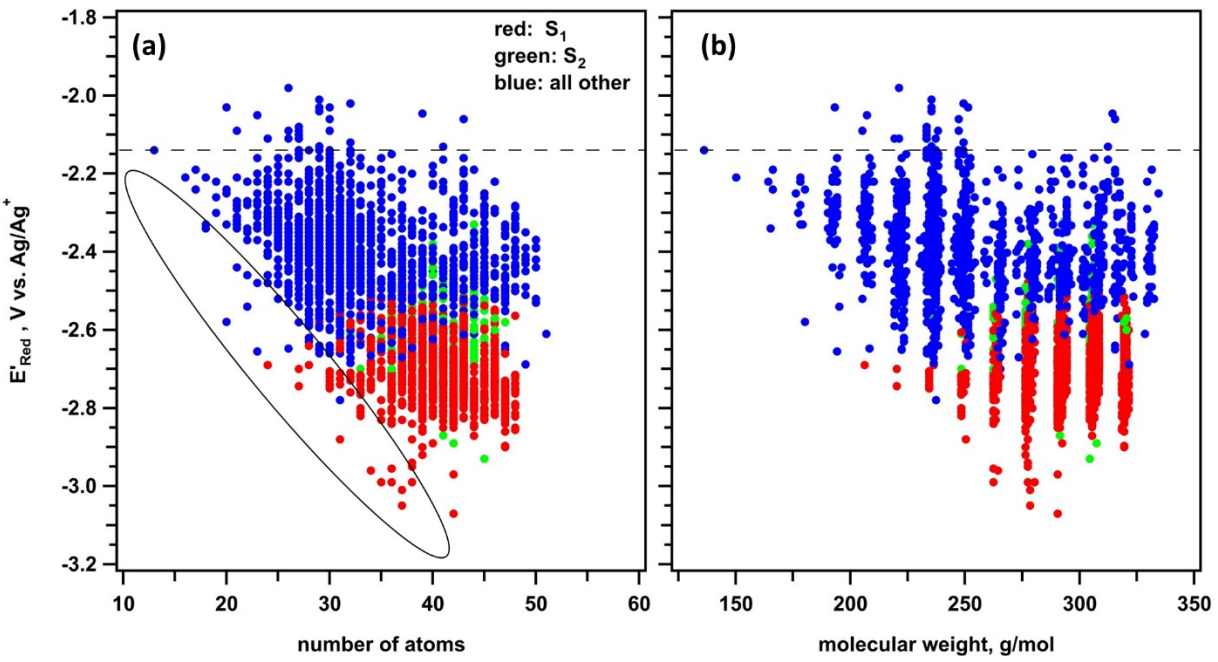


**Figure S17.** The redox potential for all computed molecules plotted vs. (a) the number of atoms and (b) the molecular weight, with the $S_1$ scaffold molecules shown in the red, the $S_2$ scaffold molecules in the green, and all other molecules in the blue. The redox potential for the parent BTZ molecule is shown by the horizontal dashed line. The oval in panel a show the molecules with the lowest redox potential in each category. Compare with Figure S2(c) for a smaller set of molecules that does not include the species discovered through AL. The fifteen molecules shown in Figure 8 in the text have been selected from such outliers.

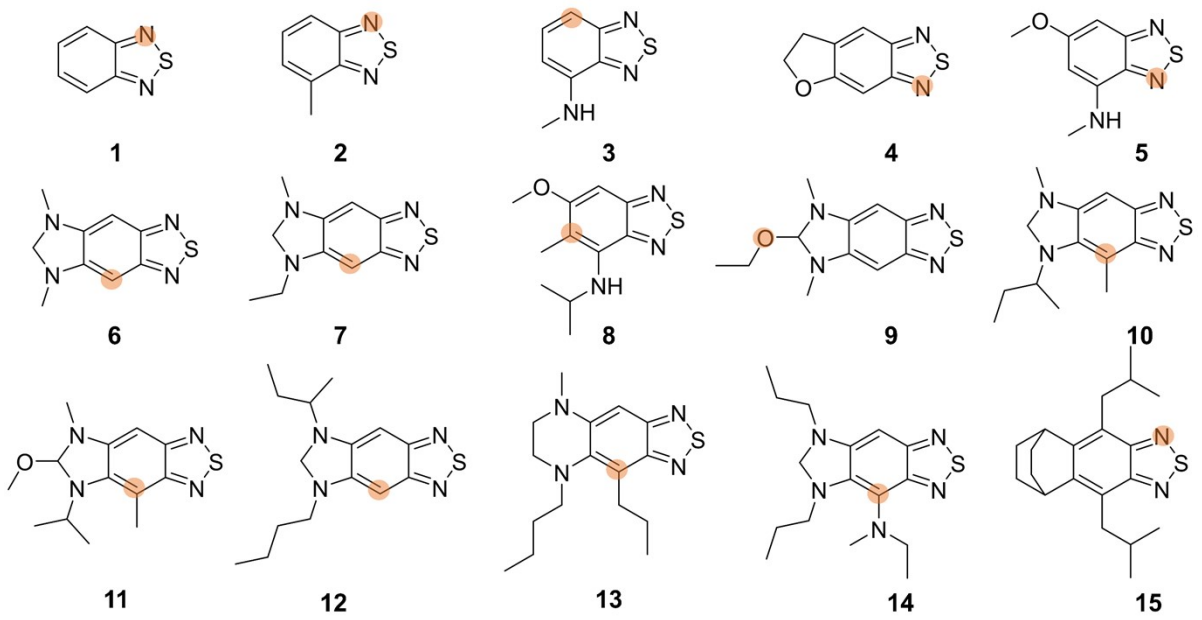


**Figure S18.** Thermodynamically preferable protonation sites in the molecules are shown in Figure 8 in the text. For molecule **11**, the methoxy group leaves when the oxygen becomes protonated.

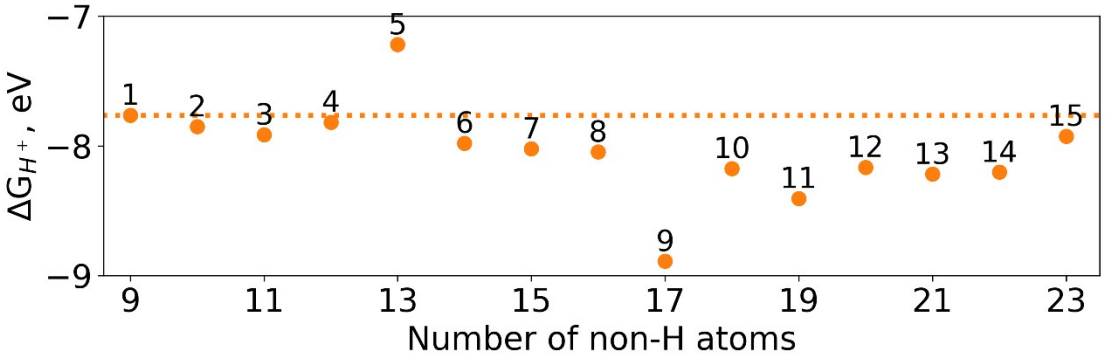


**Figure S19.** The free energy for protonation of the radical anion for molecules shown in Figures 8 and S18. The dashed lines show the computed value for the parent BTZ molecule (**1**). See eq. 4 for the definition of $^{\Delta G}{}_{H^+}$.

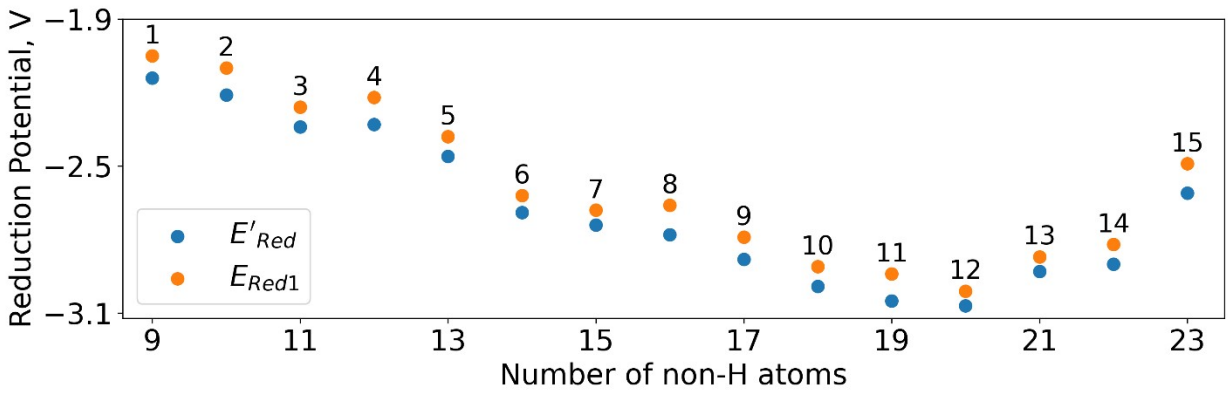


**Figure S20.** The computed 1e⁻ redox potential $E_{Red1}$ and $E_{Red}^{'}$ of 15 molecules shown in Figure 8a. The variance in the difference between $E_{Red1}$ and $E_{Red}^{'}$ (the vibrational and entropy corrections) is small, and $E_{Red1} - E_{Red}^{'} \approx$ 0.1 V.

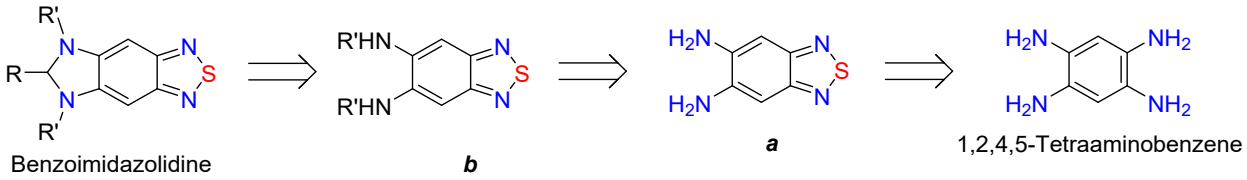


**Figure S21.** Retrosynthetic analysis of the axisymmetric $S_1$ molecules. The benzoimidazolidine is decyclized to *N*-alkyated *ortho*-substituted BTZ ***b***, which is further derived from 5,6-diamino-2,1,3-benzothiadiazole ***a***, which can be prepared from commercially-available 1,2,4,5-tetraaminobenzene.

## S3. Supporting References:


1 H. Moriwaki, Y.-S. Tian, N. Kawashita and T. Takagi, *J. Cheminformatics*, 2018, **10**, 1–14.
2 G. Landrum, *RDKit Open-Source Cheminformatics Httpswwwrdkitorg*.
3 S. A. Wildman and G. M. Crippen, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 868–873.
4 G. Ertl, H. Knözinger, F. Schüth and J. Weitkamp, 2008, **Vol. 1**, 342–356.
5 L. H. Hall and L. B. Kier, *Rev. Comput. Chem.*, 1991, 367–422.
6 P. Labute, *J. Mol. Graph. Model.*, 2000, **18**, 464–477.