Supplementary Information

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

Jingjing Zhang,[1,2] Lily A. Robertson,[*1,2] Ilya A. Shkrob,[*1,2] Rajeev S. Assary,[1,3] and Lu Zhang[1,2]

[1] *Joint Center for Energy Storage Research, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*

[2] *Chemical Sciences and Engineering Division, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439-4837, USA*

[3] *Materials Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439-4837, USA*

*** Corresponding authors:**

Ilya A. Shkrob, shkrob@anl.gov, Tel. 630-252-9516
Lily Robertson, robertla@anl.gov, Tel. 630-252-8956

Table of Contents

Materials and methods

All experiments described here pose little danger when performed safely and do not require any unusual precautions.

Synthesis and electrochemical characterization. The dialkoxyarenes in Figure 2 and Table 1 were synthesized in one pot from their phenol precursors by reaction with the corresponding haloalkanes (Section S1). The compounds were purified by chromatography and ethanol recrystallization, which gave the compounds in 99.0–99.9% purity as determined by chromatography – mass spectrometry (GCMS; Agilent Technologies Model 7890B chromatograph equipped with a Model 5977 mass detector) and high-performance liquid chromatography (HPLC; Agilent 1260 Infinity HPLC G1311C Quat Pump VL, with detectors G1314B Variable Wavelength Detector (VWD) and G7115A Diode Array Detector (DAD) WR) (Section S9 and the figures therein). Some of these molecules readily yield high quality crystals from ethanol, which were analyzed using X-ray diffraction (Section S2 and tables and figures therein). These structural data can be obtained from the Cambridge Crystallographic Data Center base (CCDC). Table S2.1 gives the corresponding entry numbers, space groups, and lattice parameters, while Table S2.2 gives the relevant bond lengths, bond angles, and dihedral angles for neutral dialkoxyarenes in the unit cells.

All electrochemical experiments were carried out at 25 °C in a dry argon glovebox with < 5 ppm O_2/H_2O levels. Cyclic voltammetry (see Section S3 and figures therein) was performed in 10 mM redoxmer in 0.5 M lithium bistriflimide (LiTFSI) in acetonitrile solutions using a CHI760D electrochemical workstation (CH Instruments, TX). The estimated half-wave redox potentials $E_{1/2}$ vs Ag/Ag^+ are given in Table 1, and Table S3.1 gives the peak anodic and cathodic current ratios. The latter were all > 0.92, suggesting electrochemical reversibility of the one-electron oxidation.

Cycle life testing was performed using a Parstat MC potentiostat (Princeton Applied Research). For each test (see Section S4 and figures therein), 5 mL of 20 mM redoxmer in 0.5 M LiTFSI solution was placed in both compartments of an H-cell equipped with a thick ceramic separator. Reticulated vitreous carbon was used as the working and counter electrodes, and the quasi-reference electrode was $Ag/AgNO_3$ (10 mM $AgNO_3$ in acetonitrile, placed as close as possible to the working electrode). The theoretical capacity was calculated from the known redoxmer concentration and referred to as 100% state-of-charge (SOC). Each charge/discharge cycle was reversed when (a) the upper/lower cutoff potential was reached (which were 1.8 and 0 V vs $Ag/AgNO_3$, respectively) or (b) the set capacity of 50% SOC was reached. Since in our experiments the capacity is set to 50% SOC, to terminate cycling due to depletion of the redoxmer, > 50% of the initial material needs to be consumed. The charge/discharge current was set at 3C, at which rate the cells reach full theoretical capacity (C) in 1/3 h. The products of electrolysis were analyzed as described in Section S7 and figures therein.

The same setup was used to oxidize molecules for calendar life evaluations (Section S5 and tables and figures therein). To this end, 5–20 mM of a redoxmer was charged at a fast rate of 5C to 100% SOC. These solutions were diluted 1:10 *v*/*v* to obtain well-resolved electron paramagnetic resonance (EPR) spectra of the radical cations, from which the isotropic hyperfine coupling constants given in Table S5.1 were estimated. The instrumentation was a continuous-wave X band (9.43 GHz) Bruker ESP300E EPR spectrometer. EPR spectroscopy was also used to follow the decay of the radical cation in the electrolyte solution over several days. The integrated EPR signal (which is proportional to the concentration of the radical cation) was biexponentially fit. The fast component corresponds to a reaction of the radical cation with impurities in the electrolyte, whereas the slow component reflects the "natural" reactivity (see, for example, Figure S5.1). The exponential half-life times for these slow components ($t_{1/2}$) were taken as a metric of calendar lifetime (Table 1).

Computational characterization. Molecular structures were preoptimized using semiempirical methods. For type-2 molecules in Figure 2, C_i symmetry was imposed in both states of charge. To calculate optimized geometries, energetics, and proton hyperfine coupling constants, density functional theory (DFT) calculations using B3LYP functional[1, 2] with 6-31G(d,p) basis set as implemented in Gaussian 09[3] was used with CPCM polarized continuum model of the solvent.[4, 5] The redox potential vs Li/Li$^+$ was estimated as described in refs.[6, 7] The optimized geometries and parameters for neutral and charged molecules obtained in this way were used to compute molecular descriptors for ML. The GA-MMLR procedures are described in Section S6 and Figure 1.

Generic descriptors. The Python-based RDKit[8, 9] with MMFF94 force field[10] was used to calculate atomic properties and topological indices for the neutral redoxmer molecules in families **1** and **2** shown in Figure 2 and Table 1. The atomic coordinates were fed into a Mordred calculator[11] that generated 1,200+ generic descriptors per neutral molecule. To reduce the number of descriptors, we excluded descriptors that had the absolute Pearson's correlation coefficients > 0.9 with each other and used interquartile range test to exclude excessive clustering. This reduction left 300–350 descriptors that were used for ML analyses described below. An example descriptor is a Moreau-Broto lag-*g* spatial autocorrelator $\sum_{ij}^{n} \delta_{ij}\, w_i w_j$ of atomic properties w_j (that can be a mass, charge, electronegativity, polarizability, etc.) calculated over all atomic pairs separated by *g* bonds.[12] Other descriptors used Moran's I[13] or Geary's C[14] autocorrelators or still more complex quantities. See Section S6 for some addition molecular shape descriptors used.

Shape characterizing ML descriptors. To characterize shape asymmetry, the gyration tensor was defined as

$$I_{ab} = \sum_i w_i \delta r_{ia} \delta r_{ib} / \sum_i w_i, \qquad \text{(M1)}$$

where δr_{ia} is the distance relative to the centroid $\delta r_{ia} = r_{ia} - \bar{r}_a$,

$$\bar{r}_a = \sum_i w_i r_{ia} / \sum_i w_i, \quad \text{(M2)}$$

w_i is the weight (which was either 1 for all atoms or their *Z*-number) , and *a*=*x*,*y*,*z*; the summation is over all atoms in the molecule

The eigenvalues I_i of this tensor (*i*=1, 2, 3) were ordered so that $I_1 \leq I_2 \leq I_3$. The gyration radius

$$r_g = (2\pi I_1 I_2 I_3)^{1/3}, \quad \text{(M3)}$$

and the following quantities were included as descriptors: ratios $\eta_{1,2} = I_{1,2}/I_3$, eccentricity $\varepsilon = (1 - \eta_1^2)^{1/2}$, asphericity

$$a = \tfrac{1}{2}((I_3 - I_2)^2 + (I_2 - I_1)^2 + (I_1 - I_3)^2)/(I_1^2 + I_2^2 + I_3^2), \quad \text{(M4)}$$

and spherocity

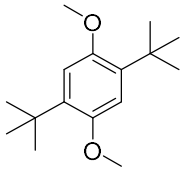
$$s = 3I_1/(I_1 + I_2 + I_3). \quad \text{(M5)}$$

S1. Synthetic procedures

The following chemicals were purchased from Sigma Aldrich and used as received: 2,5-di-*tert*-butylbenzene-1,4-diol, 2,5-di-*tert*-butyl-4-methoxyphenol, iodomethane, bromoethane, 1-bromopropane, 2-bromopropane, 2-bromobutane, benzyl bromide, and cesium carbonate. Anhydrous *N*,*N*-dimethylformamide (DMF) was obtained from VWR. All synthetic manipulations were performed under argon atmosphere with magnetic stirring. Flash chromatography using silica gel (230–400 mesh) as the stationary phase was used to isolate the desired products. Proton (^{1}H) and carbon-13 (^{13}C) nuclear magnetic resonance (NMR) spectra shown in Section S8 were acquired in $CDCl_3$ using a Bruker Avance DPX-300 spectrometer. Chemical shifts in parts per million (ppm) are reported relative to the residual solvent peak. The purity (> 99.0–99.9%) was established using ^{1}H NMR and GCMS as well as ^{13}C NMR (Section S8) and HPLC (Section S9) for compounds previously unreported (**MeEt**, **DiEt**, **MePr**, **DiPr**, **MeiPr**, **DiiPr**, **MesBu**, **DisBu**, **MeBn**, and **DiBn**).

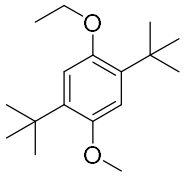
Non-symmetric dialkoxyarenes (procedure 1). 2,5-Di-*tert*-butyl-4-methoxyphenol (5 g, 21.1 mmol, 1.0 equiv.) was dissolved in DMF (80 mL) at room temperature followed by the addition of Cs_2CO_3 (11 g, 33.8 mmol, 1.6 equiv). The reaction mixture was stirred for 2 h, and the haloalkane (63.3 mmol, 3.0 equiv) was added dropwise. The reaction mixture was heated to 80 °C for 8 h, cooled room temperature, and extracted with CH_2Cl_2 and water. The organic phase was washed with water (3x), dried ($MgSO_4$), filtered, and the solvent removed by rotary evaporation. The crude product was purified via flash chromatography on silica gel using CH_2Cl_2: hexanes mixture (1:1, *v/v*) and recrystallized from ethanol.

Symmetric dialkoxyarenes (procedure 2). 2,5-di-*tert*-butylbenzene-1,4-diol (5 g, 22.5 mmol, 1.0 equiv) was dissolved in DMF (80 mL) at room temperature followed by the addition of Cs_2CO_3 (22 g, 67.6 mmol, 3.0 equiv.). The reaction mixture was stirred for 2 h, and the haloalkane (90.0 mmol, 4 equiv.) was added dropwise. Workup and purification were performed analogously to that of the non-symmetric dialkoxyarenes.



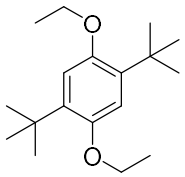
DiMe

1,4-Bis(1,1-dimethylethyl)-2,5-dimethoxybenzene, DiMe. Prepared as reported in ref.[15] in 70% yield. ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.85 (s, 2H), 3.82 (s, 6H), 1.38 (s, 18H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 152.11, 136.46, 111.83, 56.08, 34.74, 29.94.



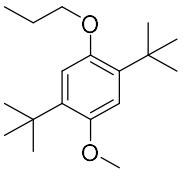
MeEt

1,4-Bis(1,1-dimethylethyl)-2-ethoxy-5-methoxybenene, MeEt. Synthesized by procedure 1 using 2,5-di-*tert*-butyl-4-methoxyphenol (5 g, 21.1 mmol), bromoethane (6.9 g, 63.3 mmol), Cs_2CO_3 (11 g, 33.8 mmol), and DMF (80 mL) to give the desired product as a white solid (4.1 g, 73%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.85 (s, 1H), 6.83 (s, 1H), 4.04 (q, J = 7.0 Hz, 2H), 3.82 (s, 3H), 1.45 (t, J = 7.0 Hz, 3H), 1.40 (s, 9H), 1.37 (s, 9H). ^{13}C NMR (75 MHz, $CDCl_3$): 151.94, 151.37, 136.34, 136.31, 112.21, 111.87, 64.26, 56.08, 34.79, 34.71, 29.98, 29.95, 15.38.



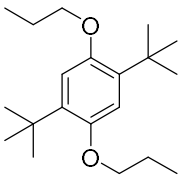
DiEt

1,4-Bis(1,1-dimethylethyl)-2,5-diethoxybenzene, DiEt. Synthesized by procedure 2 using 2,5-di-*tert*-butylbenzene-1,4-diol (5 g, 22.5 mmol), bromoethane (9.8 g, 90 mmol), Cs_2CO_3 (22 g, 67.6 mmol), and DMF (80 mL) to give the desired product as a white solid (4.3 g, 69%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.83 (s, 2H), 4.03 (q, J = 7.0 Hz, 4H), 1.44 (t, J = 7.0 Hz, 6H), 1.38 (s, 18H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 151.21, 136.19, 112.26, 64.26, 34.76, 29.99, 15.38.



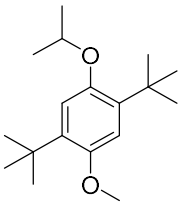
MePr

1,4-Bis(1,1-dimethylethyl)-2-methoxy-5-propoxy-benzene, MePr. Synthesized by procedure 1 using 2,5-di-*tert*-butyl-4-methoxyphenol (5 g, 21.1 mmol), 1-bromopropane 7.8 g (63.3 mmol), Cs_2CO_3 (11 g, 33.8 mmol), and DMF (80 mL) to give the desired product as a white solid (3.5 g, 60%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.85 (s, 1H), 6.82 (s, 1H), 3.93 (t, J = 6.3 Hz, 2H), 3.82 (s, 3H), 1.92–1.82 (m, 2H), 1.39 (s, 9H), 1.37 (s, 9H), 1.09 (t, J = 7.4 Hz, 3H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 151.81, 151.45, 136.30, 136.06, 112.01, 111.64, 70.11, 56.13, 34.81, 34.72, 29.99, 29.95, 23.24, 11.29.



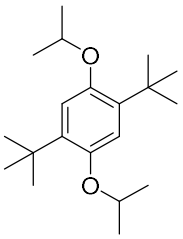
DiPr

1,4-Bis(1,1-dimethylethyl)-2,5-dipropoxy-benzene, DiPr. Synthesized by procedure 2 using 2,5-di-*tert*-butylbenzene-1,4-diol (5 g, 22.5 mmol), 1-bromopropane (11.1 g, 90 mmol), Cs_2CO_3 (22 g, 67.6 mmol), and DMF (80 mL) to give the desired product as a white solid (4.2 g, 61%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.82 (s, 2H), 3.92 (td, J = 6.8, 5.8, 2.6 Hz, 4H), 1.89–1.82 (m, 4H), 1.39 (s, 18H), 1.12–1.06 (m, 6H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 151.16, 135.92, 111.82, 70.15, 34.80, 30.00, 23.25, 11.28



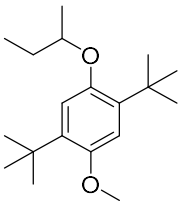
MeiPr

1,4-Bis(1,1-dimethylethyl)-2-methoxy-5-(1-methylethoxy)-benzene, MeiPr. Synthesized by procedure 1 using 2,5-di-*tert*-butyl-4-methoxyphenol (5 g, 21.1 mmol), 2-bromopropane (7.8 g, 63.3 mmol), Cs_2CO_3 (11 g, 33.8 mmol), and DMF (80 mL) to give the desired product as a white solid (3.7 g, 63%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.86 (s, 1H), 6.81 (s, 1H), 4.66–4.59 (m, 1H), 3.82 (s, 3H), 1.40 (s, 9H), 1.37 (s, 9H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 151.39, 149.51, 136.29, 136.13, 112.20, 111.71, 68.74, 56.08, 34.80, 34.71, 30.00, 22.34.



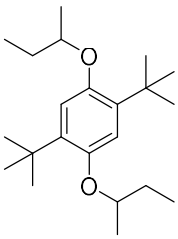
DiiPr

1,4-Bis(1,1-dimethylethyl)-2,5-bis(1-methylethoxy)-benzene, DiiPr Synthesized by procedure 2 using 2,5-di-*tert*-butylbenzene-1,4-diol (5 g, 22.5 mmol), 2-bromopropane (11.1 g, 90 mmol), Cs_2CO_3 (22 g, 67.6 mmol), and DMF (80 mL) to give the desired product as a white solid (4.9 g, 57%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.81 (s, 2H), 4.65–4.55 (m, 2H), 1.38 (s, 18H), 1.37–1.35 (m, 6H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 148.76, 135.89, 111.99, 68.66, 34.76, 30.04, 22.38.



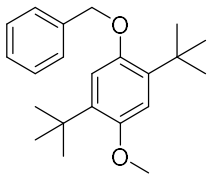
MesBu

1,4-bis(1,1-dimethylethyl)-2-methoxy-5-(1-methylpropoxy)-benzene, MesBu. Synthesized by procedure 1 using 2,5-di-*tert*-butyl-4-methoxyphenol (5 g, 21.1 mmol), 2-bromobutane (8.7 g, 63.3 mmol), Cs_2CO_3 (11 g, 33.8 mmol), and DMF (80 mL) to give the desired product as a white solid (4.2 g, 68%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.87 (s, 1H), 6.79 (s, 1H), 4.42 (h, J = 5.9 Hz, 1H), 3.82 (s, 3H), 1.88–1.68 (m, 2H), 1.40 (s, 9H), 1.38 (s, 9H), 1.04 (t, J = 7.5 Hz, 3H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 151.13, 149.40, 136.03, 135.93, 112.14, 111.25, 73.30, 55.97, 34.70, 34.58, 29.88, 29.87, 29.40, 18.95, 10.08.



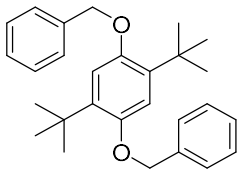
DisBu

1,4-Bis(1,1-dimethylethyl)-2,5-bis(1-methylpropoxy)-benzene DisBu. Synthesized by procedure 2 using 2,5-di-*tert*-butylbenzene-1,4-diol (5 g, 22.5 mmol), 2-bromobutane (12.3 g, 90 mmol), Cs_2CO_3 (22 g, 67.6 mmol), and DMF (80 mL) to give the desired product as a white solid (4.1 g, 54%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.80 (s, 2H), 4.40 (h, J = 6.0 Hz, 2H), 1.89–1.59 (m, 4H), 1.38 (s, 18H), 1.28 (dd, J = 6.0, 2.9 Hz, 6H), 1.03 (td, J = 7.5, 2.3 Hz, 6H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 148.66, 135.73, 135.68, 111.76,111.68, 73.33, 34.82, 30.07, 29.62, 29.52, 19.17, 19.10, 10.24, 10.20.



MeBn

1,4-Bis(1,1-dimethylethyl)-2-methoxy-5-(phenylmethoxy)-benzene MeBn. Synthesized by procedure 1 using 2,5-di-*tert*-butyl-4-methoxyphenol (5 g, 21.1 mmol), (bromomethyl)benzene (10.8 g, 63.3 mmol), Cs_2CO_3 (11 g, 33.8 mmol), and DMF (80 mL) to give the desired product as a white solid (3.9 g, 57%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.51–7.33 (m, 5H), 6.90 (s, 1H), 6.88 (s, 1H), 5.09 (s, 2H), 3.83 (s, 3H), 1.41 (s, 9H), 1.35 (s, 9H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 152.20, 151.09, 138.13, 136.55, 136.38, 128.59, 127.69, 127.45, 112.65, 111.89, 71.01, 56.04, 34.84, 34.74, 30.04, 29.90.



DiBn

1,4-Bis(1,1-dimethylethyl)-2,5-bis(phenylmethoxy)benzene DiBn. Synthesized by procedure 2 using 2,5-di-*tert*-butylbenzene-1,4-diol (5 g, 22.5 mmol), (bromomethyl)benzene (15.4 g, 90 mmol), Cs_2CO_3 (22 g, 67.6 mmol), and DMF (80 mL) to give the desired product as a white solid (4.6 g, 51%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.52–7.32 (m, 10H), 6.95 (s, 2H), 5.12 (s, 4H), 1.40 (s, 18H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 151.20, 138.09, 136.48, 128.60, 127.70, 127.45, 112.74, 70.97, 34.84, 30.01.

S2. X-ray crystallography

Low-temperature diffraction data were collected using a Bruker SMART APEX II diffractometer coupled to an APEX II CCD detector and with Mo Kα radiation (λ = 0.71073 Å). The Bruker suite of programs on the APEX II was used to integrate the data (Bruker AXS Inc., Madison, WI, USA, 2007.65) and SADABS was used for absorption correction (Bruker AXS Inc., Madison, WI, USA, 2001) The structure was solved by direct methods and refined using the SHELXTL. All non-hydrogen atoms were refined with anisotropic thermal parameters, and the refinements converged for $I > 2\sigma(I)$. The structures were deposited with the Cambridge Crystallographic Data Centre (CCDC), and the corresponding indices are given in **Table S2.1**. Molecular geometries in the crystals are shown in **Figures S2.1 and S2.2**, while selected bond lengths, bond angles, and torsion angles are given in **Table S2.2.**

Table S2.1. Selected unit cell parameters for dialkoxyarenes.

Molecule	CCDC database index	RO group (1)	R'O group (2)	Space group	*Z*	*a*, Å	*b*, Å	*c*, Å	α, °	β, °	γ, °	Volume, Å
DiMe	645127 [b]	Me	Me	P 21/c	2	10.621	6.786	11.033	90	105.688	90	765.6
DiEt	1938227 [c]	Et	Et	P 21/c	2	6.342	8.038	16.689	90	99.68	90	838.7
MeiPr	1938230 [c]	Me	*iso*-Pr	P 21/n	4	11.211	10.187	15.053	89.98	90.83	90.01	1718.9
DiiPr	1938228 [c]	*iso*-Pr	*iso*-Pr	P 1	2	8.9774	12.2155	9.435	90	110.9303	90	966.4
DisBu	1938229 [c]	*sec*-Bu	*sec*-Bu	P 21/n	2	10.5235	9.5398	11.6534	90	113.901	90	1069.6
DiBn	1938226 [c]	Ph	Ph	P 21/c	2	11.581	10.120	11.349	90	113.929	90	1215.8

Table S2.2. Comparison of the computed and crystallographic geometry parameters for neutral dialkoxyarenes.

Molecule	CCDC database index	B3LYP/6-31+G(d,p) optimized geometry			Estimated from crystallographic data		
		C*-O-Cα-Hα	C*-O-Cα-Cβ	C*-O-Cα-Hα	C*-O-Cα-Hα	C*-O-Cα-Cβ	C*-O-Cα-Hα
		RO	RO	MeO	RO	RO	MeO
DiMe	645127 [b]	180.00		180	169.20		
DiEt [d]	1938227 [c]	45.20	80.80		58.12	178.20	
MeiPr	1938230[c]	45.84	162.19	179.19	24.82	147.57	178.70
DiiPr	1938228 [c]	44.85	161.23		55.72	173.15	
DisBu	1938229 [c]	43.32	159.04		41.46	171.53	
DiBn	1938226 [c]	59.51	179.96		59.32	179.27	
Molecule	**CCDC database index**	**C*-O-Cα**	**C*-O, O-Cα**	**C*-O-Cα**	**C*-O-Cα**	**C*-O, O-Cα**	**C*-O-Cα**
		RO	**RO**	**MeO**	**RO**	**RO**	**MeO**
DiMe	645127 [b]	119.36	1.38, 1.43		117.78	1.38, 1.42	
DiEt	1938227 [c]	121.21	1.38, 1.44		118.39	1.37, 1.42	
MeiPr	1938230[c]	121.76	1.38, 1.45	119.32	115.32	1.40, 1.47	116.63
DiiPr	1938228 [c]	121.70	1.38, 1.45		119.82	1.38, 1.44	
DisBu	1938229 [c]	121.84	1.38, 1.45		121.96	1.38, 1.41	
DiBn	1938226 [c]	119.24	1.38, 1.44		119.21	1.38, 1.43	

a) The angles are in degree and the bond lengths are in Å. C^* is the ring carbon at the *tert*-butyl position, C_α is the nearest carbon in the alkoxy group, and C_β is the second nearest carbon in this group.
b) From [16]
c) This study
d) Positional uncertainty due to a large thermal ellipsoid for ethoxy groups.

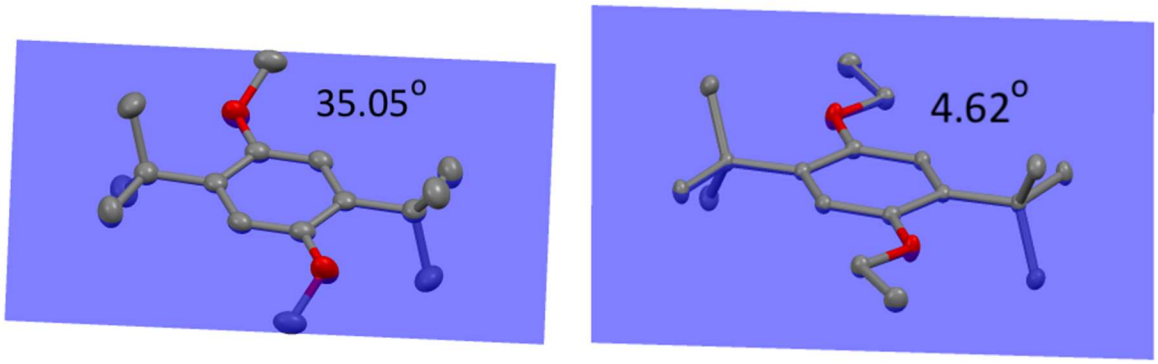


Figure S2.1. X-ray crystal structures for **DiMe** (left) and **DiEt** (right). The torsion angles given in the plot are between the planes of the C-O-C$_\alpha$ and the arene ring. Thermal ellipsoids for selected carbon atoms are shown. Hydrogen atoms are omitted for clarity.

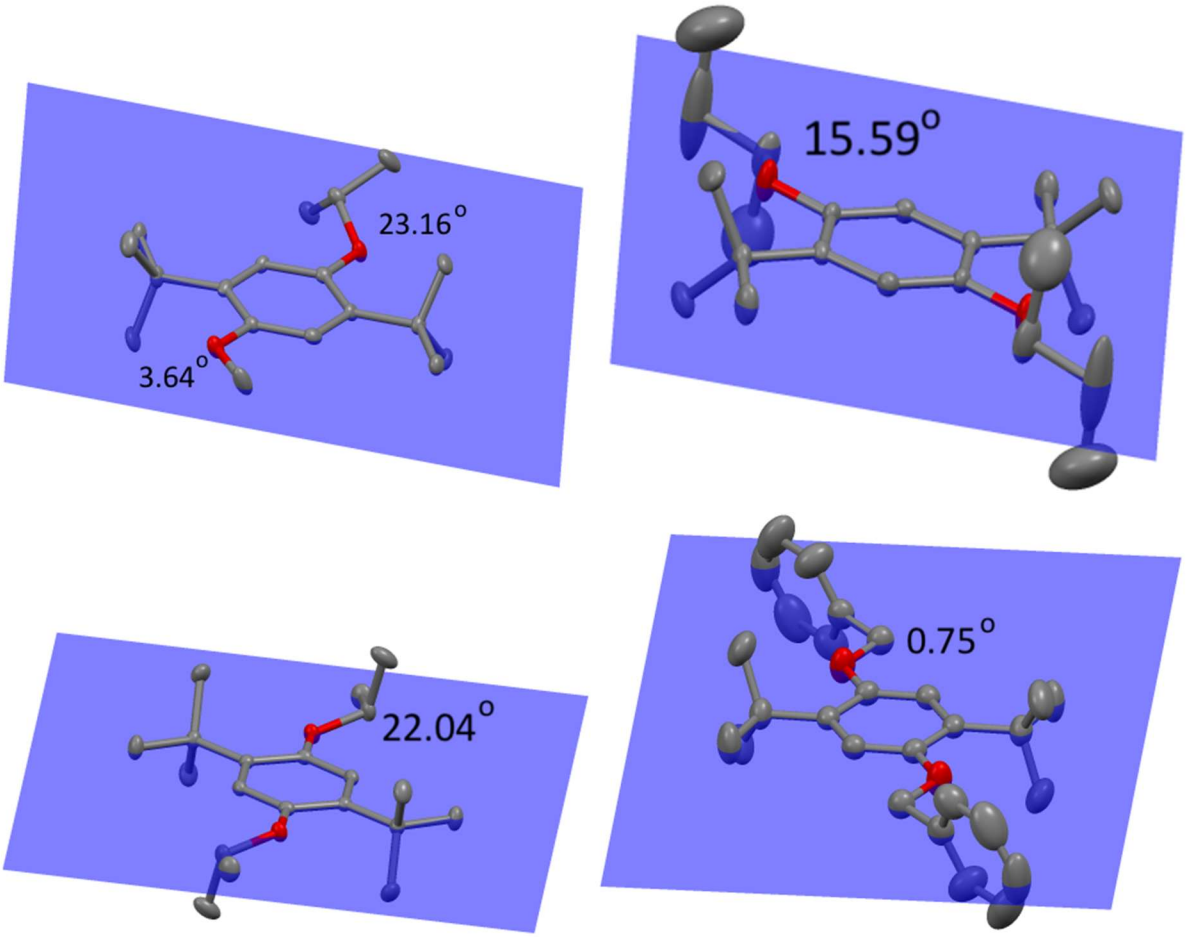


Figure S2.2. Like **Figure S2.1** for (*clockwise from the top left*): **MeiPr**, **DisBu**, **DiiPr**, and **DiBn**.

S3. Cyclic voltammetry

All electrochemical experiments were carried out in argon, using 0.5 M LiTFSI in dry acetonitrile as electrolyte. The cyclic voltammograms of 10 mM dialkoxyarene solutions were obtained at potential sweep rates of 25–500 mV/s using a CHI760D electrochemical workstation (CH Instruments, TX) in a three-electrode configuration with a Teflon encased glassy carbon disk working electrode, a platinum wire counter electrode, and an $Ag/AgNO_3$ quasi-reference electrode containing 10 mM $AgNO_3$ in acetonitrile. **Table S3.1** gives the redox potentials and the electrode current ratios at 500 mV/s. The cyclic voltammetry sweeps are shown in **Figures S3.1 and S3.2.**

Table S3.1. The half-wave redox potentials and electrode current ratios for selected dialkoxyarenes (10 mM) as obtained by cycling voltammetry in 0.5 M LiTFSI/acetonitrile (500 mV/s).

Molecule	RO group (1)	RO group (2)	$E_{1/2}$ (V vs Ag/Ag^+)	i_a/i_c [a] electrode current ratio
DiMe	methyl	methyl	0.723	0.962
MeEt	ethyl	methyl	0.673	0.914
DiEt	ethyl	ethyl	0.657	0.943
MeiPr	*iso*-propyl	methyl	0.644	0.921
DiiPr	*iso*-propyl	*iso*-propyl	0.598	0.942
MesBu	*sec*-butyl	methyl	0.648	0.941
DisBu	*sec*-butyl	*sec*-butyl	0.594	0.932
MeBn	benzyl	methyl	0.710	0.986
DiBn	benzyl	benzyl	0.731	0.877

a) Ratio of anode and cathode currents (Pt/Pt cell, 25 mV/s sweep for 10 mM solute at 25 °C).

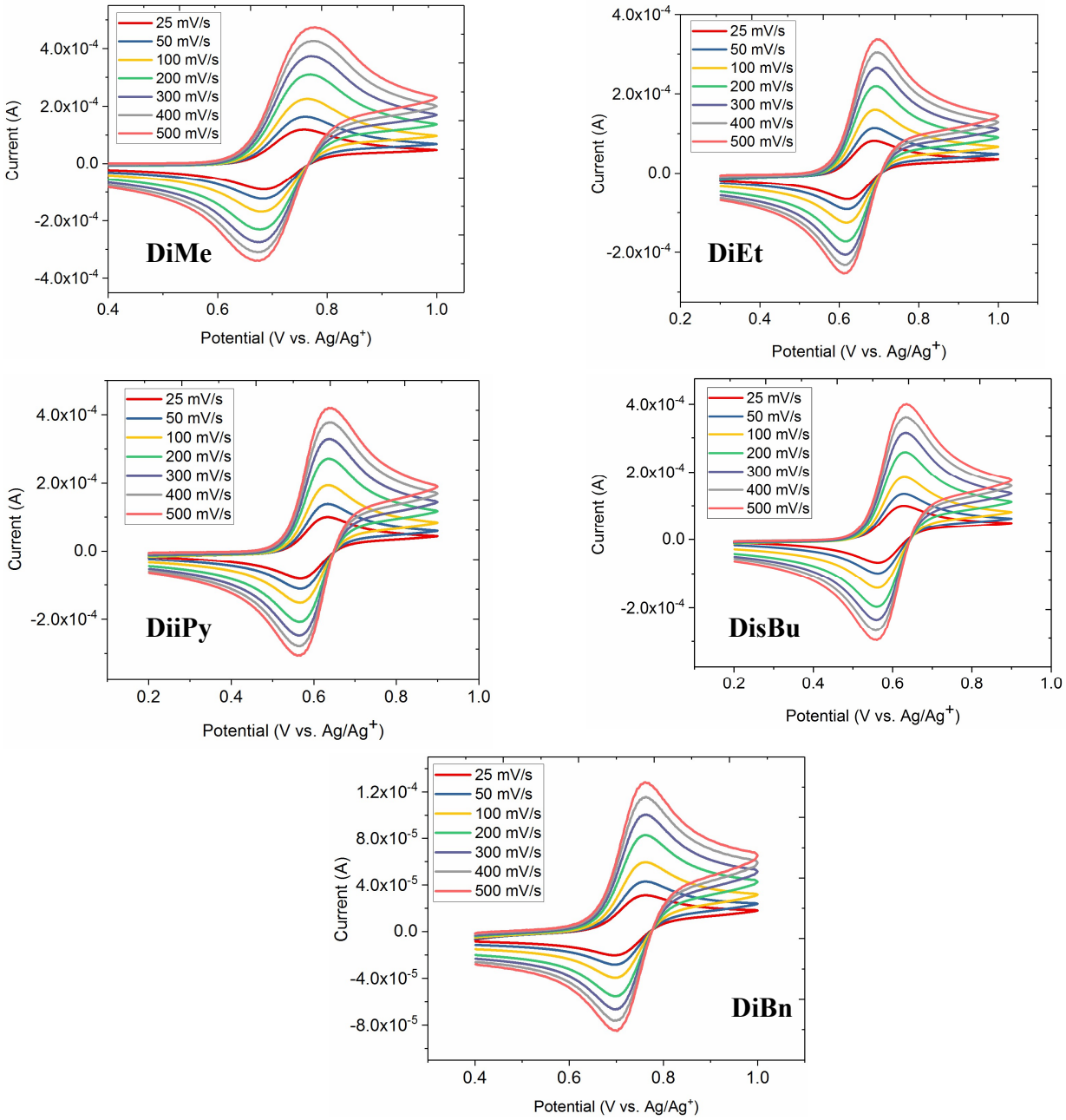


Figure S3.1. Cyclic voltammetry sweeps for 10 mM symmetric dialkoxyarenes in 0.5 M LiTFSI in acetonitrile (~25 °C, Pt/Pt cell). The sweep rates are indicated in the plot.

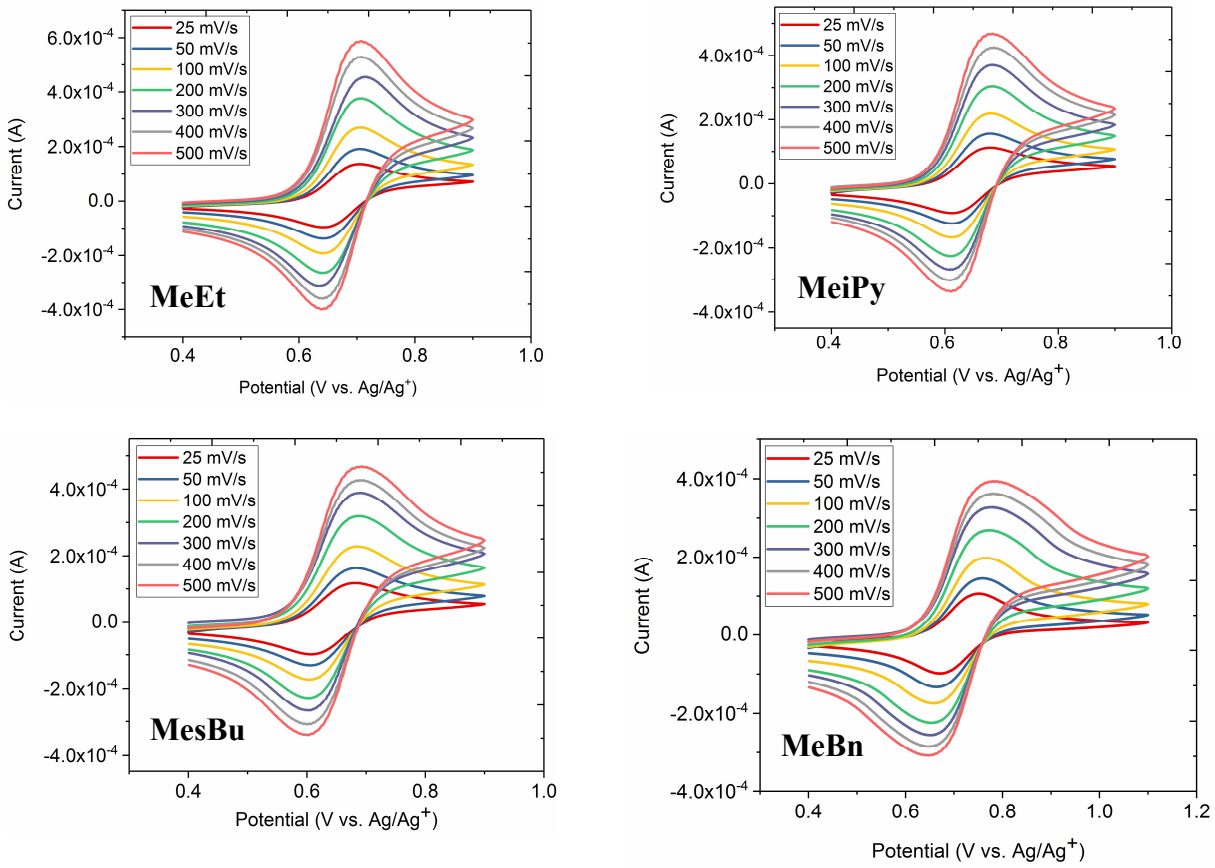


Figure S3.2. Cyclic voltammetry sweeps for 10 mM non-symmetric dialkoxyarenes in 0.5 M LiTFSI in acetonitrile (~25 °C, Pt/Pt cell). The sweep rates are indicated in the plot.

S4. Bulk electrolysis

Constant-current capacity-controlled cycling was conducted in borosilicate glass H-cells;[17, 18] the cell fluids were agitated by a magnetic stirring rod during electrolysis and cycling. Each cell compartment contained ≈ 5 mL of the same solution. The H-cell was equipped with a ceramic separator with P5 frit (2 mm, 1.0–1.6 μm pore size, 50% porosity, Adams & Chittenden). Reticulated vitreous graphite electrodes (100 PPI, ERG Aerospace) were used in both cell chambers. Parstat MC potentiostat (Princeton Applied Research) was used to control the H-cells. The potential of the counter electrode was controlled by the potentiostat to maintain a constant current rate of 3C, where 1C corresponds to full theoretical discharge in 1 h.

During the first charge, the solvent in the counter chamber of the H-cell becomes decomposed as there are not anolyte molecules to reduce while the catholyte molecules oxidizes, so the fluid in the counter chamber was replaced with a fresh solution. In this cell, when a catholyte molecule is oxidized in one chamber, the converse reaction occurs in the opposite chamber. The electrode potential is controlled only in the *workin*g electrode chamber; the potential of the counter electrode is constantly adjusted to supply the constant current. To control the potential of the working electrode, the cell was equipped with an $Ag/AgNO_3$ reference electrode placed as close as possible to the working electrode. All cells were charged to 50% SOC (state of charge, 100% SOC = theoretical capacity). The lower and the upper cut potentials were set 0.0 and 1.8 V vs Ag/Ag^+. The direction of the current reverses once the set capacity is attained or the set voltage limits are reached, whatever event comes first. The typical data are shown in **Figures S4.1 to S4.8.**

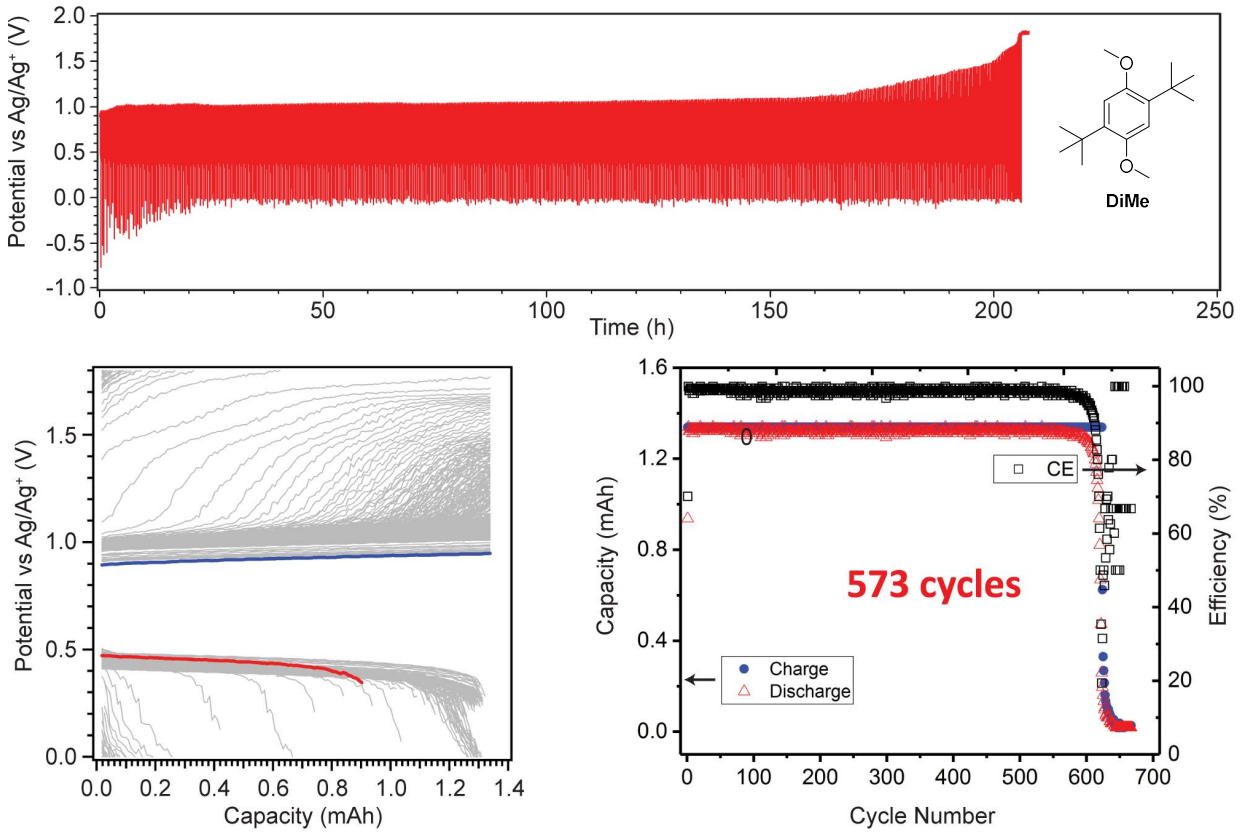


Figure S4.1. Bulk electrolysis of 5 mL electrolyte in a symmetric cell equipped with reticulated vitreous carbon electrodes and containing 20 mM **DiMe** in 0.5 M LiTFSI (3C rate, 50% SOC). The upper panel shows the potential of the working electrode (V vs Ag/Ag^+) plotted as a function of cycling time. The lower left plot shows the potential as a function of capacity for all cycles (the first cycle shown in color). The lower right plot shows time evolution of charge (*filled cicles*) and discharge (*open triangles*) capacity (*to the left*) and Coulombic efficiency (*to the right*) plotted as a function of cycle number. Note the gradual increase in the terminal charge potential after 150 h. Eventually the potential increases over the set upper cutoff potential of 1.8 V vs Ag/Ag^+ and the run is automatically terminated.

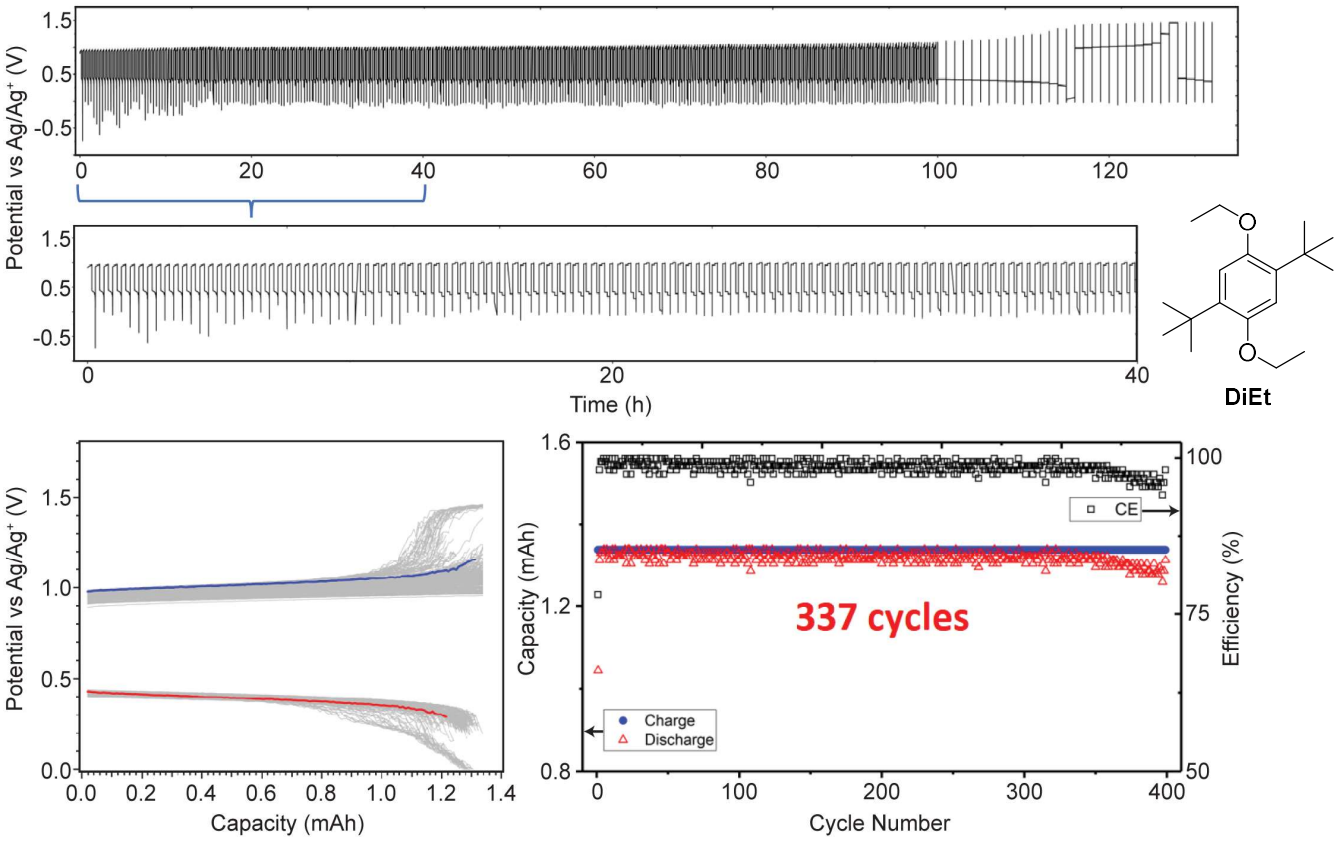


Figure S4.2. Like Figure S4.1, for **DiEt.**

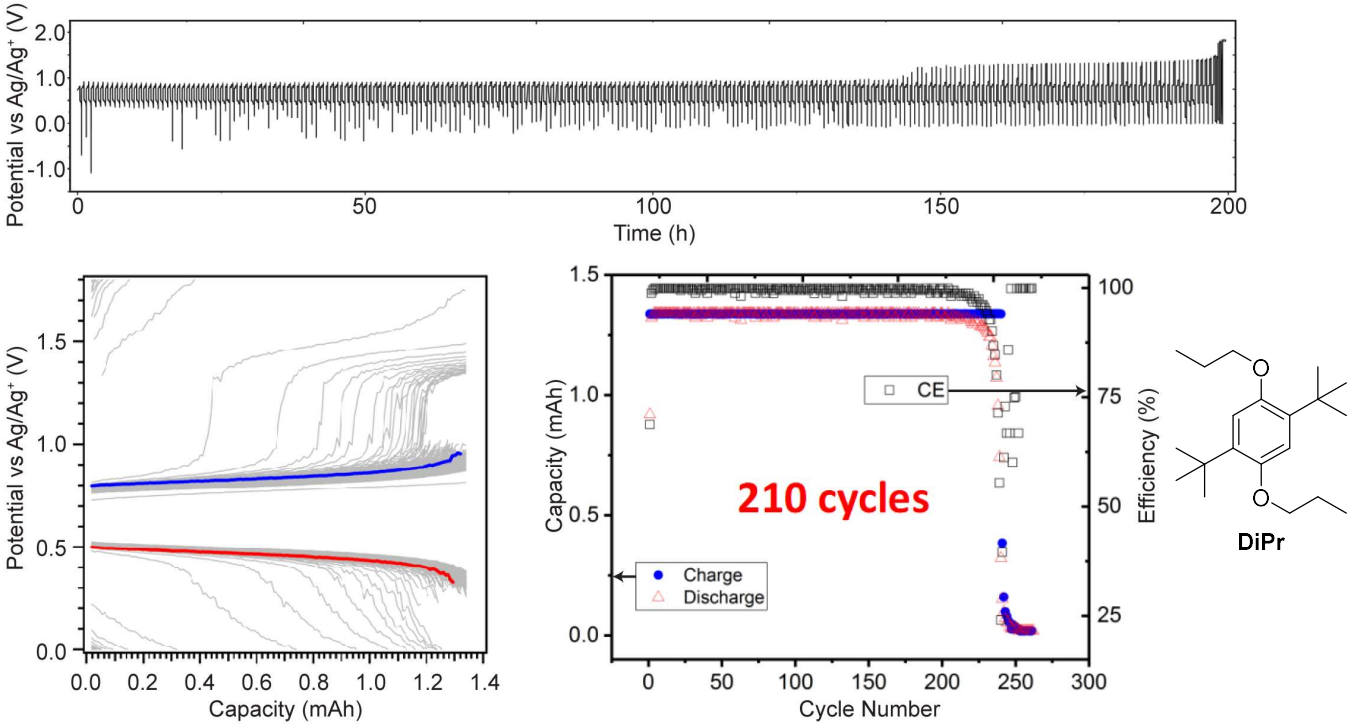


Figure S4.3. Like Figure S4.1, for **DiiPr.**

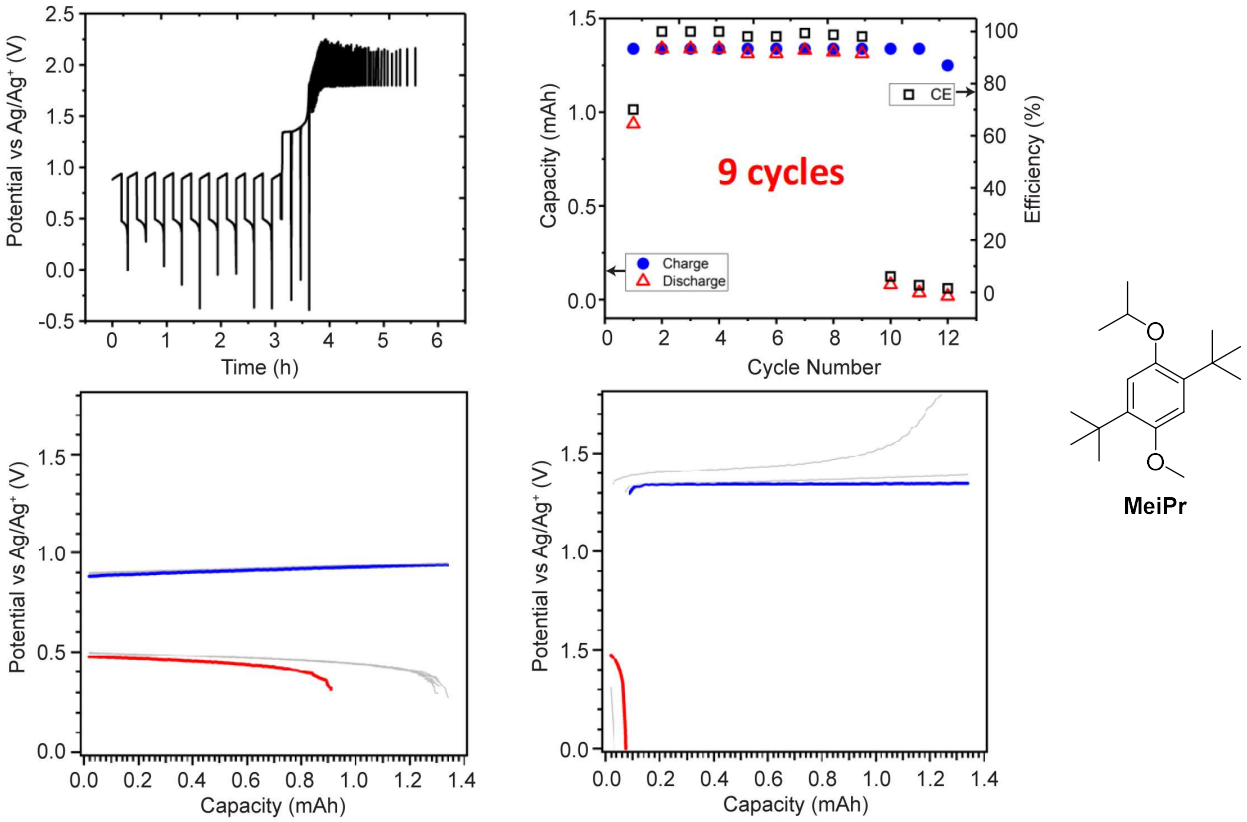


Figure S4.4. Like Figure S4.1, for **MeiPr.**

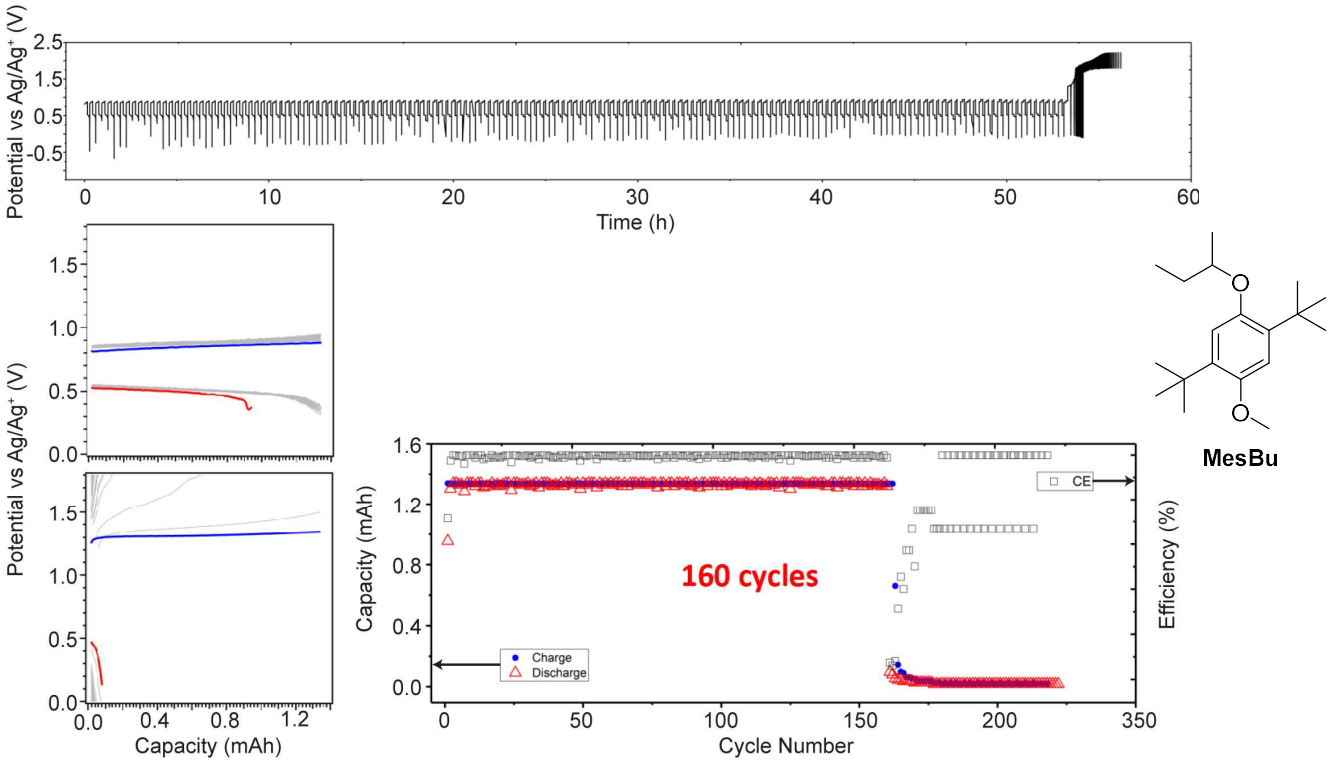


Figure S4.5. Like Figure S4.1, for **MesBu.**

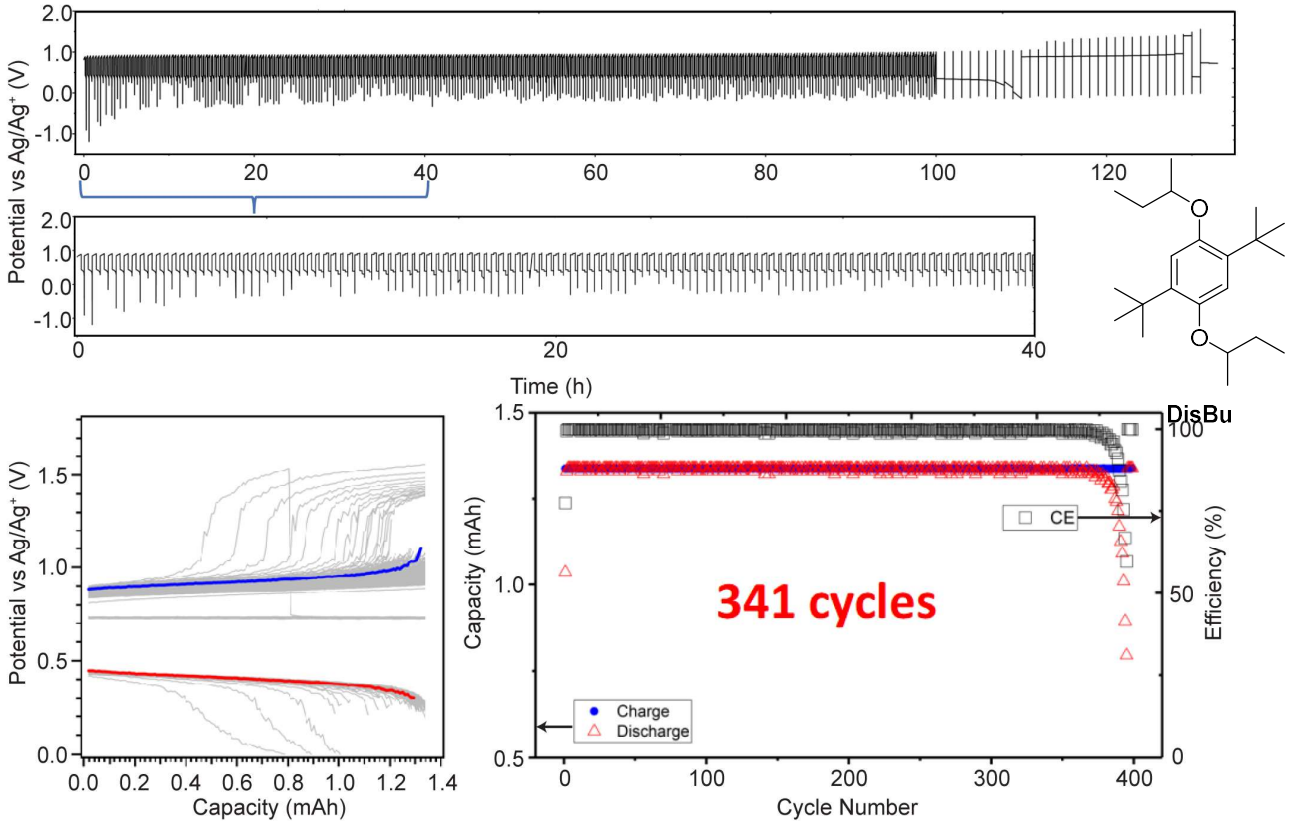


Figure S4.6. Like Figure S4.1, for **DisBu.**

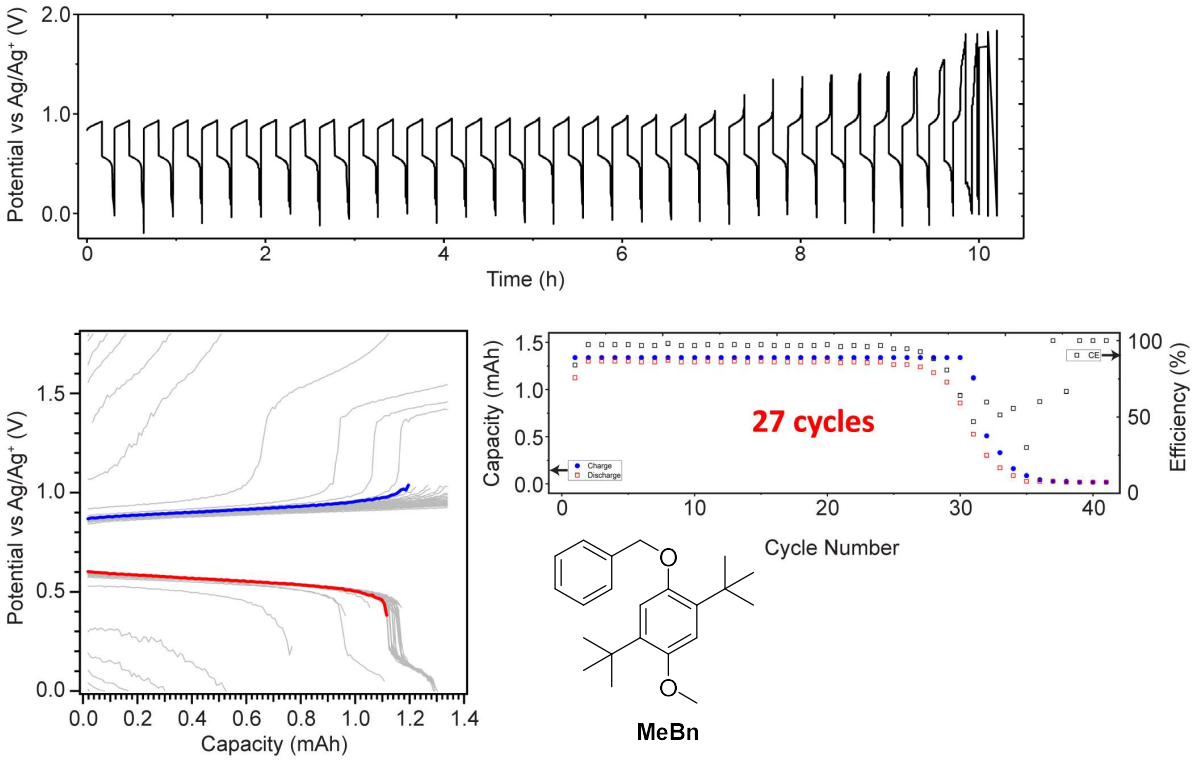


Figure S4.7. Like Figure S4.1, for **MeBn.**

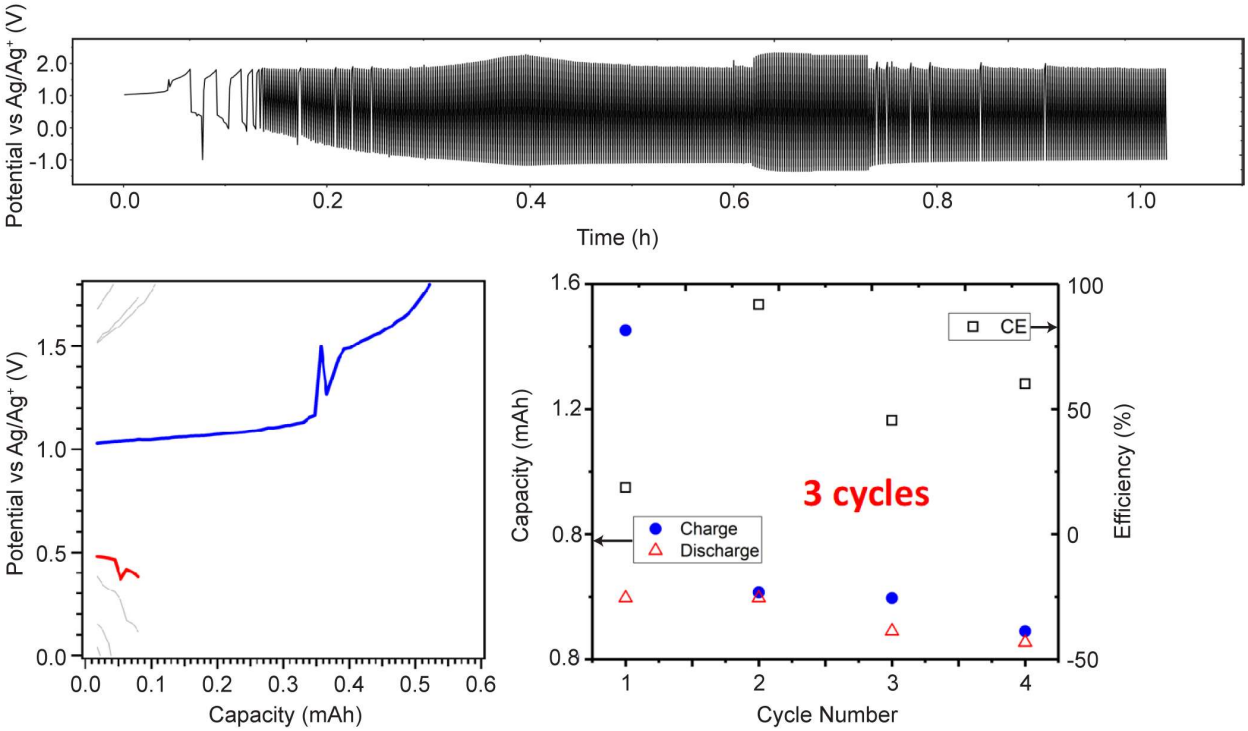


Figure S4.8. Like Figure S4.1, for **DiBn.**

S5. Kinetic measurements

As radical ions have unpaired electrons, spin resonance was used to characterize their calendar stability in electrolyte bulk. To this end, the electrolyte containing 20 mM redoxmer in 0.5 M LiTFSI in acetonitrile was charged at a rate of 5C to 100% SOC. Electrochemically-generated radical cations were observed using continuous-wave electron paramagnetic resonance (cw EPR) spectroscopy in the X-band using 100 kHz field modulation.[6, 18-23] A 100-μL aliquot of this liquid sample was placed in a glass capillary and sealed in a glass tube equipped with Teflon piston seal (Wilmad-LabGlass model 734-LPV-7). The first-derivative EPR spectra were collected at room temperature using a Bruker ESP300E X-band spectrometer operating at 9.43 GHz; field modulation of 0.2 G (1 G = 10^{-4} T) was used.

The hyperfine coupling constants in **Table S5.1** were estimated from these spectra using WinSim suite; the electrolyzed solution was diluted 1:20 v/v to improve spectra resolution for this analysis. For analyses of decay kinetics (no dilution used), the EPR spectra were recorded at fixed time intervals, centered, background corrected, and doubly integrated. These decay kinetics were fit by biexponential functions and the half-life time $t_{1/2}$ of the slower component were used as a metric for calendar lifetime (the fast component is due to residual charge scavenging impurities in the solvent). **Figures S5.1** to **S5.6** contain examples of normalized decay kinetics for dialkoxyarenes in this study.

Table S5.1. Absolute isotropic hyperfine coupling constants (in Gauss, 1 G=10^{-4} T) for arene and α-alkoxy protons in radical cations of dialkoxyarenes (1 mM) in acetonitrile solution containing 0.5 M LiTFSI. [a]

Molecule	B3LYP/631+G(d,p) calculation				EPR spectroscopy			
		Arene		Arene		Arene		Arene
Proton	RO (α)	H	RO (α) [b]	H	RO (α)	H	RO (α)	H
Position	1	6	4	3	1	6	4	3
DiMe	3.76	0.43	3.76	0.43	3.19	1.02	3.19	1.02
MeEt	5.80	0.55	3.62	0.33	3.74	1.11	3.13	0.99
DiEt	5.21	0.47	5.21	0.47	3.69	1.03	3.69	1.03
MeiPr	1.97	0.70	3.62	0.37	1.78	1.13	3.10	0.96
DiiPr	1.85	0.62	1.85	0.62	1.71	1.05	1.71	1.05
MesBu	1.81	0.70	3.62	0.36	1.67	1.13	3.11	0.97
DisBu	1.76	0.63	1.76	0.63	1.64	1.08	1.64	1.08
MeBn	5.16	0.46	3.74	0.39	3.63	1.00	3.17	0.97
DiBn	5.13	0.44	5.13	0.44	3.74	0.97	3.74	0.97

a) The first-derivative EPR spectra were obtained using a Bruker ESP300E spectrometer operating at 9.4 GHz and simulated using WINSIM suite; 100 kHz field modulation at 25 °C (2 mW, 0.2 G).1:10 v/v dilution of 10 mM electrolyzed solution (5C to 100% charge) of the corresponding molecule. b) Methoxy group side.

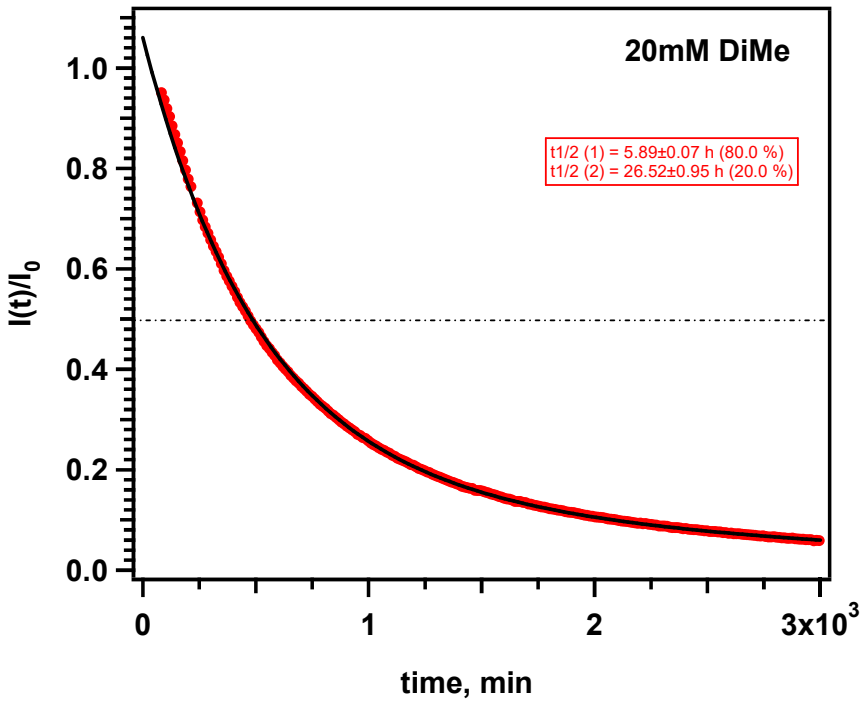


Figure S5.1. Decay kinetics for the doubly integrated first-derivative EPR signal $I(t)$ from the radical cation of **DiMe** in 0.5 M LiTFSI in acetonitrile (red circles). This signal is normalized to the initial EPR signal $I_0 = I(t = 0)$ after electrolysis and is proportional to the concentration of the radical cation (0.2 G modulation at 100 kHz, microwave poser of 2 mW). The solid line is a biexponential fit. The fast component is due to reactions of the radical cation with impurity, while the slow component corresponds to the natural lifetime of the radical cation. The radical cation was generated by electrolysis of 20 mM **DiMe** solution at 5C to 100% SOC.

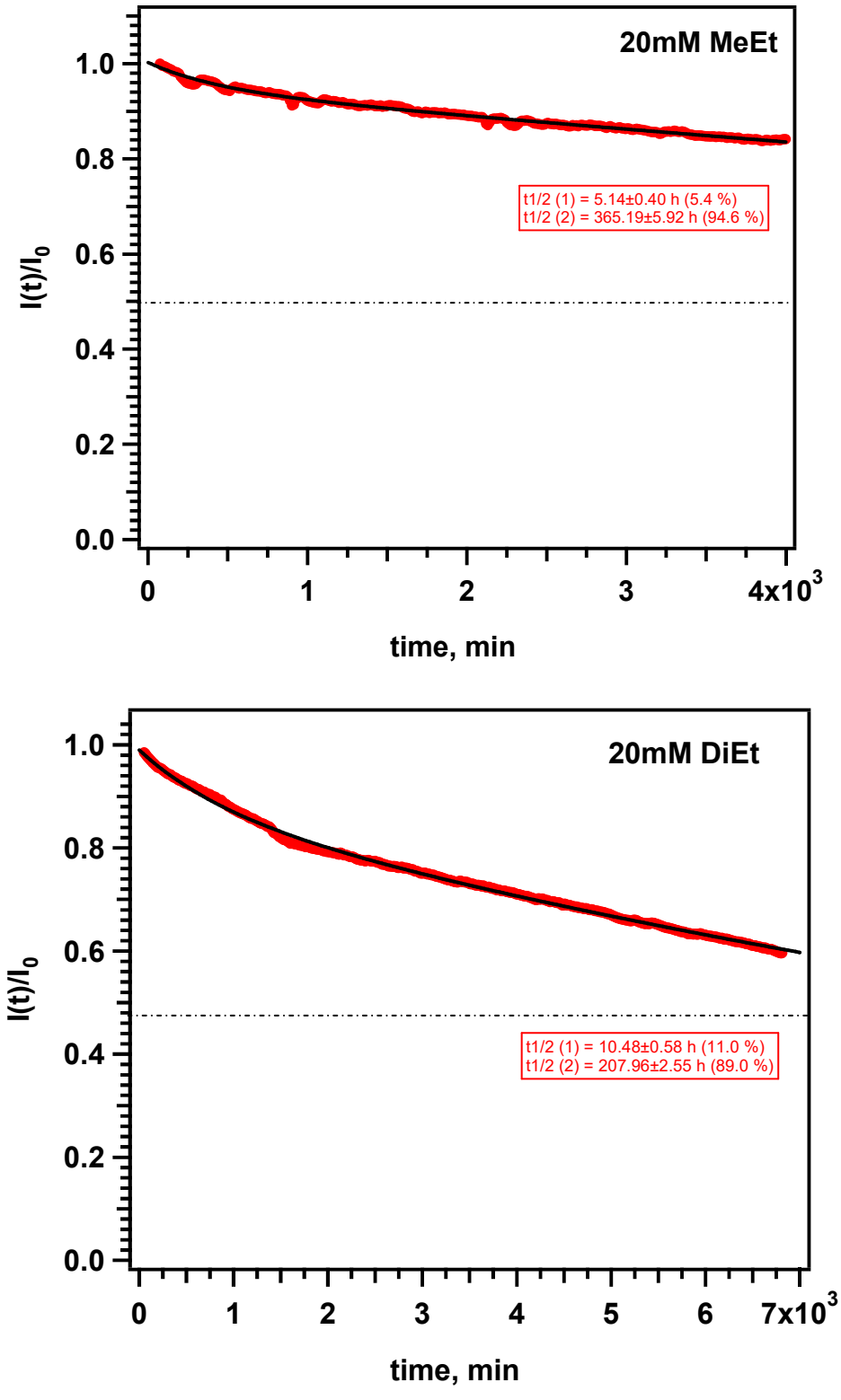


Figure S5.2. Like Figure S5.1 for **MeEt** and **DiEt**.

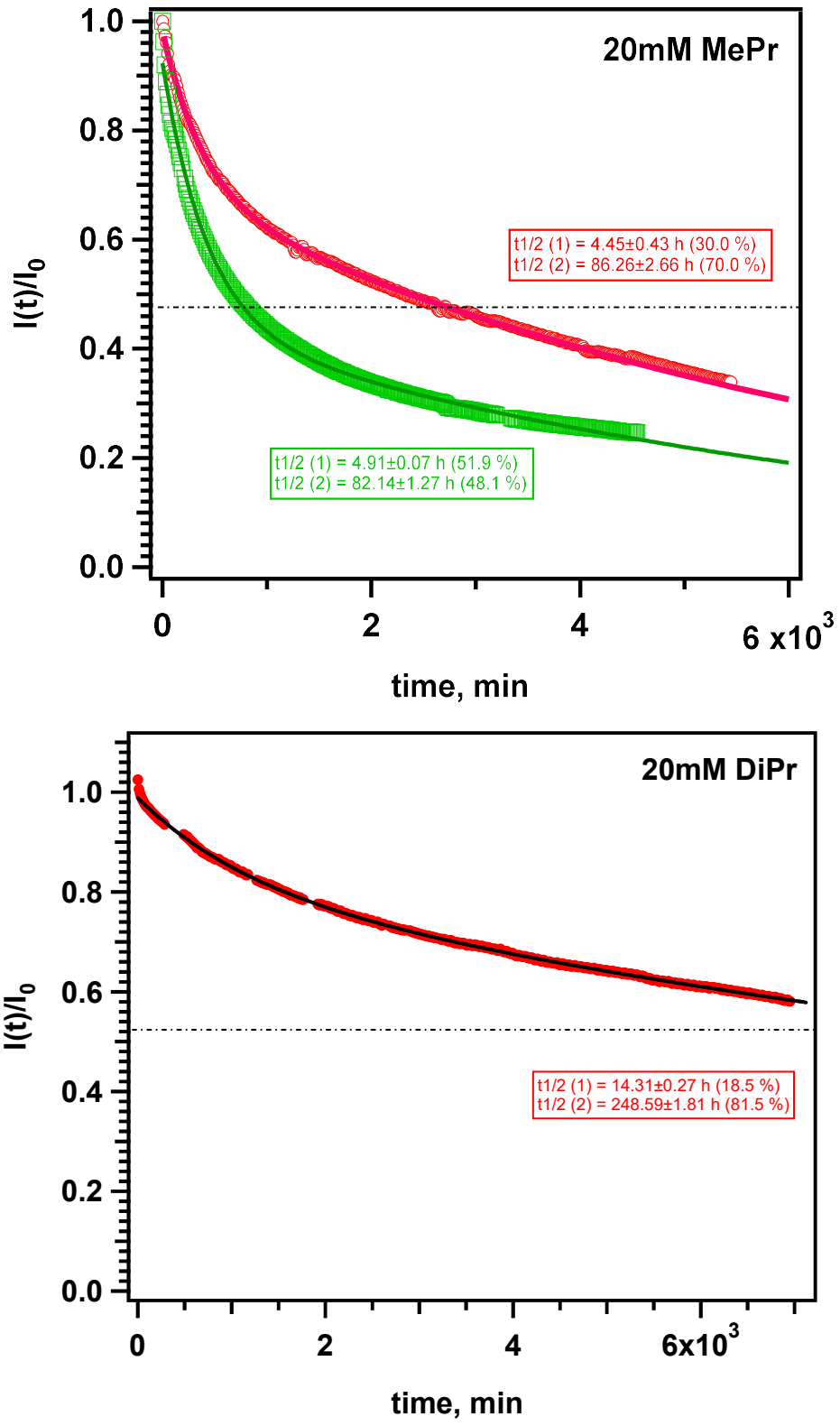
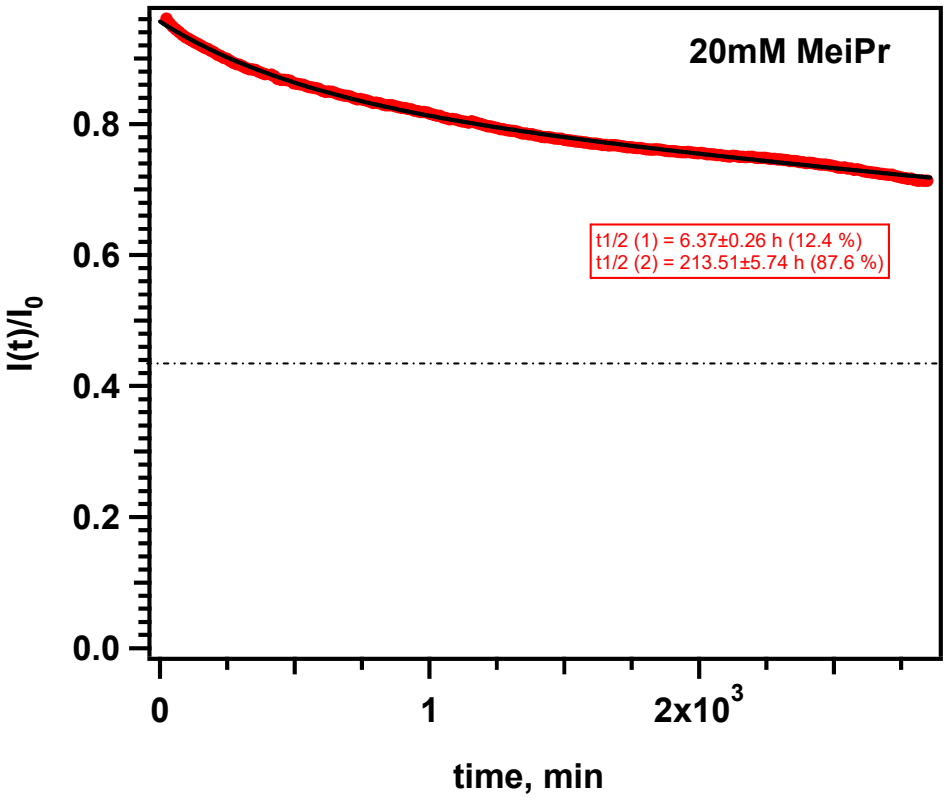


Figure S5.3. Like Figure S5.1 for **MePr** and **DiPr**. For **MePr**, two runs are shown. While the fast decay depends on the concentration of impurity, the slow decay is identical for both of these two runs.



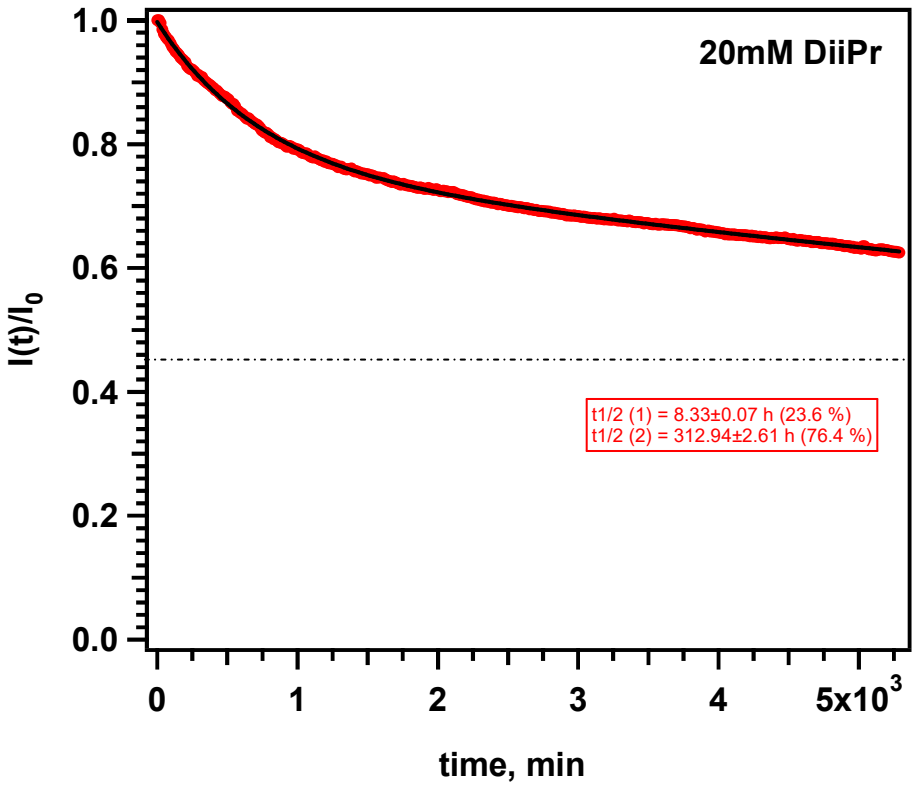


Figure S5.4. Like Figure S5.1 for **MeiPr** and **DiiPr**.

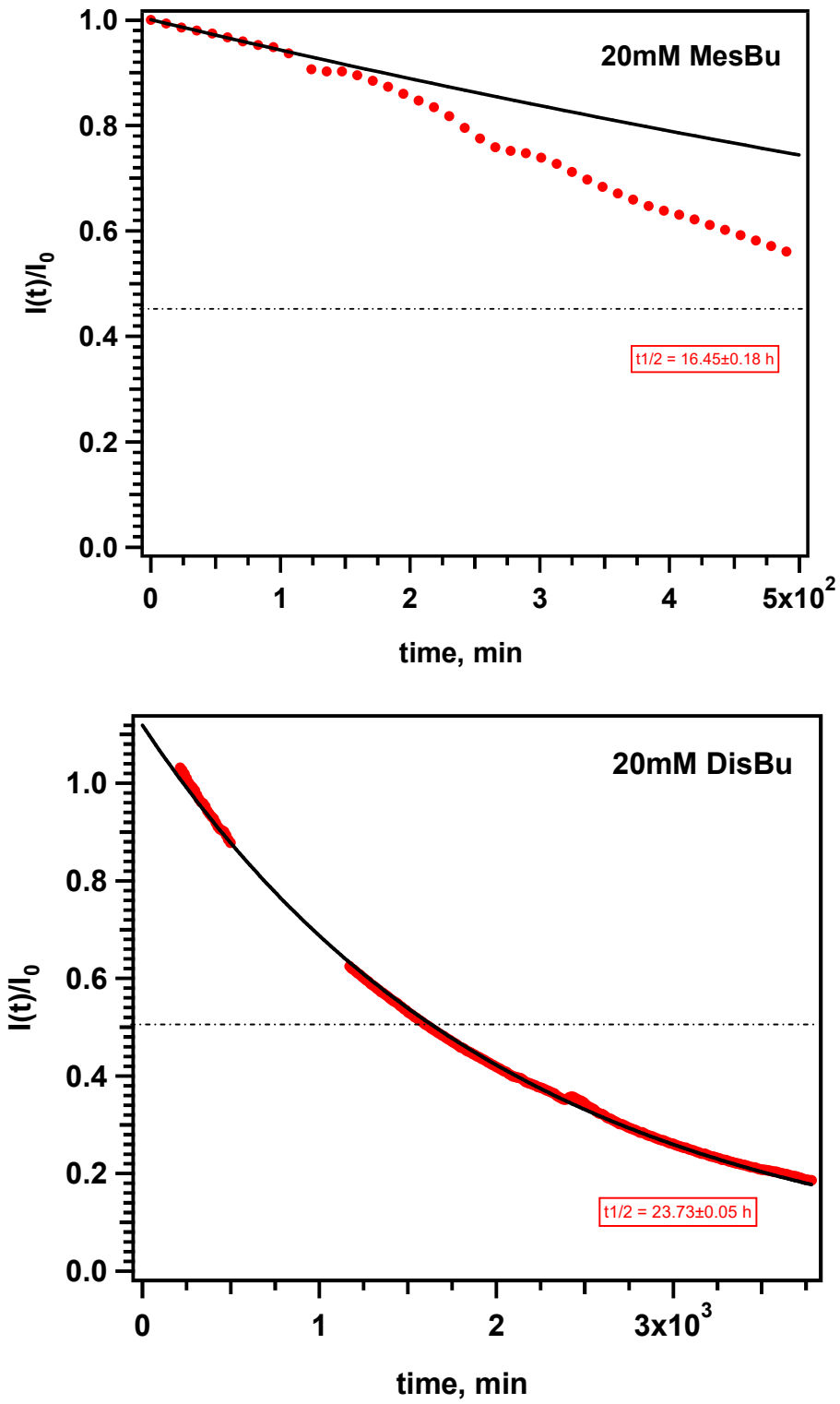
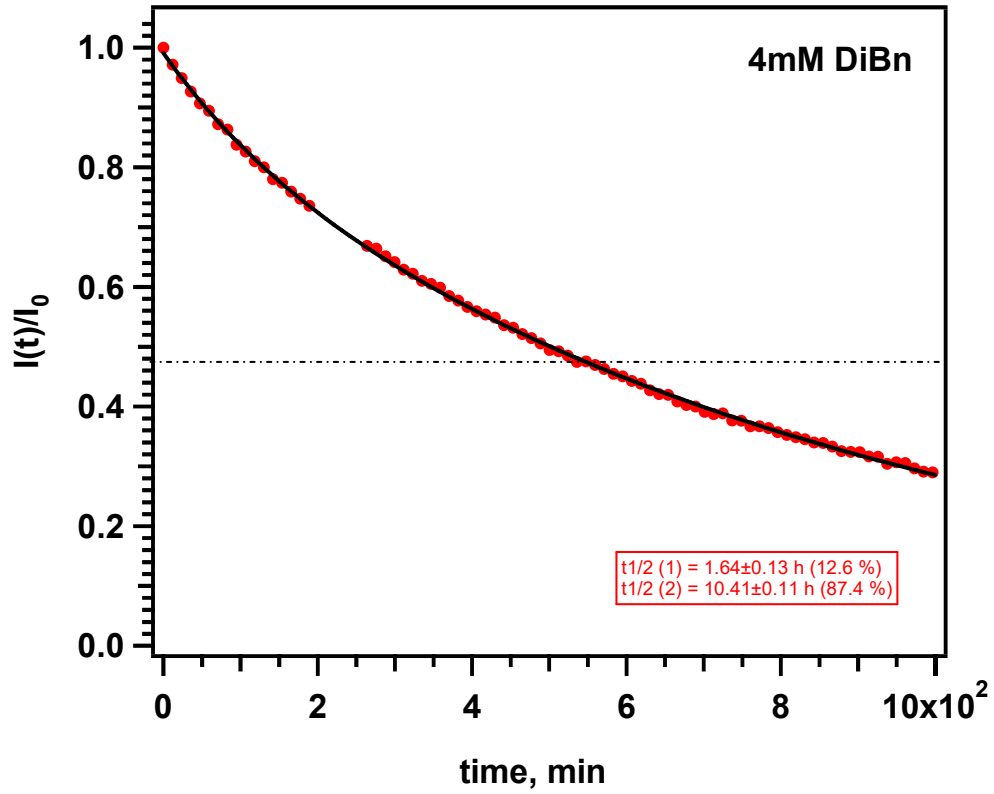


Figure S5.5. Like Figure S5.1 for **MesBu** and **DisBu**.



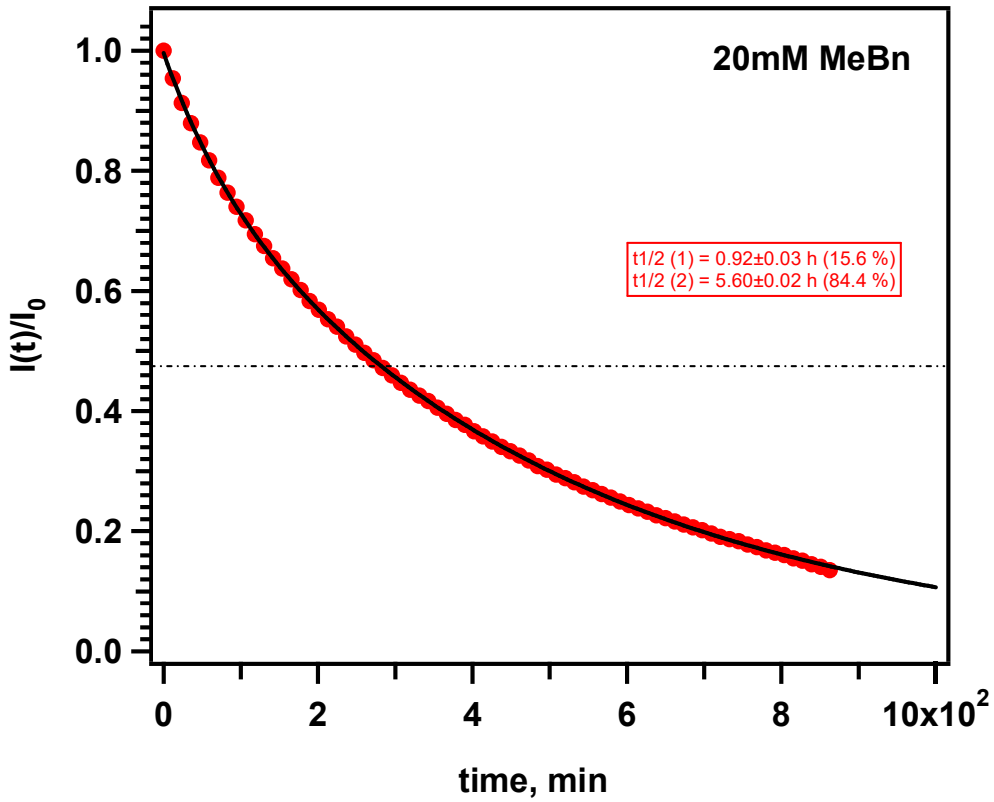


Figure S5.6. Like Figure S5.1 for **MeBn** and **DiBn**.

S6. MMLR-GA modeling.

S6.1. Models. Multiple multivariate linear (MMLR) regression in structural descriptors was used as our base model. The experimental measurements $\boldsymbol{Y}=\{y_{ik}\}$ (i=1,…N_y variables for k=1,…n measurements) are correlated with the explanatory variables $\boldsymbol{X}=\{x_{jk}\}$ (j=1,…N_x variables) that can include structural descriptors or environmental parameters:

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \text{(S6.1)}$$

where $\boldsymbol{\beta}$ is a coefficients matrix to be determined and $\boldsymbol{\varepsilon}$ is the residuals matrix, whose norm $\|\boldsymbol{\varepsilon}\|$ is minimized using the calculus of variations Eq. S6.1 relates to the *standard* (reduced) form variables, in which the means of n measurements are subtracted from each $\boldsymbol{x}$ and $\boldsymbol{y}$ variable and the resulting differences are normalized by the standard deviation for each variable. The optimum matrix $\boldsymbol{\beta}$ is given by[24]

$$\boldsymbol{\beta} = \mathbf{S}_{xx}^{-1}\mathbf{S}_{yx} \qquad \text{(S6.2)}$$

where $\mathbf{S}_{xx}=\boldsymbol{X}^t\boldsymbol{X}$ is the matrix of covariations, $\mathbf{S}_{yx}=\boldsymbol{X}^t\boldsymbol{Y}$, and

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\,\boldsymbol{\beta}\,, \qquad \text{(S6.3)}$$

so that $\boldsymbol{\varepsilon}=\boldsymbol{Y}-\hat{\boldsymbol{Y}}$. The coefficient of determination r^2, which is the measure of multivariate association for multivariate linear regression (MLR, N_y=1), cannot be uniquely defined for MMLR,[24] and we used the Wilks pseudo-r^2 measure (also known as the Hotelling-Cramer measure) given by

$$r^2 = det[\mathbf{S}_{\varepsilon\varepsilon}]/det[\mathbf{S}_{yy}] \qquad \text{(S6.4)}$$

where $\mathbf{S}_{\varepsilon\varepsilon}=\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon}$ and $\mathbf{S}_{\hat{y}\hat{y}}=\hat{\boldsymbol{Y}}^t\hat{\boldsymbol{Y}}$. For univariate regression with N_y=1, this equation gives the usual r^2 coefficient based on variance decomposition, but in the multivariate case, it does not have this intuitive meaning. It measures reduction of generalized variance rather than prediction accuracy, it aggregates effects across multiple response variables, it depends strongly on correlations among responses, and it increases with model flexibility.

Our goal was to find the smallest set of explanatory variables that accounts for most of variation in the data. This task can be achieved using penalization terms in Eq. S6.1 via ridge, Lasso,[25] or elastic net regularization.[26] However, these common methods proved to be ineffective for small data sets, we used the genetic algorithms (GA) introduced in refs.[27-29]

In this approach (Figure 1), multiple subsets χ of $m<N_x$ variables ("chromosomes," with each explanatory variable representing a "gene") are generated at random and eqs. S6.2 and S6.3

are solved for each subset with the optional ridge regularization. For quadratic regression, this step is preceded by removal of tightly correlated products of the variables (with the absolute Pearson's correlation coefficient over the training set > 0.9). The root mean squares deviation (rms) for each "chromosome" serves as the measure of its "fitness." At each iteration, 50 "chromosomes" are sorted and 10 most fit "chromosomes" are selected at random according to the precedence of their fitness. Two "mutations" (random substitutions of the "genes") are introduced per "chromosome," and random fragments of these chromosomes are interchanged (Figure 2). Some "genes" can become duplicated during this exchange, which is equivalent to reducing the effective size m of subset χ. These variations are introduced at each generation, as the population evolves searching for the global minimum on the fitness landscape. Typically, (0.1–2)x10^3 iterations were sufficient to locate the optimal subset. For this subset, the rank r_χ for the covariance matrix $\mathbf{S}\chi\chi$ is calculated and, if $r_\chi < m$, m is reduced, and this procedure is repeated.

The same approach was used for logistic regression to maximize the log-likelihood for prediction of odds. To this end, n measurements are divided into two classes labeled 0 and 1 containing n_0 and n_1 elements, respectively. The classifier is given by the linear log-odds $logit(p) = \boldsymbol{X}\boldsymbol{\beta} + \beta_0$, where p_i is the estimated probability of belonging to class-1, and the linear coefficients correspond to the maximum log-likelihood $L(\beta_0, \boldsymbol{\beta})$. Variation of the latter gives the nonlinear matrix equation that was solved for each trial set χ using iteratively re-weighted least squares method until the convergence criterion for the relative mean deviation in the absolute coefficients was met. Ridge regularization was used to suppress possible linear correlations between the variables in the "chromosomes." McFadden pseudo-r^2 defined as

$$r^2 = 1 - L(\beta_0, \boldsymbol{\beta})/L_0 \qquad \text{(S6.5)}$$

where the null model

$$L_0 = n_0 \log\frac{n_0}{n} + n_1 \log\frac{n_1}{n}, \qquad \text{(S6.7)}$$

was used as a goodness-of-fit criterion. Our GA routine maximizes r^2 computed from eq. S6.5, which is equivalent to minimizing pseudo-rms given by

$$\varepsilon_k^2 \,/\, 2 = y_k \, log\frac{y_k}{p_k} + \bar{y}_k \, log\frac{\overline{y_k}}{\overline{p_k}}, \qquad \text{(S6.8)}$$

where ε_k is the deviance residual.

S6.2. Cross validation. MMLR–GA optimizes the problem of finding the minimal set of descriptors yielding the minimum prediction variance, and in this sense the solution is uniquely defined by the data and the descriptor pool. However, removal of data points or addition of new

data can lead to new optimal descriptor sets. Since the usefulness of the optimal descriptors as guides for mechanistic inference depends on their reliable identification, cross-validation is necessary to demonstrate the robustness of this optimal set. To characterize this robustness, bootstrapping was used. The data set was randomly resampled by retaining 90% of the data points (200 trials total), and the MMLR–GA optimization with the same number of descriptors was performed for each trial. Figure S6.5 shows the results of this analysis for the calendar time data presented in Figure 5b of the main text. Panel a shows the averages of the absolute magnitudes of the linear coefficients β across all trials (for the reduced descriptor sets). The more frequently a descriptor appears in this series of trials and the larger the linear response associated with it, the more prominently it appears in the histogram. It is seen that the three descriptors selected by MMLR–GA for the full data set are also the most statistically prevalent in the incomplete data sets.

Panel b further quantifies this outcome by introducing a mismatch score between the optimal descriptor sets obtained in each trial and the original descriptor set for full data. Two such scores are defined. The semantic score quantifies the literal distinction between these sets. For example, if the trial set were abc and the original set were aef, the mismatch score would be 2/3. This score, however, can be misleading because different descriptor sets can define similar linear spaces as these descriptors are not linearly independent over the data sets. Thus, a more meaningful metric is the extent to which the two linear subspaces of the same dimension differ from each other. Such metrics are known as Grassmannian metrics, and they can be constructed in several ways, analogous to the taxicab and Euclidean metrics in ordinary space. We used a mismatch metric based on the Binet–Cauchy distance, normalized such that the score is 0 for identical spaces and 1 for orthogonal ones (i.e., when all principal angles between the linear subspaces are 90°). Panel b illustrates the critical difference between the semantic and Grassmannian scoring: while multiple trials yield descriptor sets that may be semantically different from the original one, the linear spaces defined by these sets remain close to the original. The higher-scoring sets have low weights in the histogram of mismatch scores. Similar results are obtained for the data shown in Figure 5a in the main text.

This cross validation example illustrates that small-scale variation in the data shown in Figure 5 does not lead to dramatic changes in the descriptor sets seen as progenitors of linear subspaces, and in this technical sense the fit is robust, so “reading” into the mechanistic content of the full set optimized descriptors is justified. That said, while the solution is robust to point deletion, one can never be certain that addition of new data points would not result in a different set of optimal descriptors, which is the case for all models.

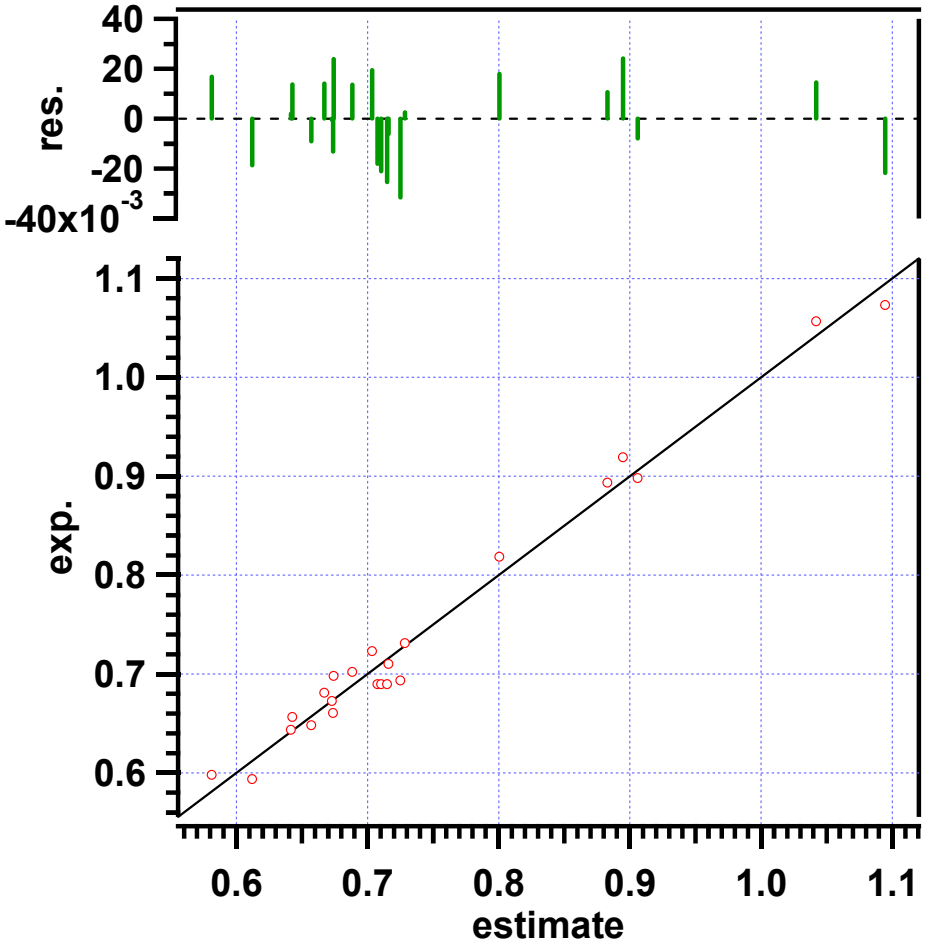


Figure S6.1. Correlation of experimental (*open circles*) and estimated $E_{1/2}$ (in V vs Ag/Ag^+) for 20 dialkoxyarenes given on the next page in 0.5 M LiTFSI in acetonitrile. The computed redox potential (*solid line*) is given by $E_{1/2} = 0.42857 + 0.03665\ MolLogP - 1.00738\ ATSC7c$, where the descriptors are explained in the text ($r^2 = 0.982$). The values of the descriptors are given on the next page. The vertical green bars show fit residues.

	Dialkoxyarene[a]	SMILES	$E_{1/2}$ (V vs Ag/Ag^{+})		Molecular descriptors	
			Exp.	Fit	*MolLogP*	*ATSC7c*
1	**DiMe**	COc1cc(C(C)(C)C)c(OC)cc1C(C)(C)C	0.723	0.704	4.2988	-0.1166
2	**MeEt**	CCOc1cc(C(C)(C)C)c(OC)cc1C(C)(C)C	0.673	0.673	4.6889	-0.0720
3	**DiEt**	CCOc1cc(C(C)(C)C)c(OCC)cc1C(C)(C)C	0.656	0.643	5.0790	-0.0278
4	**MePr**	CCCOc1cc(C(C)(C)C)c(OC)cc1C(C)(C)C	0.702	0.688	5.0790	-0.0732
5	**DiPr**	CCCOc1cc(C(C)(C)C)c(OCCC)cc1C(C)(C)C	0.660	0.674	5.8592	-0.0302
6	**MeiPr**	COc1cc(C(C)(C)C)c(OC(C)C)cc1C(C)(C)C	0.643	0.642	5.0774	-0.0268
7	**DiiPr**	CC(C)Oc1cc(C(C)(C)C)c(OC(C)C)cc1C(C)(C)C	0.598	0.581	5.8560	0.0615
8	**MesBu**	CCC(C)Oc1cc(C(C)(C)C)c(OC)cc1C(C)(C)C	0.648	0.657	5.4675	-0.0279
9	**DisBu**	CCC(C)Oc1cc(C(C)(C)C)c(OC(C)CC)cc1C(C)(C)C	0.594	0.612	6.6362	0.0592
10	**MeBn**	COc1cc(C(C)(C)C)c(OCc2ccccc2)cc1C(C)(C)C	0.710	0.716	5.8692	-0.0719
11	**DiBn**	CC(C)(C)c1cc(OCc2ccccc2)c(C(C)(C)C)cc1OCc1ccccc1	0.731	0.728	7.4396	-0.0271
12	**MeEO**	COCCOc1cc(C(C)(C)C)c(OC)cc1C(C)(C)C	0.689	0.708	4.3154	-0.1200
13	**DiEO**	COCCOc1cc(C(C)(C)C)c(OCCOC)cc1C(C)(C)C	0.689	0.71	4.3320	-0.1223
14	**MeEO2**	COCCOCCOc1cc(C(C)(C)C)c(OC)cc1C(C)(C)C	0.689	0.715	4.3320	-0.1267
15	**DiEO2**	COCCOCCOc1cc(C(C)(C)C)c(OCCOCCOC)cc1C(C)(C)C	0.693	0.725	4.3652	-0.1356
16	**JH1**	COc1ccc(OC)c(C)c1C	0.698	0.674	2.3206	-0.1593
17	**JH2**	COc1cc(C)c(OC)cc1C	0.681	0.667	2.3206	-0.1522
18	**FMe**	COc1cc(C(C)(C)C)c(OCC(F)(F)F)cc1C(C)(C)C	0.893	0.883	5.2313	-0.2606
19	**DiFMe**	CC(C)(C)c1cc(OCC(F)(F)F)c(C(C)(C)C)cc1OCC(F)(F)F	1.056	1.042	6.1638	-0.3847
20	**FEt**	COc1cc(C(C)(C)C)c(OCC(F)(F)C(F)(F)F)cc1C(C)(C)C	0.898	0.906	5.8666	-0.2604
21	**DiFEt**	CC(C)(C)c1cc(OCC(F)(F)C(F)(F)F)c(C(C)(C)C)cc1OCC(F)(F)C(F)(F)F	1.073	1.095	7.4344	-0.3908
22	**FBn**	COc1cc(C(C)(C)C)c(OCc2c(F)c(F)c(F)c(F)c2F)cc1C(C)(C)C	0.818	0.801	6.5647	-0.1305
23	**DiFBn**	CC(C)(C)c1cc(OCc2c(F)c(F)c(F)c(F)c2F)c(C(C)(C)C)cc1OCc1c(F)c(F)c(F)c(F)c1F	0.919	0.895	8.8306	-0.1417

a) Entries 12–23 are addition dialkoxyarenes under study in our laboratory.

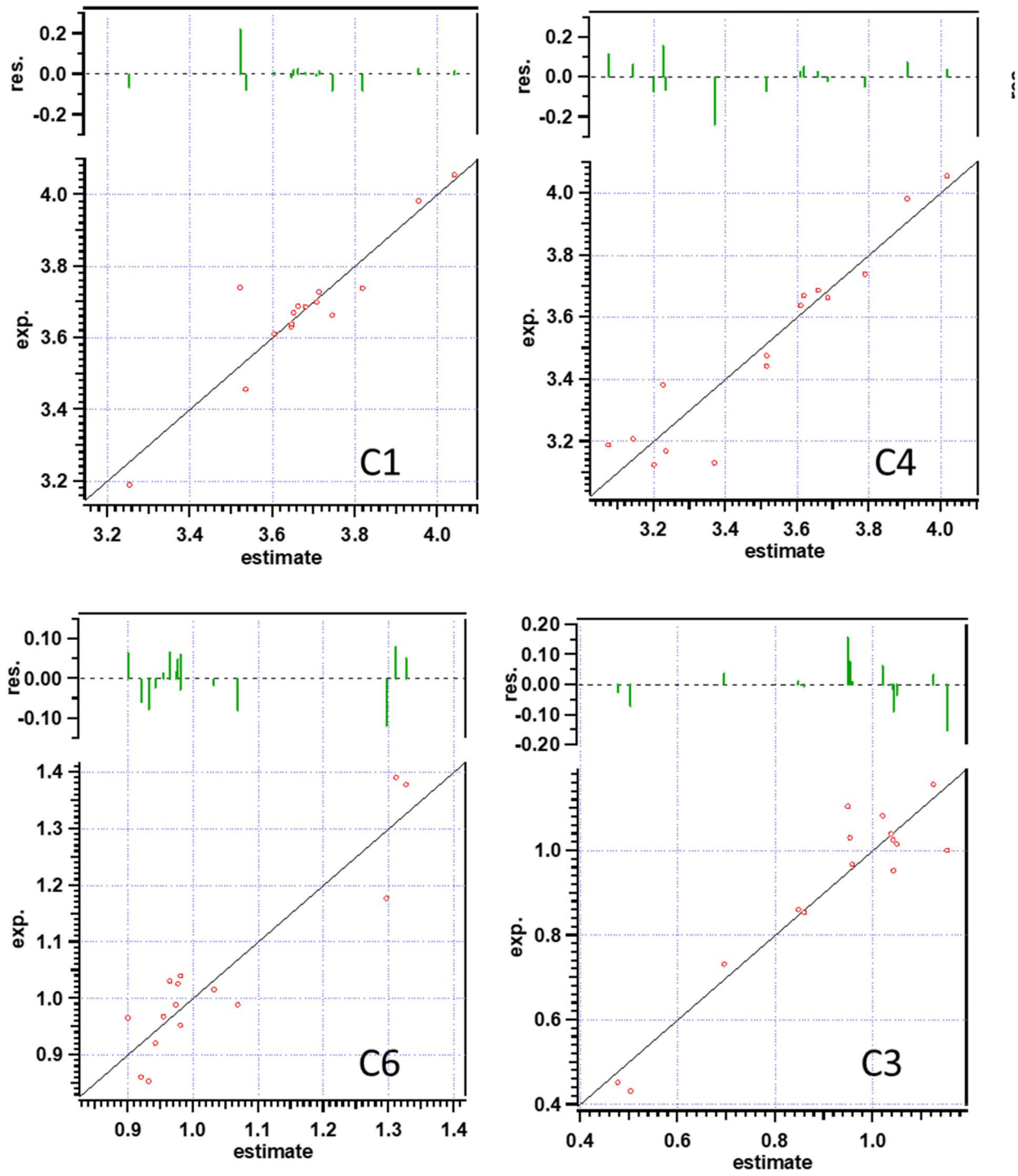
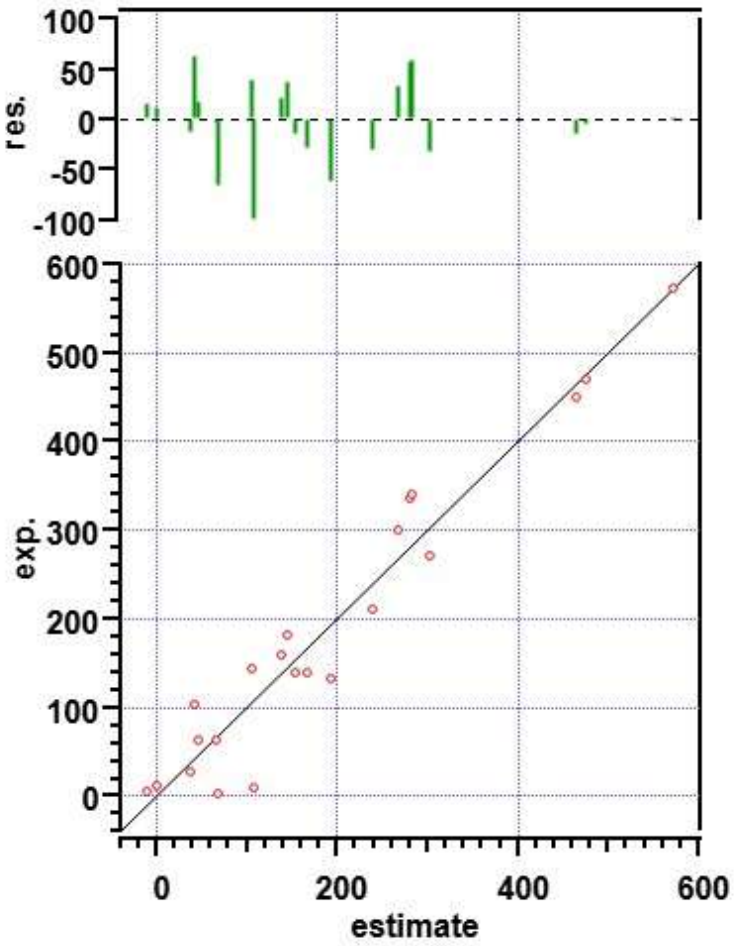


Figure S6.2. Multiple multivariate linear regression (MMLR) of the four proton hfcc's (at carbons -1, -4, -6, and -3 from left to right) given in Table S4 for radical cations of dialkoxyarenes shown in Figure S6.1. The table on the next page gives linear coefficients for the six generic descriptors and their types.

Descriptor name	Linear coefficients for proton hfcc's at carbon positions 1	4	6	3
ATS8se	0.001028	0.000026	0.000998	-3.3E-05
AATSC5d	-1.25547	1.810024	-1.92984	-3.01627
MATS5i	-4.3912	-3.68247	-0.65618	4.562641
GATS1i	-0.55845	-0.32598	-0.65876	0.653768
RPCG	-2.5943	-6.88385	-0.25545	6.654242
CIC3	-0.8702	-0.11425	0.664839	0.070668
Linear offset	2.20307	-0.5674	2.003658	-0.46164

Descriptor name	Descriptor type
ATS8se	2D: Moreau-Broto autocorrelation of lag 8 weighted by Sanderson electronegativity
AATSC5d	2D: averaged and centered Moreau-Broto autocorrelation of lag 5 weighted by σ electrons
MATS5i	2D: Moran coefficient of lag 5 weighted by ionization potential
GATS1i	2D: Geary coefficient of lag 1 weighted by ionization potential
RPCG	2D: relative positive charge
CIC3	2D: 3-ordered complementary information content



Descriptor name	Linear coefficient	Descriptor type
eccentricity	2034.49	eccentricity
ATSC3dv	-1.34582	2D: centered Moreau-Broto autocorrelation of lag 3 weighted by valence electrons
ATSC1d	137.0966	2D: centered Moreau-Broto autocorrelation of lag 1 weighted by σ electrons
RASA	2586.248	3D: relative hydrophobic surface area
MAXssO	-827.018	2D: max of ssO
JGI8	6779.873	2D: 8-ordered mean topological charge
Linear offset	-1440.08	

Figure S6.3. An example of fitting cycling stability metrics (N_{90}, *empty circles on the right*) using general-use molecular descriptors (for the same set of 20 molecules shown in Figure S6.3) with r^2 = 0.935. To achieve this confidence using MMLR, it was necessary to use at least six descriptors (see the table below the graph). Not only does this result in poor predictability but also most of these descriptors have no meaningful connection to physical and chemical properties of the radical cations.

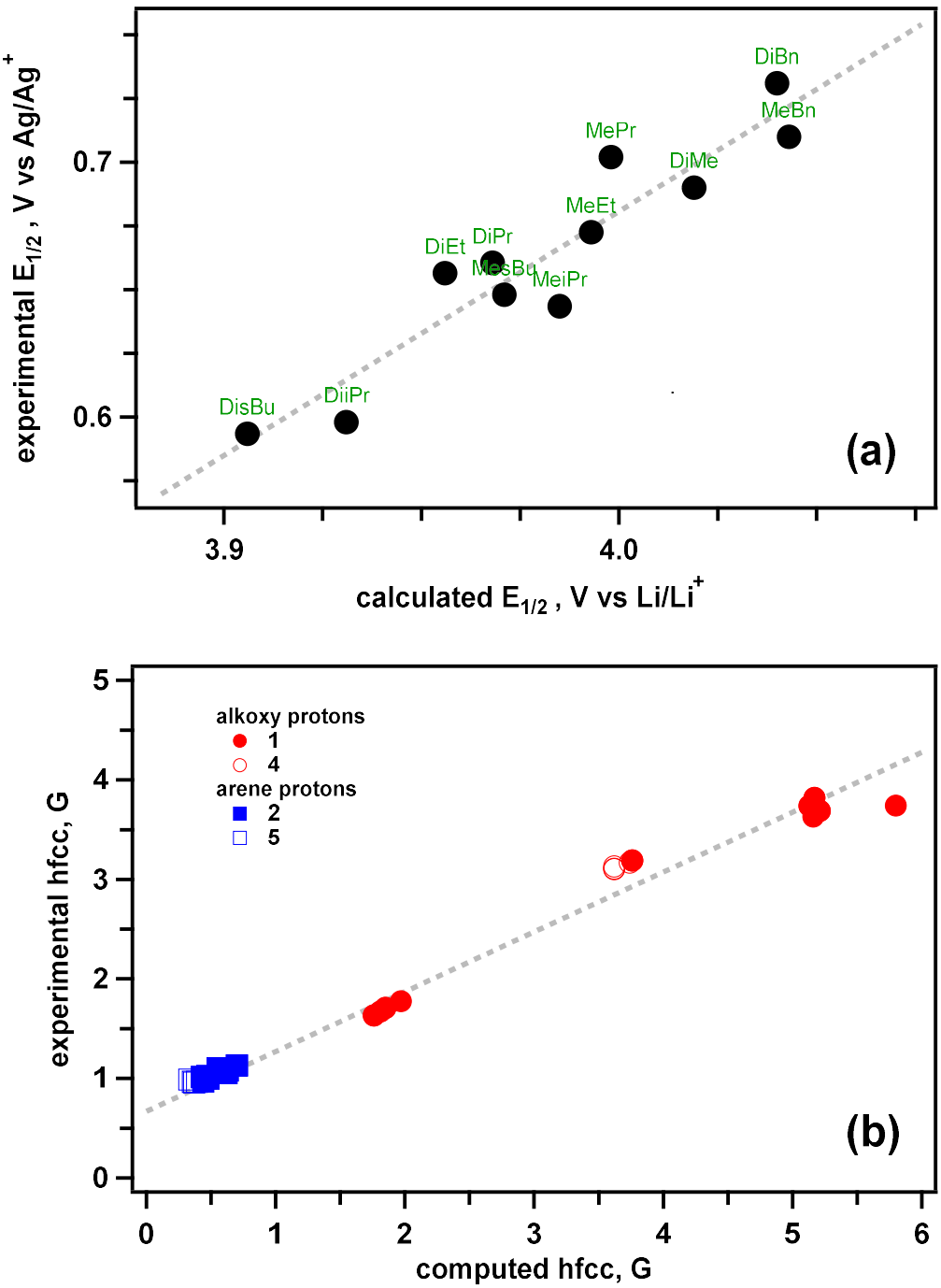


Figure S6.4. (a) Linear correlation of the experimental and computed redox potentials $E_{1/2}$ for the dialkoxyarenes shown in Table 1 (r^2 = 0.885). (b) The same for the experimental and computed hyperfine coupling constants (hfcc's) in the alkoxy α-protons (*red*) and arene ring protons (*blue*) for the radical cations of these dialkoxyarenes. The positions in the ring are indicated in the plot.

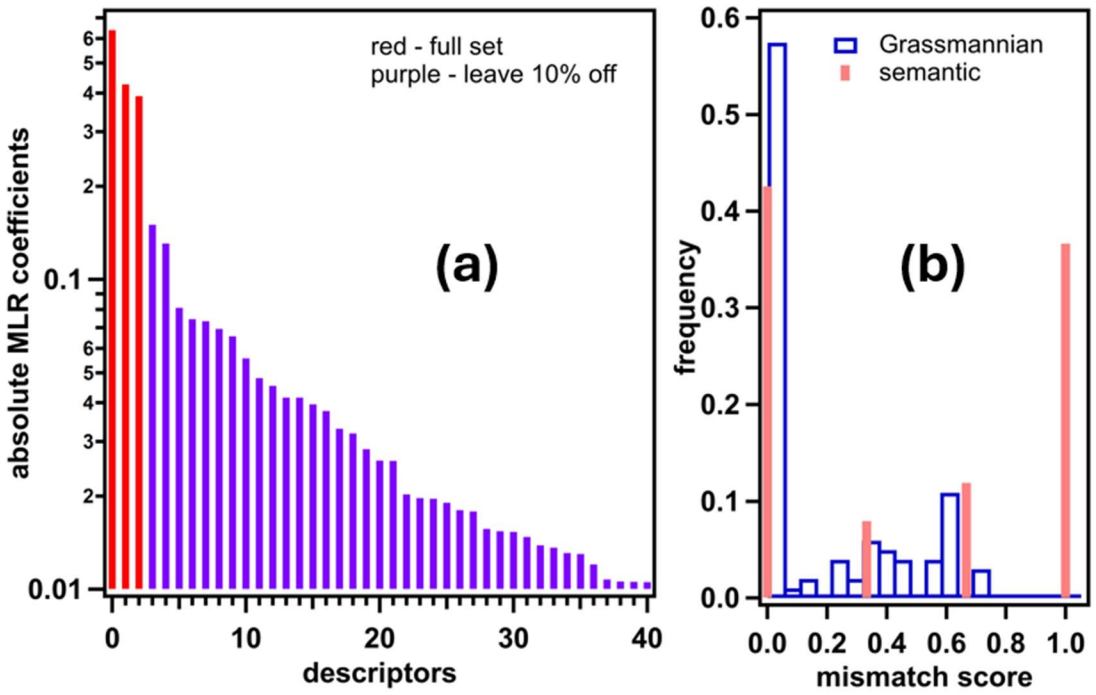


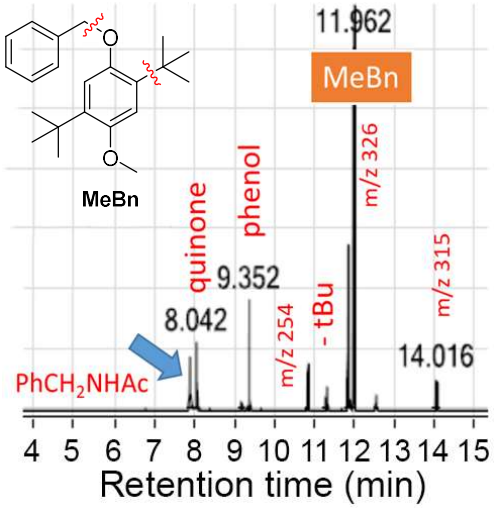
Figure S6.5. Cross-validation of the MMLR–GA fit for log($t_{1/2}$) in Figure 5b (*main text*) using random 90% data subsampling over 200 trials. (a) The absolute MLR coefficients for the reduced optimized descriptors obtained across the entire trial series are plotted in decreasing order; the original descriptors for the full data set are indicated in red. It is seen that these full-set descriptors are indeed the most important for this data set, as they are consistently selected across all subsets of incomplete data. (b) Histograms of the descriptor-set mismatch scores defined in the section above. The semantic score is shown in red and the Grassmannian score (which accounts for linear dependence between descriptors) is shown in blue. While the descriptors may differ semantically, with 50% of the sets showing complete mismatch, the linear spaces defined by these descriptor sets are largely identical. Thus, the descriptors obtained from the full data set adequately represent the variance in the data.

S7. Analysis of decay products

The products of H-cell electrolysis (Section S4) in both compartments were analyzed using GCMS and ^{1}H NMR spectroscopy. For GCMS analyses, the solvent was removed in vacuum, water was added, and reaction products and unreacted redoxmers were extracted into dichloromethane. The organic layer was twice washed with water to remove the remaining LiTFSI salt. **Figures S7.1 and S7.2** below show examples of the GCMS data. For these analyses, 1 μL liquid sample was loaded on an HP-5MS (bore 0.25 μm, length 30 m) column using an Agilent Technologies Model 7890B chromatograph equipped with a Model 5977 mass detector. The typical program include a 15-min hold at 30 °C, a 15-min hold at 50 °C, followed by a 20 °C/min ramp to 250 °C followed by another 20 min hold.

For high-performance liquid chromatography (HPLC) analyses with the light absorption detection, the samples were diluted 1:20 v/v with acetonitrile. The degree of decomposition was determined by comparison of the peaks of the dialkoxyarene compound in the degraded material with the solution before electrolysis. For these analyses, isocratic elution of the reaction mixture (5 μL aliquot) at 0.25 mL/min on Supelco LC-PAH column length 25 mm, bore 4.6 mm, particle size (5 μm); the absorbance was detected using a photodiode array detector using a ThermoScientific Accela suite. The analysis took 20 min/sample. This method was used to quantify the depletion of the active material.

Additional fractioning of the extracted materials was sometimes necessary due to presence of multiple products. Short silica columns were used to this end with dichloromethane serving as an eluent; acetonitrile was used for column stripping. The organic solvent was removed in vacuum, and the products were analyzed by NMR and GCMS in $CDCl_3$ or CD_3CN (see **Figure S7.3** for an example). As these procedures can change the composition of the constituents, we also diluted electrolyte with 1:10 *v*/*v* CD_3CN and examined these solutions with ^{1}H NMR and HPLC.



Peak	Start	Ret. time	End	Height	Area	Area %
1	7.83	7.87	7.945	177275.85	325824.94	17.01
2	8.019	8.042	8.802	222779.85	192485.49	10.05
3	9.33	9.352	9.427	363643.85	314209.25	16.4
4	10.789	10.812	10.852	156795.85	145929.09	7.62
5	11.206	11.241	11.315	43571.85	120838.91	6.31
6	11.767	11.819	11.842	552187.85	725257.3	37.85
7	11.842	11.859	11.91	38403.85	59477.55	3.1
8	11.91	11.962	12.03	1347899.85	1915949.19	100
9	12.471	12.511	12.574	50371.85	77172.55	4.03
10	13.964	14.016	14.125	102107.85	174833.26	9.13

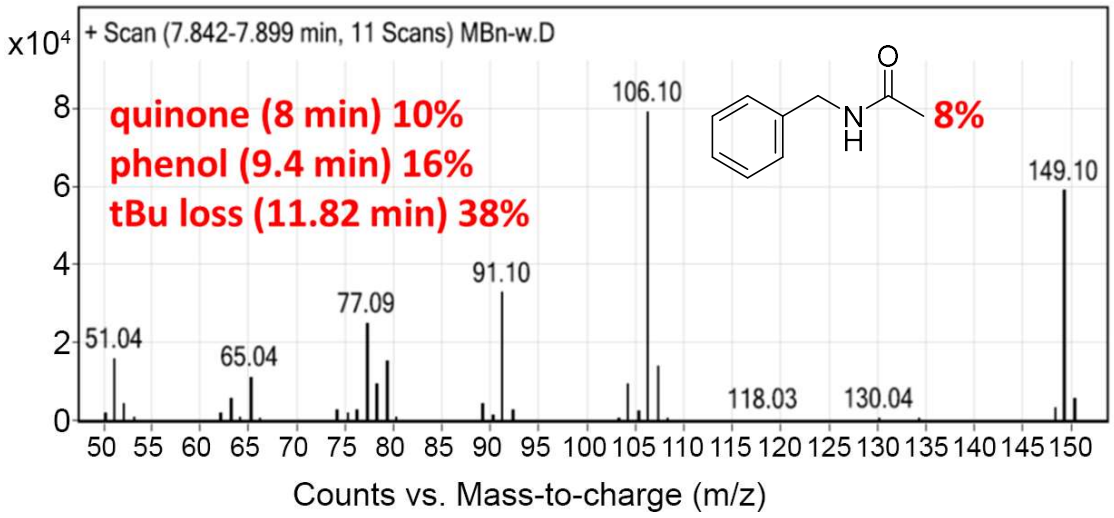
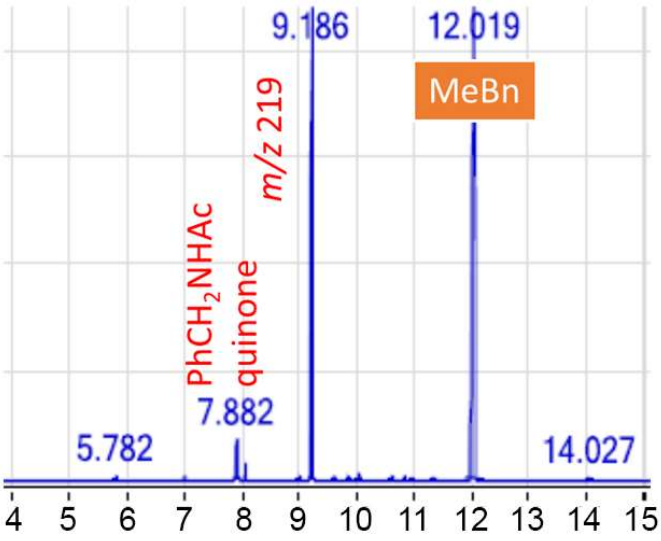
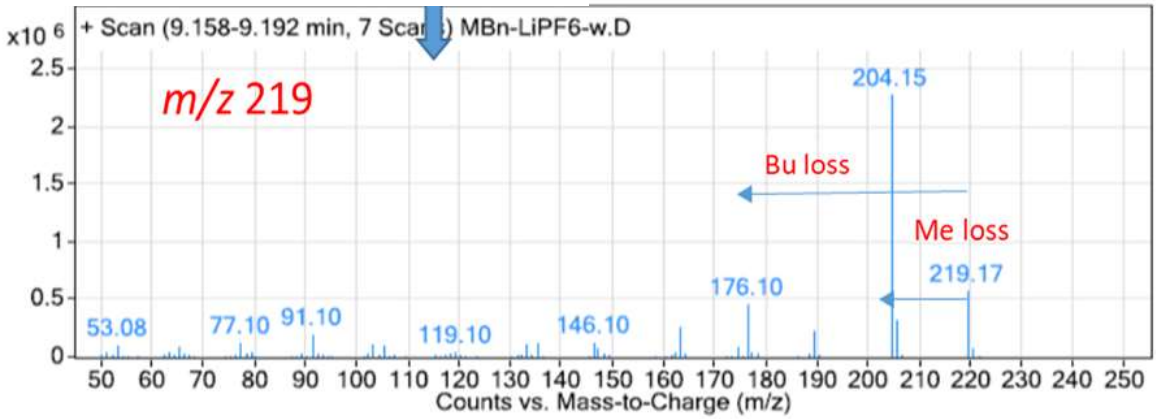


Figure S7.1. Representative GCMS analysis of volatile products generated in electrolysis of 20 mM **MeBn** in 0.5 M LiTFSI (5C charge to 100% SOC; working electrode chamber). The peaks from the phenol **(4** in Figure 4), quinone (**5** in Figure 4), *N*-benzylacetamide (*indicated with the arrow*), benzoxazole (**6** in Figure 4 with *m/z* 295), and *tert*-butyl loss products (**3** in Figure 4) are seen in the chromatograms. $PhCH_2NHAc$ is the product of hydrolysis of the nitrenium cation generated when the leaving benzyl carbocation reacts with the solvent. The mass spectrum of $PhCH_2NHAc$ is shown separately.



Peak	Start	Ret. time	End	Height	Area	Area %
1	5.761	5.782	5.822	46204.8	36383.34	0.25
2	6.961	6.984	7.004	35062.71	32454.08	0.22
3	7.836	7.882	7.956	383760.78	755899.34	5.11
4	8.019	8.042	8.059	163021.93	138146.64	0.93
5	8.946	8.975	9.015	54304	48056.62	0.33
6	9.141	9.186	9.227	4405505.77	5600487.98	37.89
7	9.553	9.576	9.627	36400	39702.14	0.27
8	9.799	9.816	9.879	57408	51974.51	0.35
9	9.976	10.016	10.056	58552	57329.03	0.39
10	10.554	10.588	10.623	48723.02	44987.95	0.3
11	10.789	10.812	10.84	34935.08	31707.3	0.21
12	10.909	10.932	10.972	20764.96	18543.56	0.13
13	11.269	11.309	11.361	28870.3	47448.78	0.32
14	11.887	12.019	12.053	4453642.51	14782209.35	100
15	12.093	12.122	12.179	16639.96	25758.82	0.17
16	13.982	14.027	14.107	26684.39	46193.5	0.31



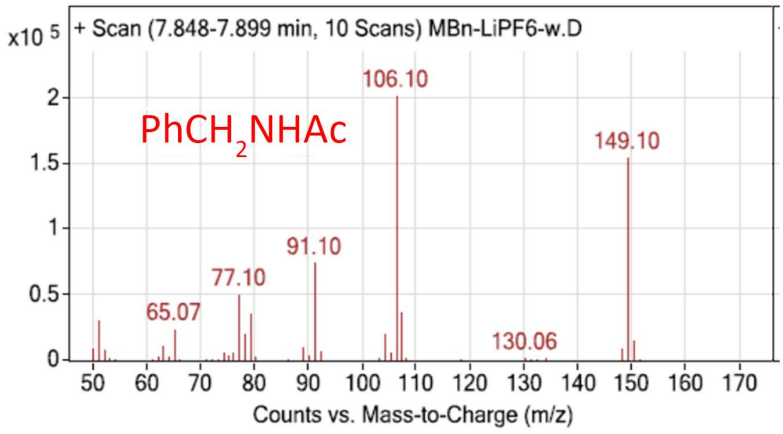


Figure S7.2. GCMS analysis of volatile products generated in electrolysis of 20 mM **MeBn** in 0.5 M $LiPF_6$ (5C charge to 100% SOC; working electrode chamber). The peaks from quinone (**5** in Figure 4), *N*-benzylacetamide, and product with *m/z* 219 (whose MS spectra are shown separately and identified with **6** in Figure 4) are indicated. The proton NMR spectrum of the latter product is shown separately in Figure S7.3.

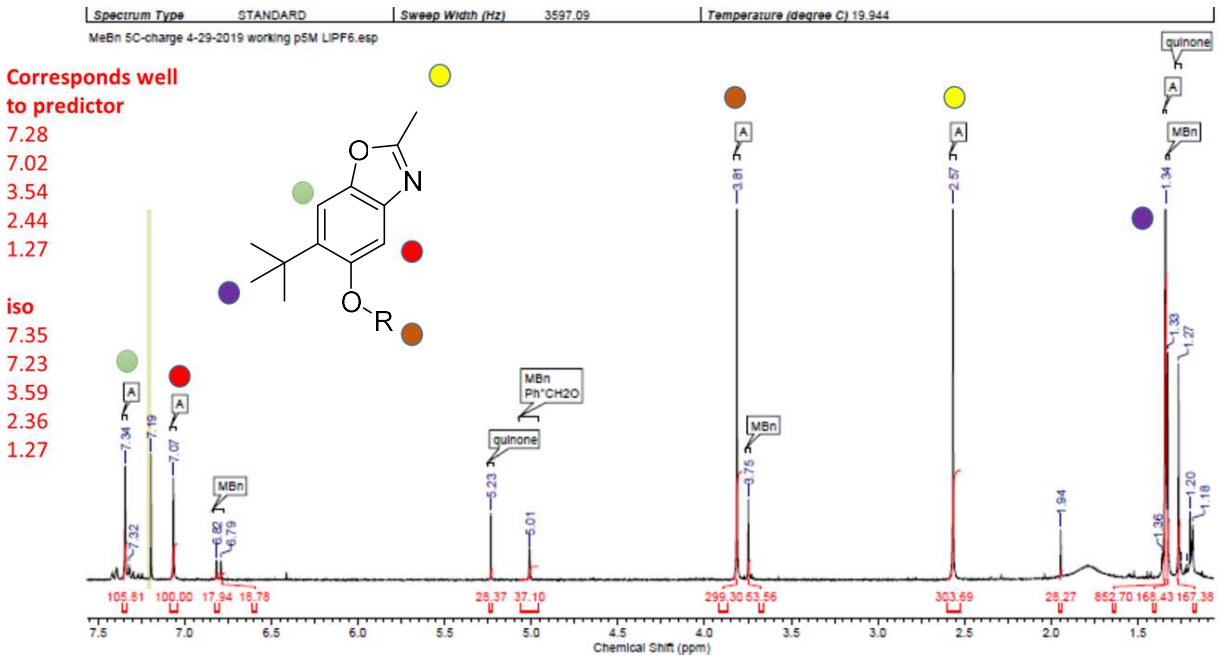


Figure S7.3. ^{1}H NMR spectrum (300 MHz) in $CDCl_3$ of chromatographically isolated *m/z* 219 product (**6** in Figure 4) generated in electrolysis of **MeBn** in acetonitrile (20 mM, 0.5 M $LiPF_6$, full charge at 5C). The product was isolated by fractioning on a silica gel column using dichloromethane as eluent. Resonance lines from the parent compound and quinone are also visible in this spectrum. The resonances of this compound (δ, ppm vs tetramethylsilane: 7.34 (1H, Ar), 7.07 (1H, Ar), 3.81(3H, -OCH_3), 2.57 (3H, -CH_3), 1.34 ppm (9H, *tert*-Bu) correspond well to the ones estimated using NMR predictor (http://www.nmrdb.org/new_predictor, *on the left side*), though there is some uncertainty whether it is a benzoxazole (as shown in the plot) or an indoxazine.

S8. ^{1}H and ^{13}C NMR spectra of new compounds

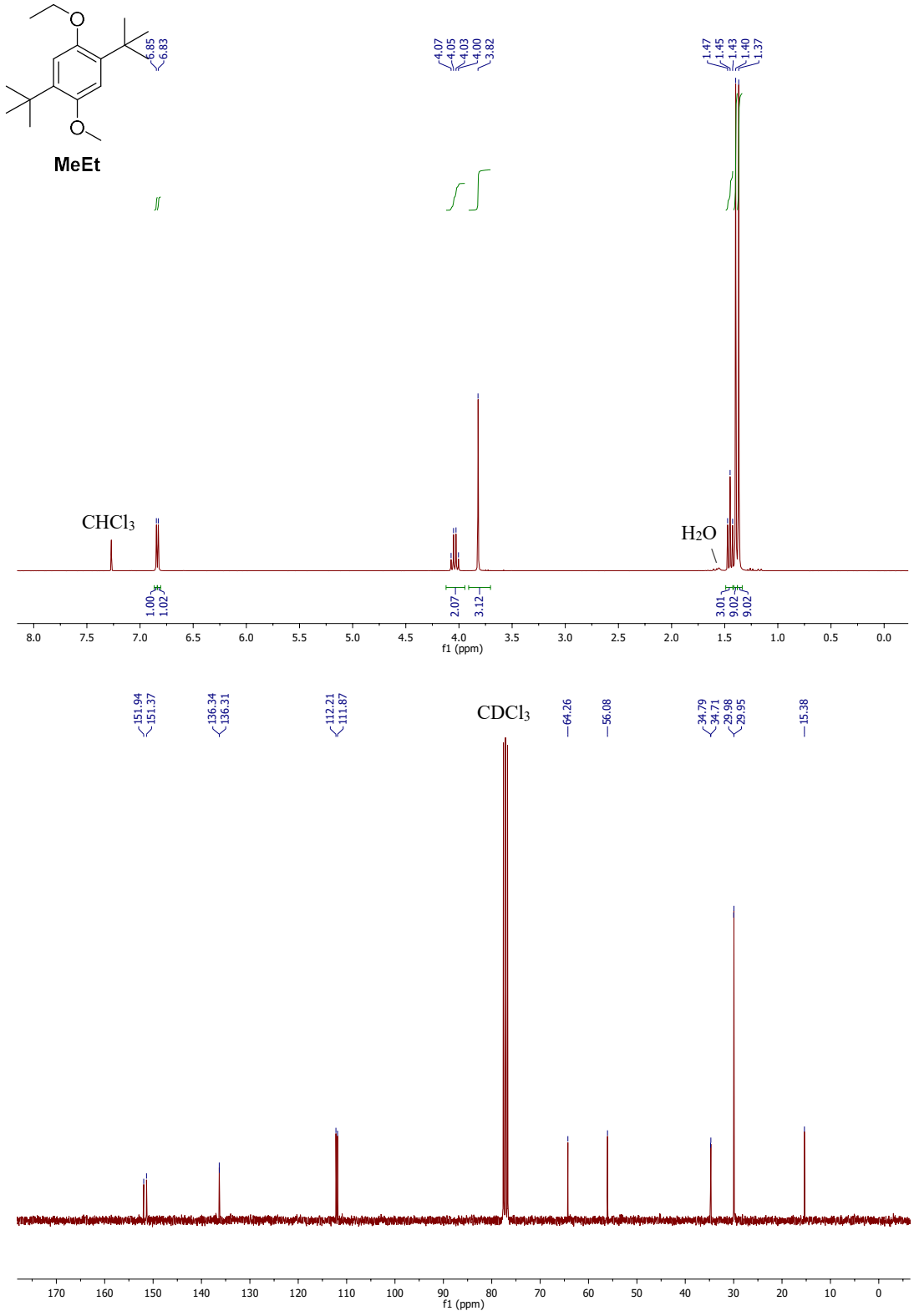


Figure S8.1 ^{1}H and ^{13}C NMR spectra of **MeEt** in $CDCl_3$ (300 and 75 MHz, respectively).

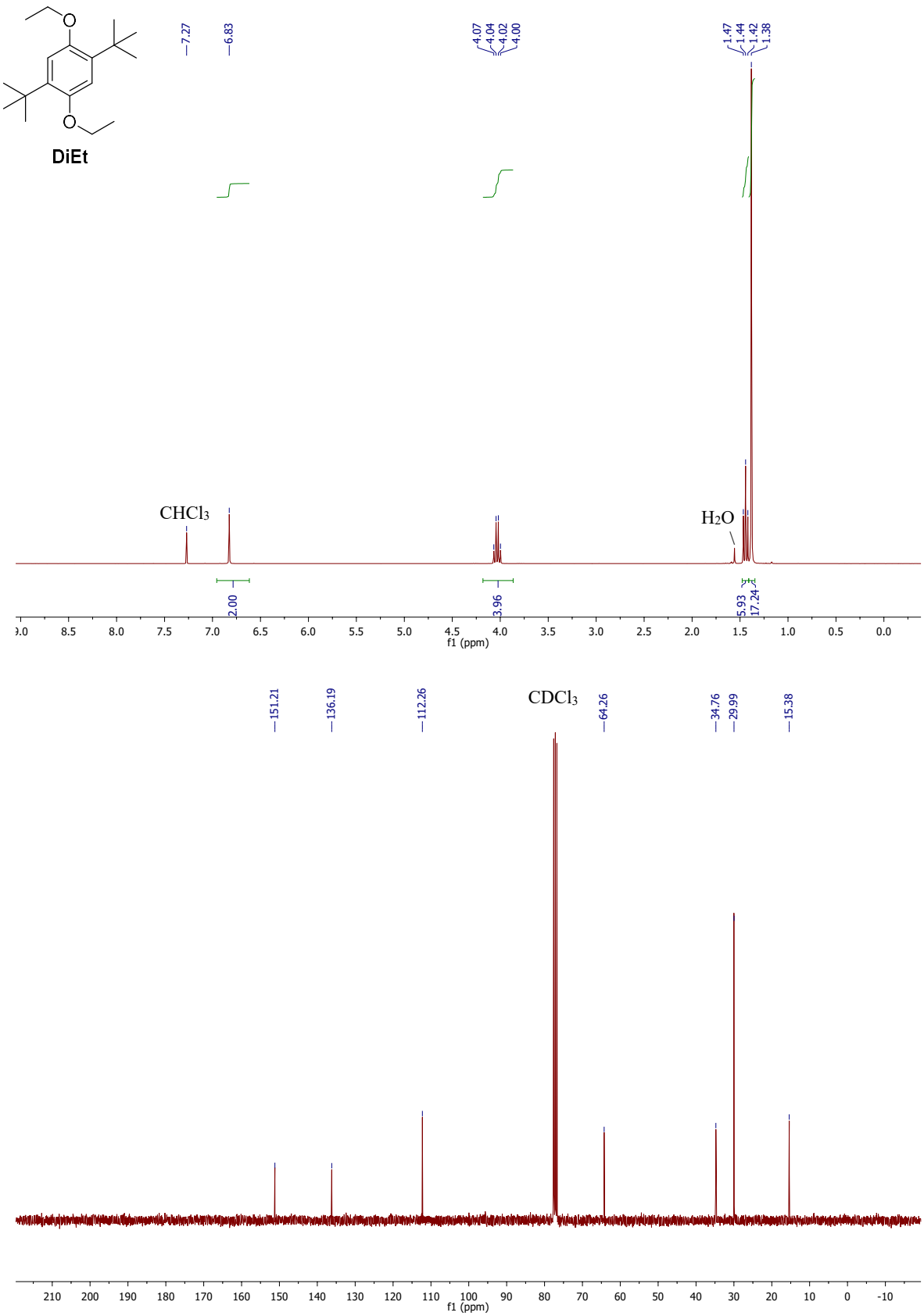


Figure S8.2. ^{1}H and ^{13}C NMR spectra of **DiEt** in $CDCl_3$ (300 and 75 MHz, respectively).

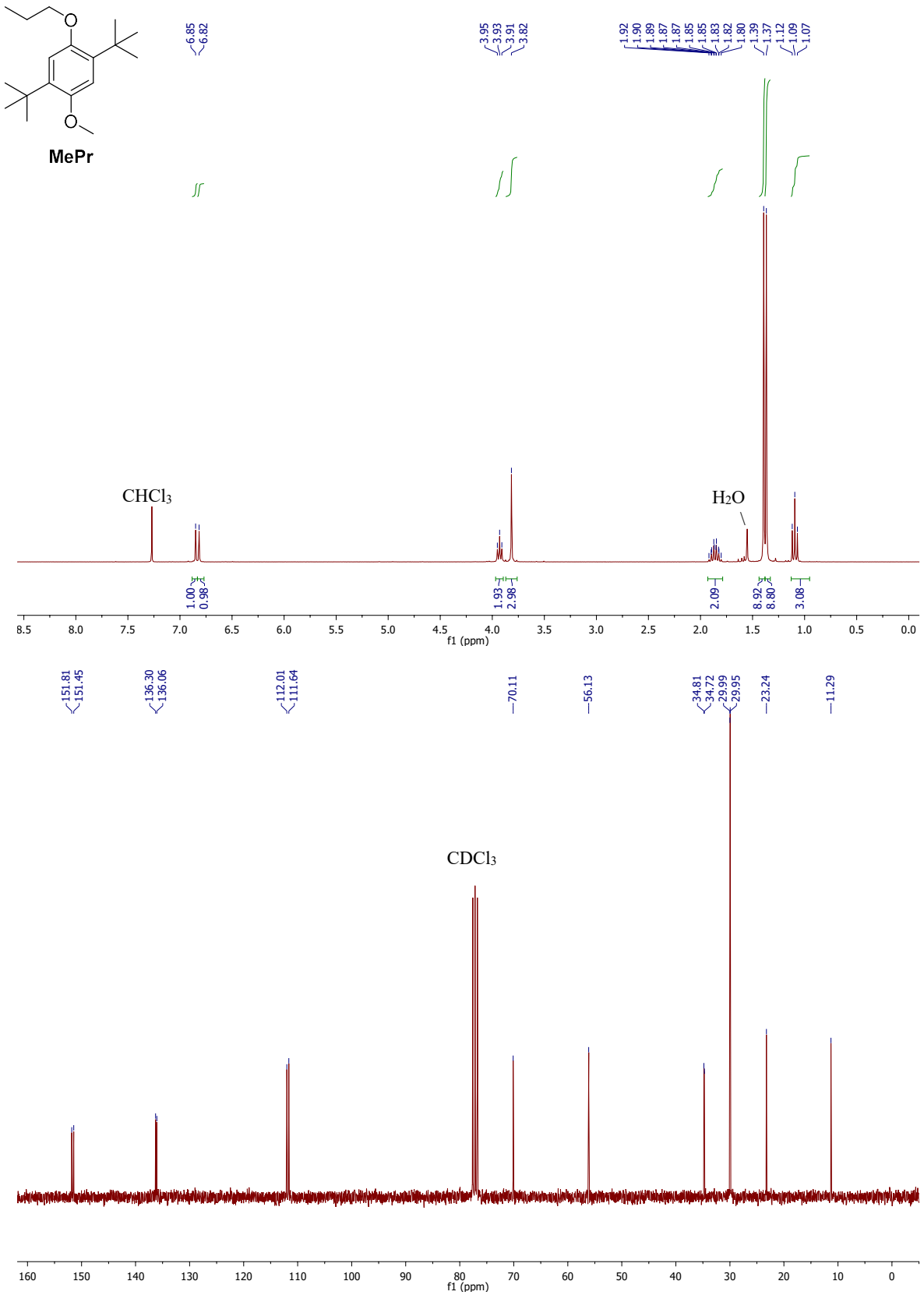


Figure S8.3. ^{1}H and ^{13}C NMR spectrum of **MePr** in $CDCl_3$ (300 and 75 MHz, respectively).

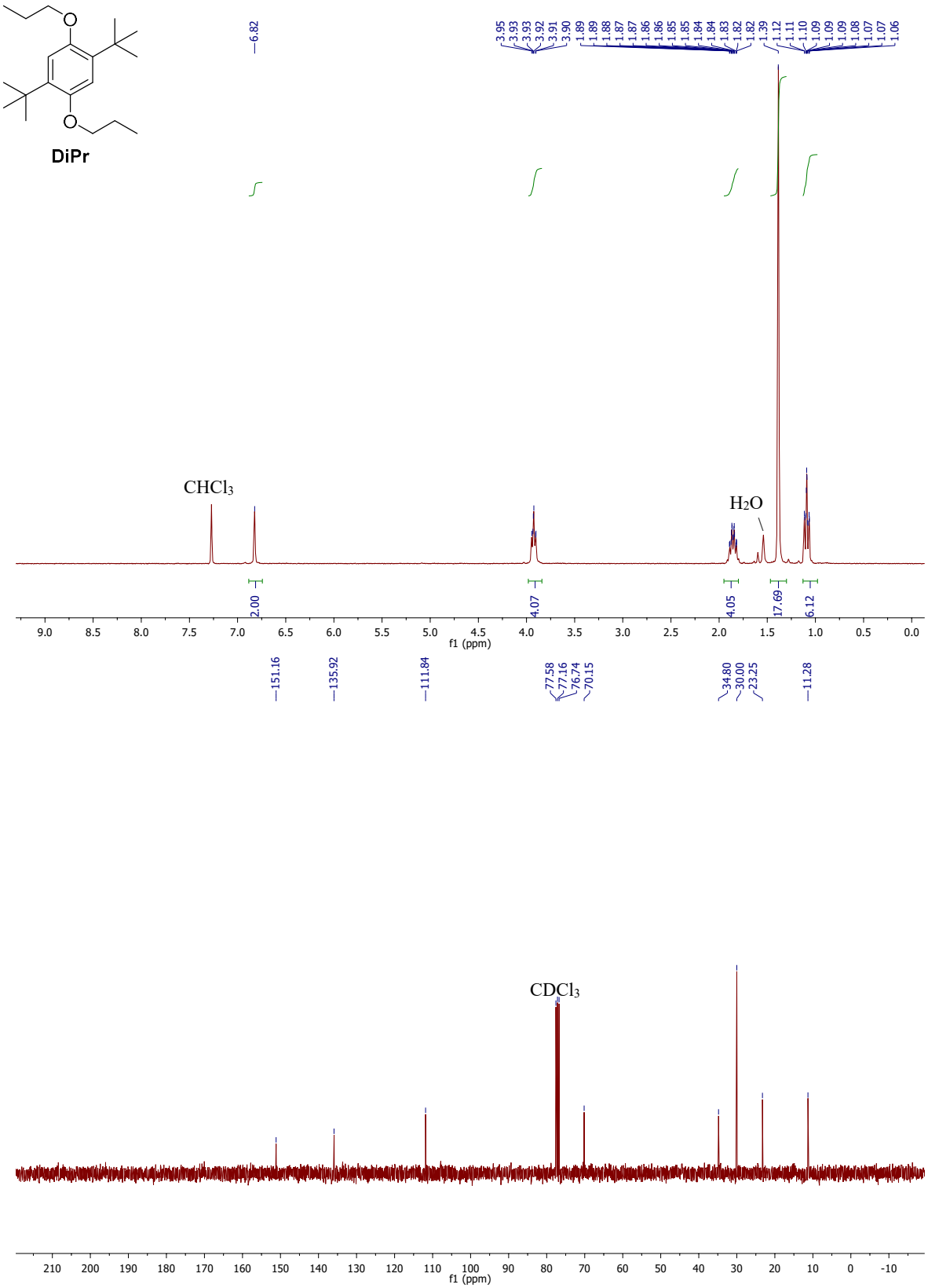


Figure S8.3. ^{1}H and ^{13}C NMR spectrum of **DiPr** in $CDCl_3$ (300 and 75 MHz, respectively).

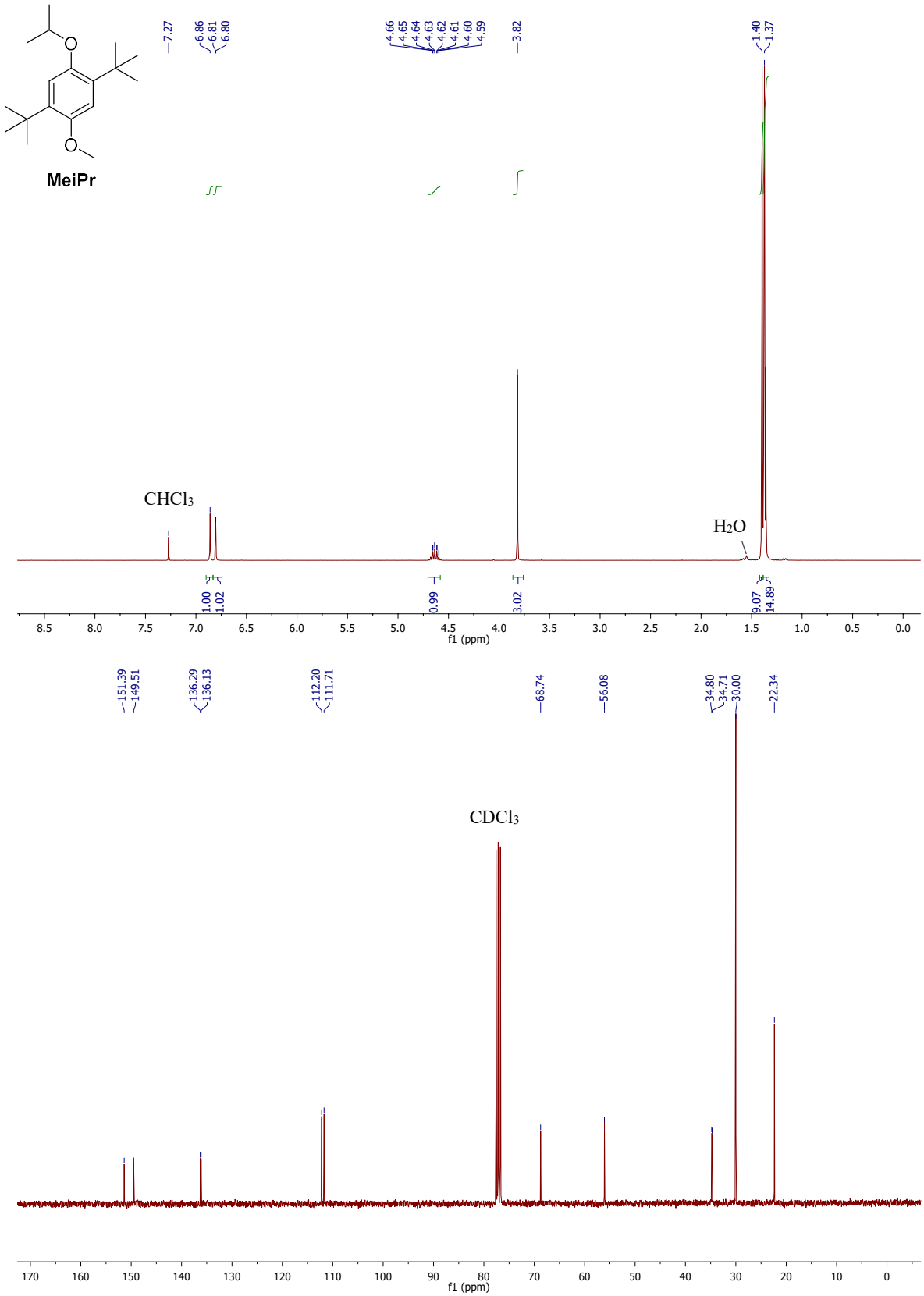


Figure S8.4. ^{1}H and ^{13}C NMR spectrum of **MeiPr** in $CDCl_3$ (300 and 75 MHz, respectively).

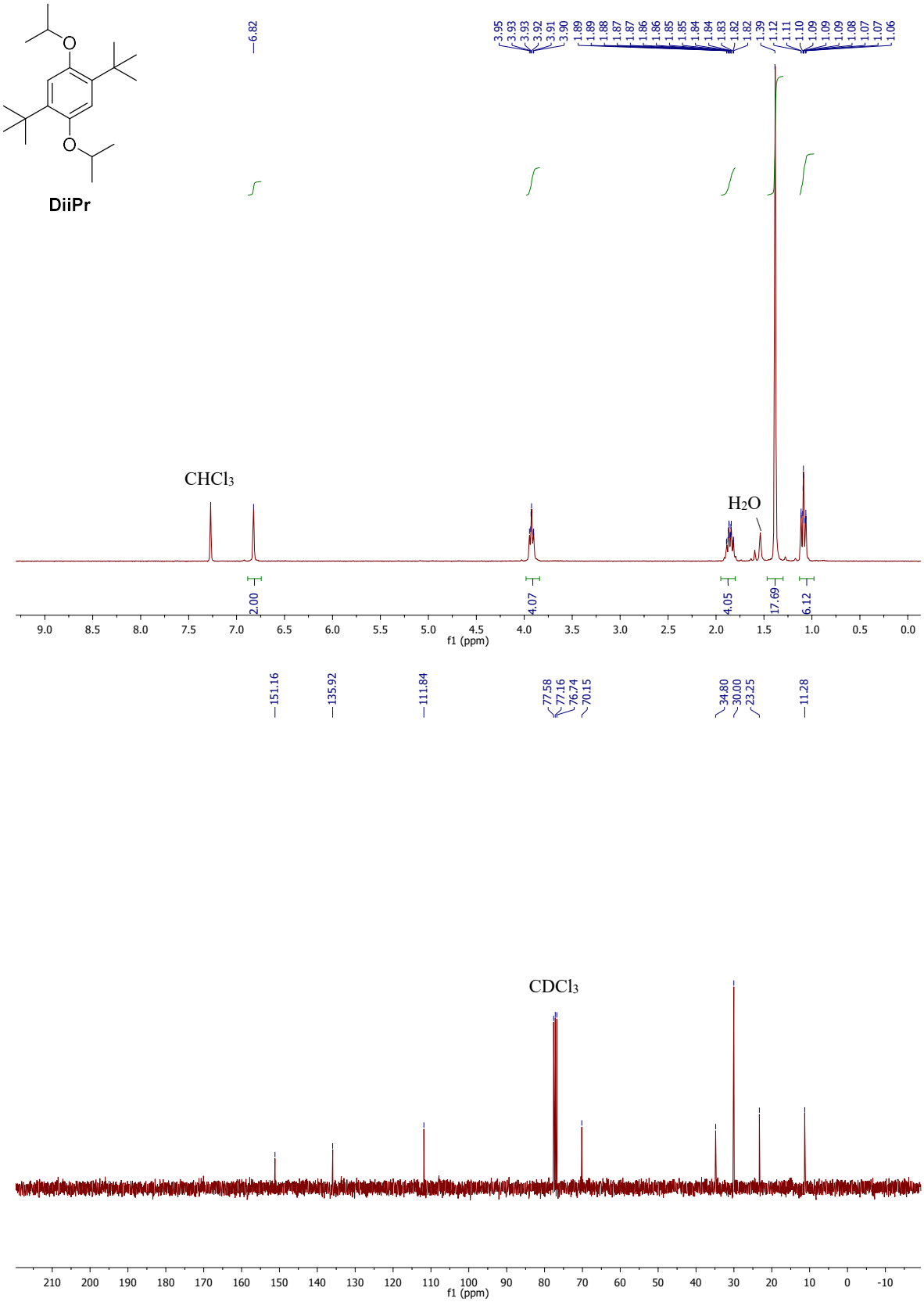


Figure S8.5. ^{1}H and ^{13}C NMR spectrum of **DiiPr** in $CDCl_3$ (300 and 75 MHz, respectively).

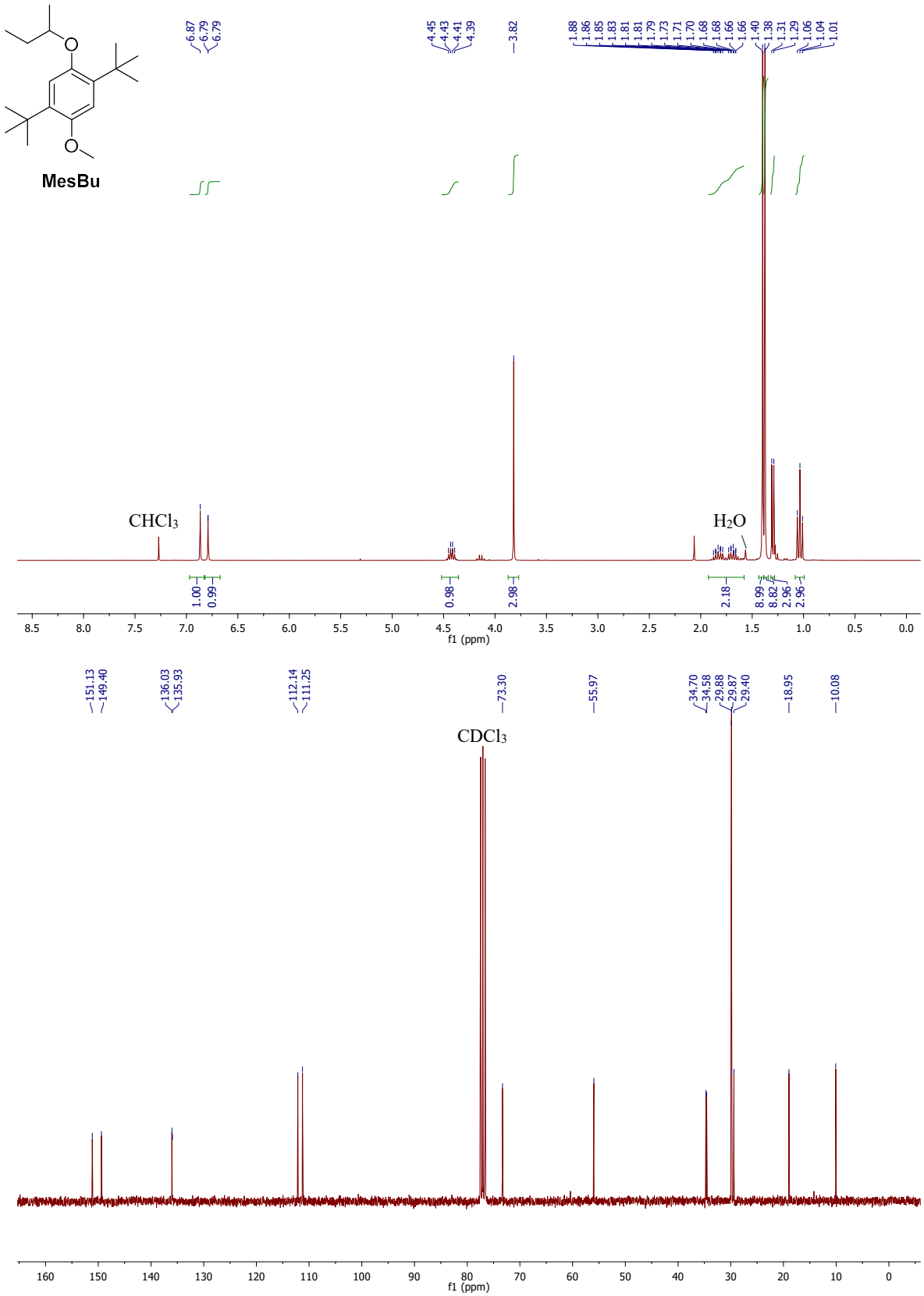


Figure S8.6. ^{1}H and ^{13}C NMR of **MesBu** in $CDCl_3$ (300 and 75 MHz, respectively).

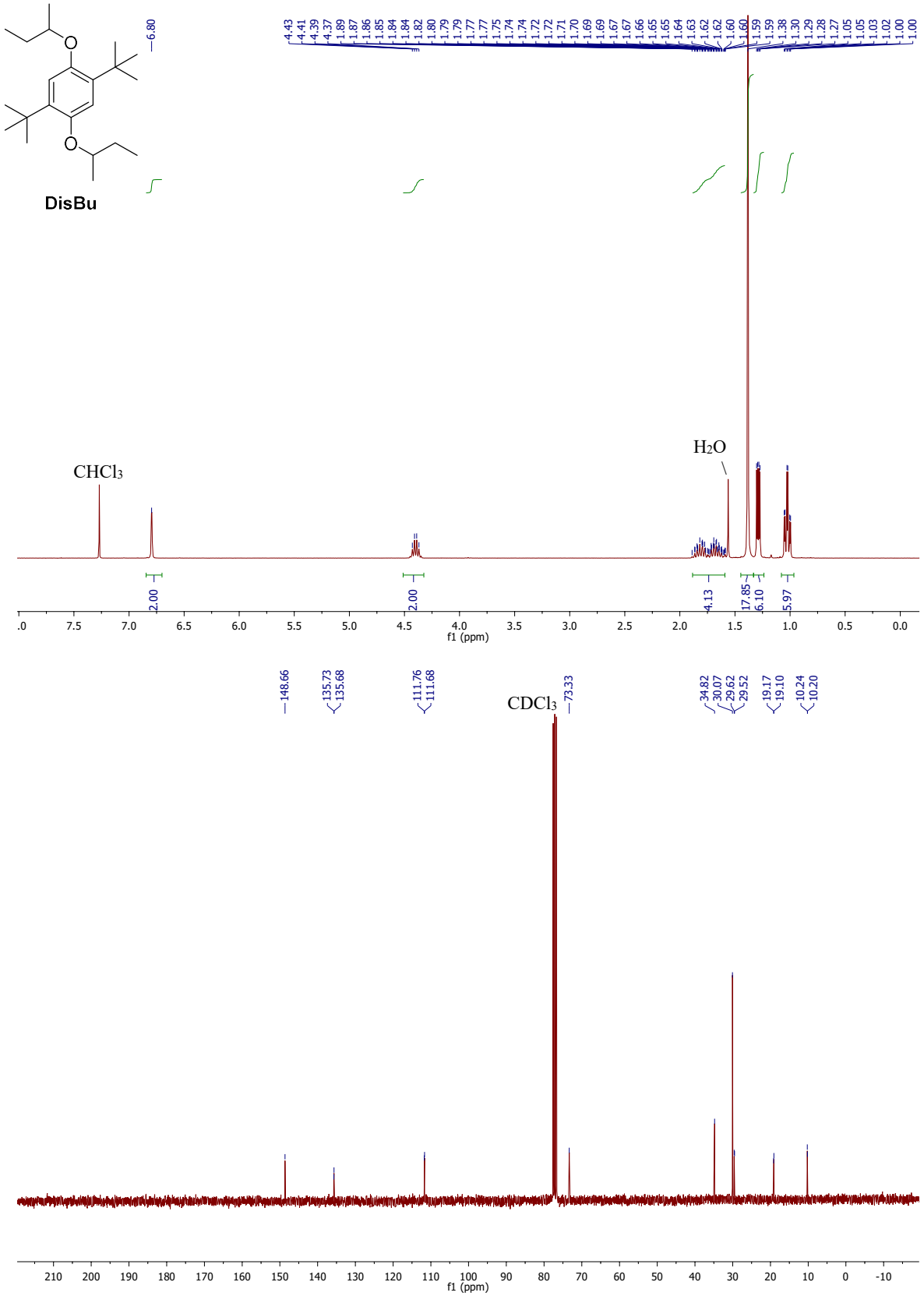


Figure S8.7. ^{1}H and ^{13}C NMR spectrum of **DisBu** in $CDCl_3$ (300 and 75 MHz, respectively).

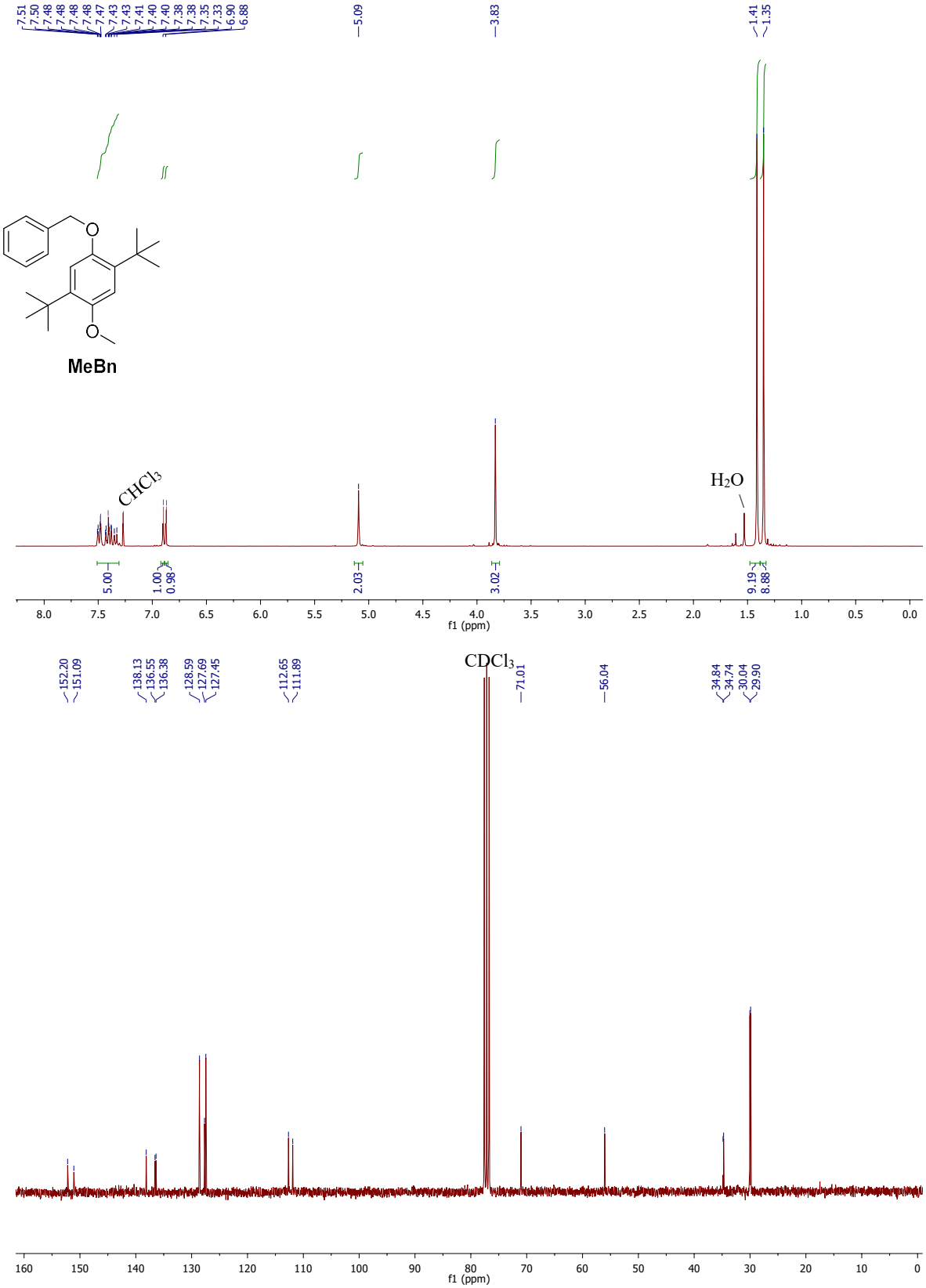


Figure S8.8. ^{1}H and ^{13}C NMR of **MeBn** in $CDCl_3$ (300 and 75 MHz, respectively).

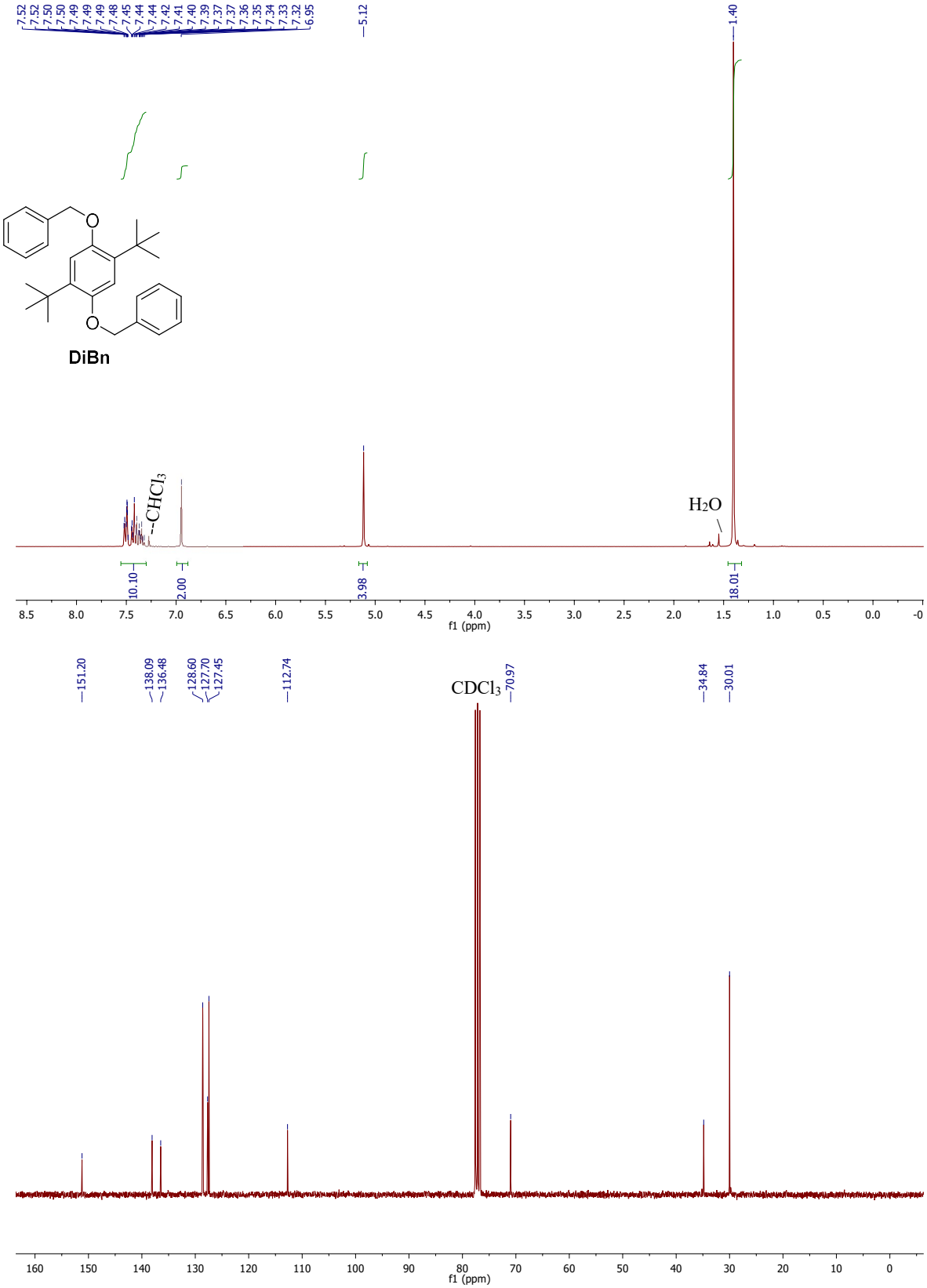


Figure S8.9. ^{1}H and ^{13}C NMR of **DiBn** in $CDCl_3$ (300 and 75 MHz, respectively).

S9. HPLC spectra of new compounds

Note, MePr and DiPr were not validated by HPLC due to limited sample remaining.

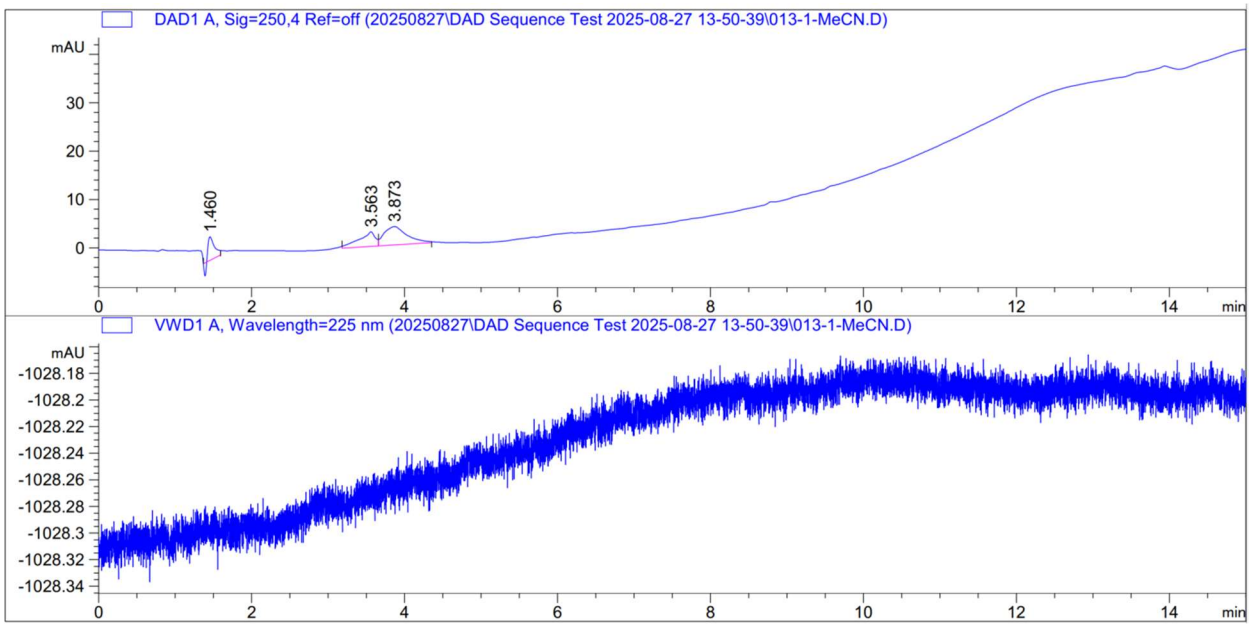


Figure S9.1. Blank spectra for HPLC detectors, diode array detector (DAD, *top*) and variable wavelength detector (VWD, *bottom*).

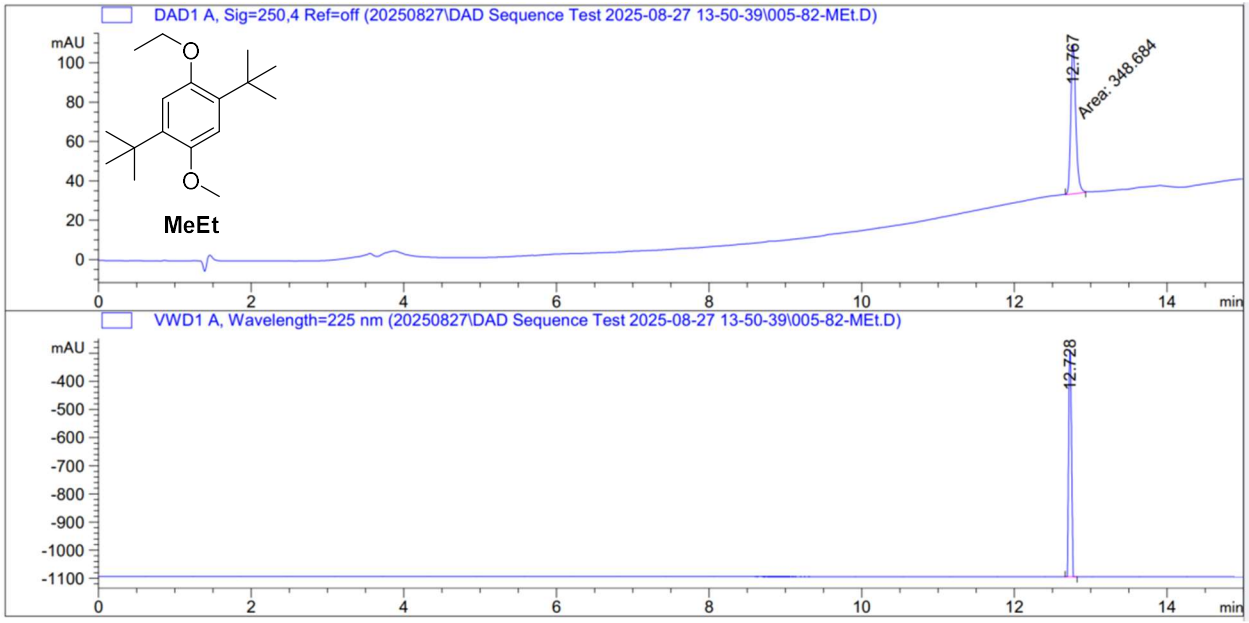


Figure S9.2. HPLC spectra for **MeEt**.

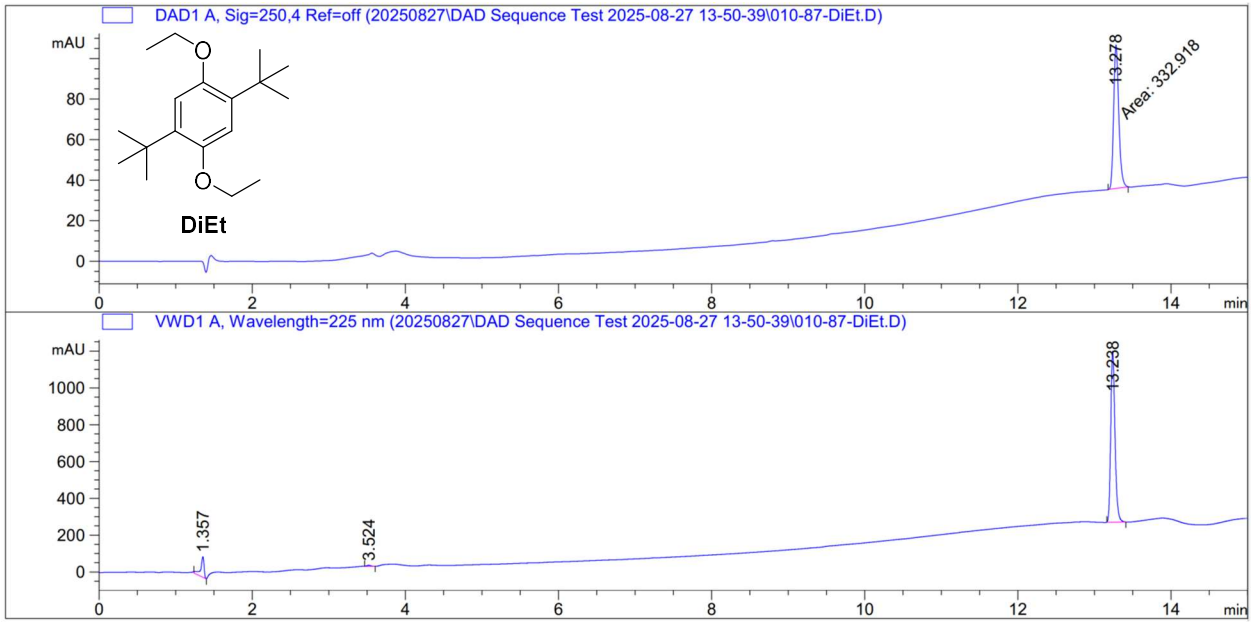


Figure S9.3. HPLC spectra for **DiEt**.

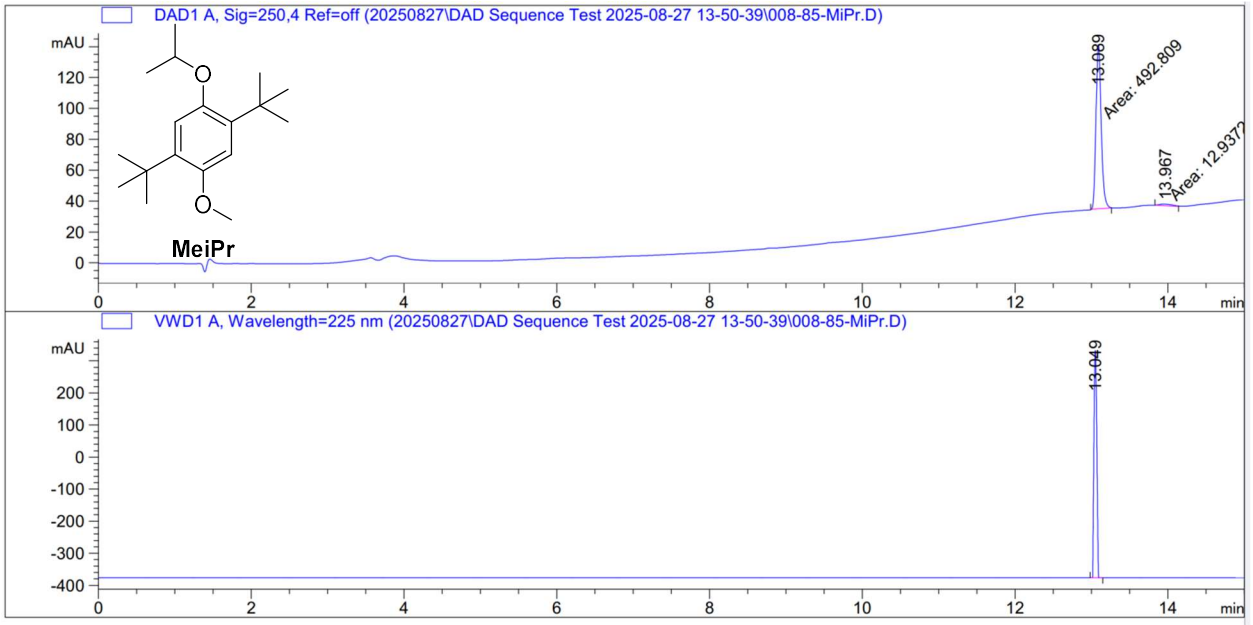


Figure S9.4. HPLC spectra for **MeiPr**.

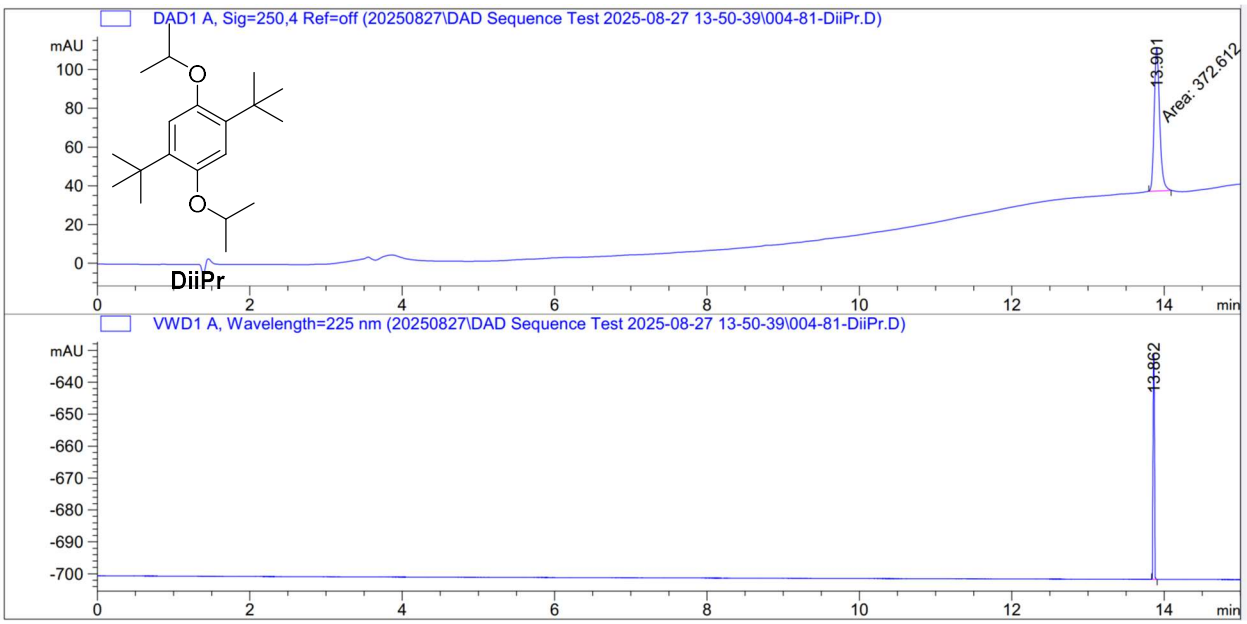


Figure S9.5. HPLC spectra for **DiiPr**.

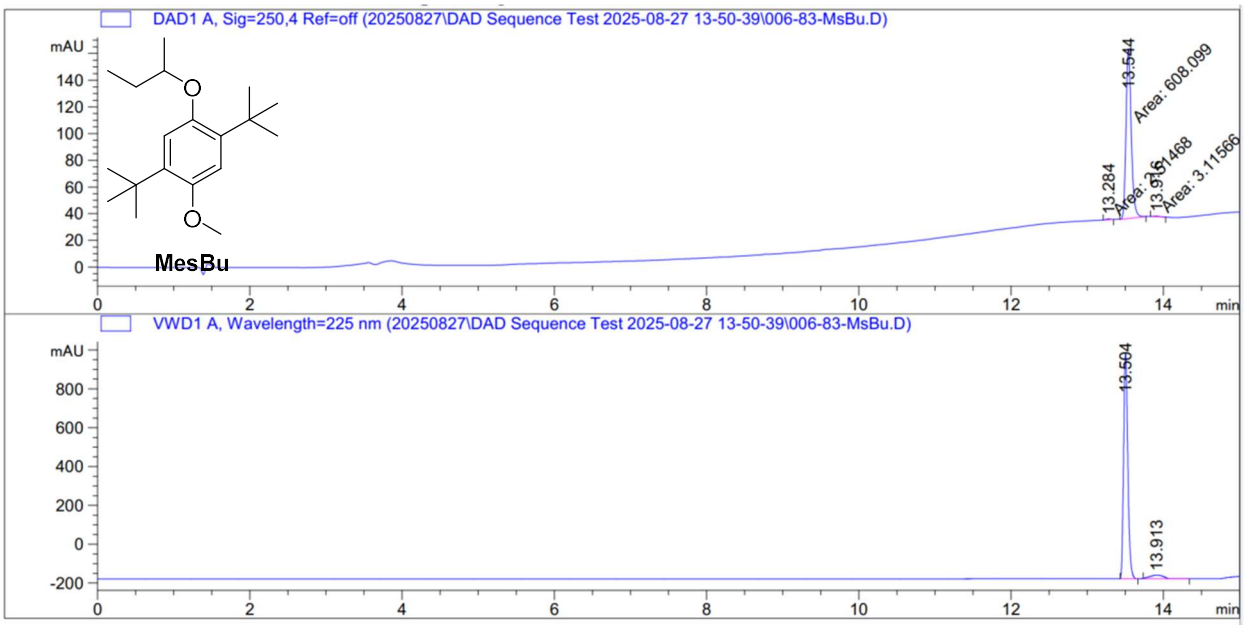


Figure S9.6. HPLC spectra for **MesBu**. For the DAD, the primary peak integrates to 99.1% while for the VWD, it integrates to 95.0%.

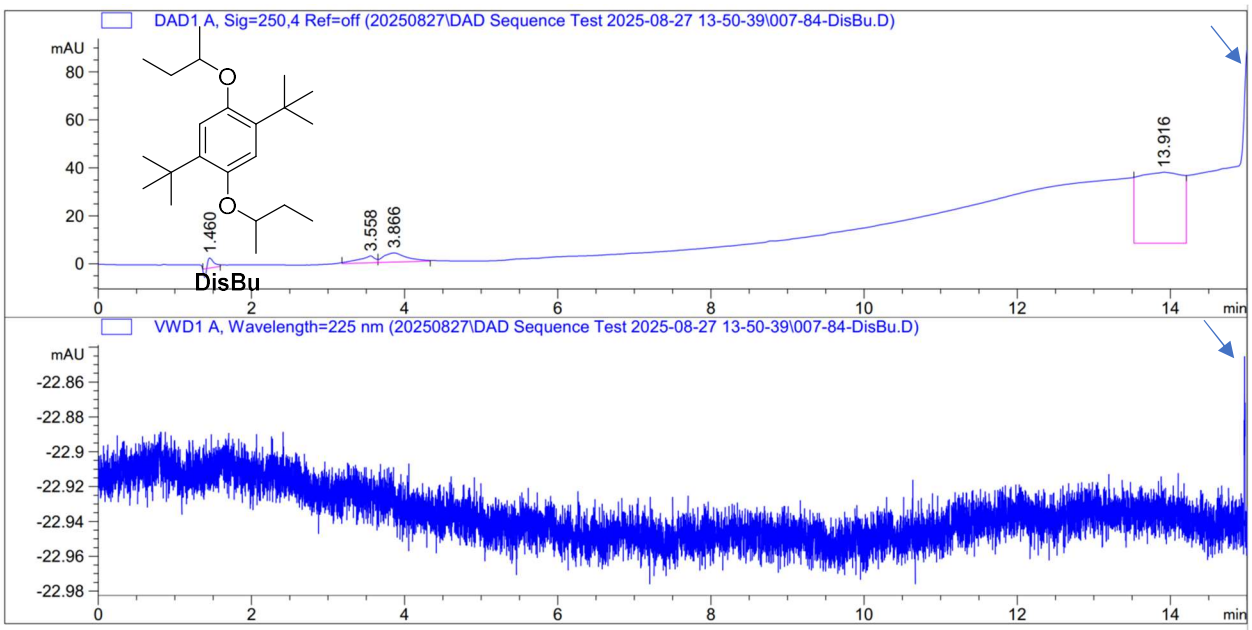


Figure S9.7. HPLC spectra for **DisBu**. Note, the primary elution peak is on the edge of the detection limit for both detectors as indicated by the arrows.

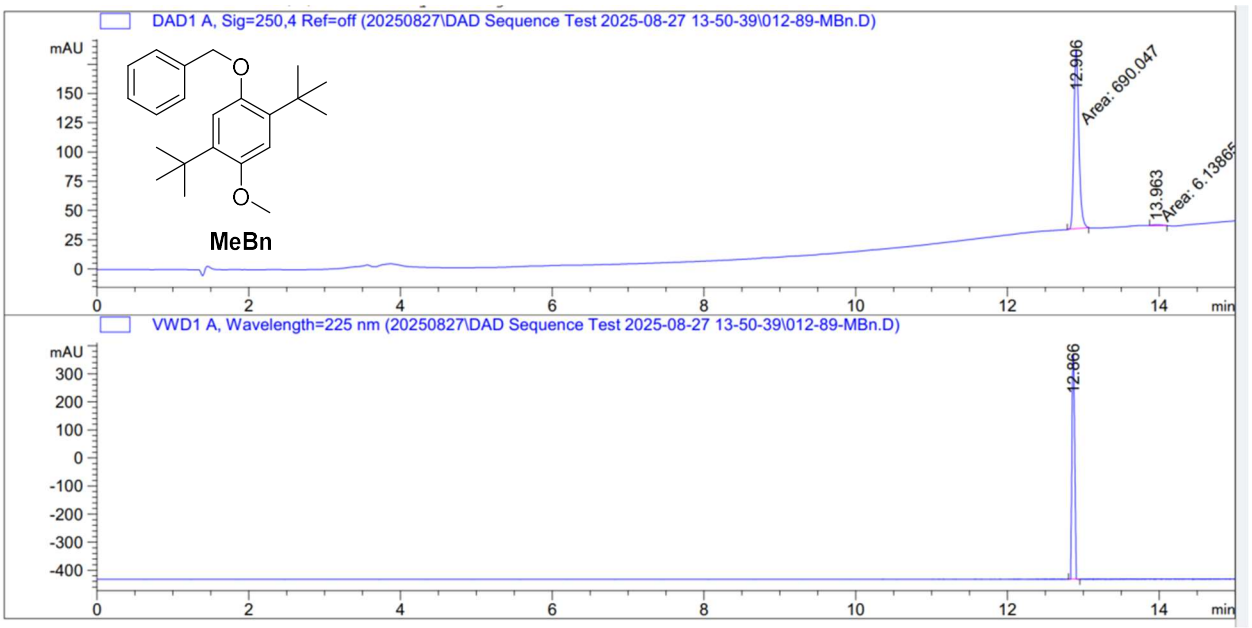


Figure S9.8. HPLC spectra for **MeBn**. For the DAD, the primary peak integrates to 99.1% while for the VWD, it integrates to 100.0%.

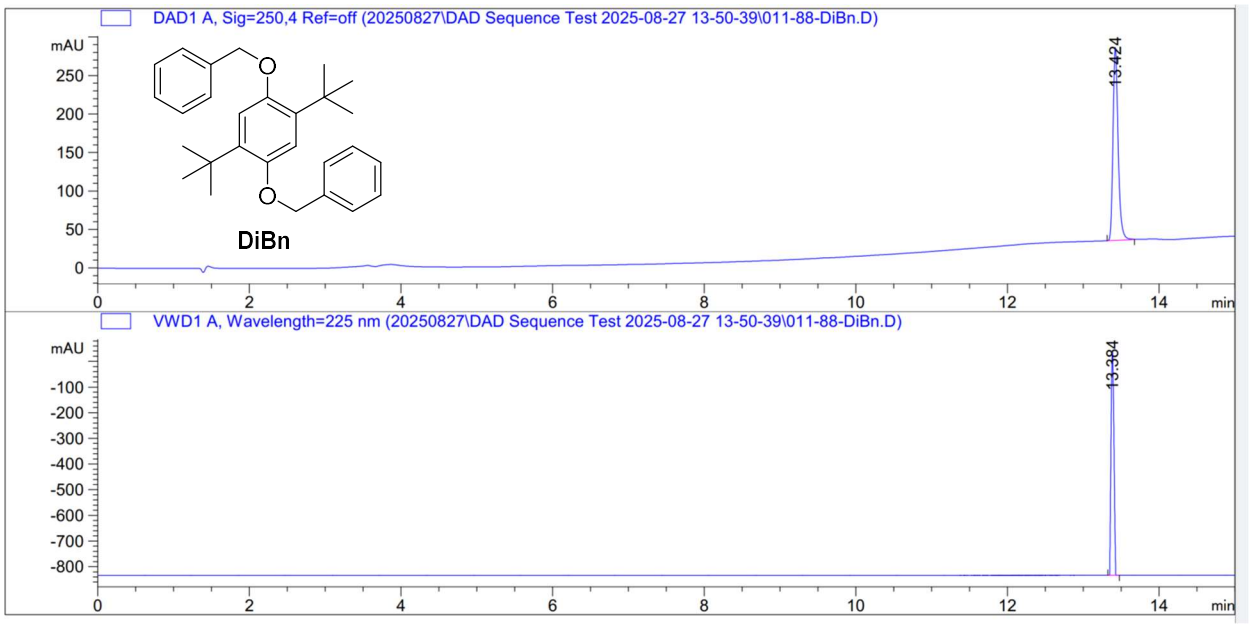


Figure S9.9. HPLC spectra for **DiBn**.

S10. Supporting References

(1) C. Lee, W. Yang and R. G. Parr, Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density, *Phys. Rev. B*, 1988, **37**, 785; A. D. Becke, Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior, *Phys. Rev. A*, 1988, **38**, 3098.

(2) *Gaussian 09, Revision D.01*; Gaussian, Inc., Wallingford CT, 2009. (accessed September 1, 2020).

(3) V. Barone and M. Cossi, Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model, *J. Phys. Chem. A*, 1998, **102**, 1995; M. Cossi, N. Rega, G. Scalmani and V. Barone, Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model, *J. Comput. Chem.*, 2003, **24**, 669.

(4) J. Zhang, I. A. Shkrob, R. S. Assary, S. o. Tung, B. Silcox, L. A. Curtiss, L. T. Thompson, L. and L. Zhang, Towards Improved Catholyte Materials for Redox Flow Batteries: What Controls Chemical Stability of Persistent Radical Cations? *J. Phys. Chem. C*, 2017, **121**, 23347.

(5) R. S. Assary, L. Zhang, J. Huang and L. A. Curtiss, Molecular Level Understanding of the Factors Affecting the Stability of Dimethoxy Benzene Catholyte Candidates from First-Principles Investigations. *J. Phys. Chem. C*, 2016, **120**, 14531.

(6) G. Landrum, *RDKit: open-source cheminformatics*. 2013. http://www.rdkit.org (accessed September 1, 2020); P. Tosco, N. Stiefl, and G. Landrum, Bringing the MMFF Force Field to the RDKit: Implementation and Validation, *J. Cheminform.*, 2014, **6**, 37.

(7) T. A. Halgren, Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94, *J. Comput. Chem.*, 1996, **17**, 490.

(8) H. Moriwaki, Y. Tian, N. Kawashita, T. Takagi, Mordred: A Molecular Descriptor Calculator, *J. Cheminform.*, 2018, **10**, 4.

(9) G. Moreau and P. Broto, Autocorrelation of a Topological Structure: A New Molecular Descriptor, *Nouv. J. Chim.*, 1980, **4**, 359.
(10) P. A. Moran, Notes on Continuous Stochastic Phenomena, *Biometrika*, 1950, **37**, 17.
(11) R. C. Geary, The Contiguity Ratio and Statistical Mapping, *Inc. Stat.*, 1954, **5**, 115.
(12) C. Buhrmester, J. Chen, L. Moshurchak, J. Jiang, R. L. Wang and J. R. Dahn, Studies of Aromatic Redox Shuttle Additives for $LiFePO_4$-Based Li-Ion Cells, *J. Electrochem. Soc.*, 2005, **152**, A2390.
(13) K. Amimoto, 1,4-Bis(1,1-dimethylpropyl)-2,5-dimethoxybenzene, *Acta Cryst.* 2011, **E67**, O2708.
(14) J. Zhang, J. Huang, L. A. Robertson, I. A. Shkrob and L. Zhang, Comparing Calendar and Cycle Life Stability of Redox Active Organic Molecules for Nonaqueous Redox Flow Batteries, *J. Power Sources*, 2018, **397**, 214.
(15) J. Zhang, J. Huang, L. A. Robertson, R. S. Assary, I. A. Shkrob, and L. Zhang, Elucidating Factors Controlling Long-Term Stability of Radical Anions for Negative Charge Storage in Nonaqueous Redox Flow Batteries, *J. Phys. Chem. C*, 2018, **122**, 8116.
(16) J. Huang, W. Duan, J. Zhang, I. A. Shkrob, R. S. Assary, B. Pan, C. Liao, Z. Zhang, X. Wei and L. Zhang, Substituted Thiadiazoles as Energy-Rich Anolytes for Nonaqueous Redox Flow Cells, *J. Mater. Chem A*, 2018, **6**, 6251; J. Zhang, Z. Yang, I. A. Shkrob, R. S. Assary, S. o. Tung, B. Silcox, W. Duan, J. Zhang, C. C. Su, B. Hu, et al., Annulated Dialkoxybenzenes as Catholyte Materials for Nonaqueous Redox Flow Batteries: Achieving High Chemical Stability through Bicyclic Substitution, *Adv. Energy Mater.*, 2017, **7**, 1701272; W. Duan, J. Huang, J. A. Kowalski, I. A. Shkrob, M. Vijayakumar, E. Walter, B. Pan, Z. Yang, J. D. Milshtein, B. Li, et al., "Wine-Dark Sea" in an Organic Flow Battery: Storing Negative Charge in 2,1,3-Benzothiadiazole Radicals Leads to Improved Cyclability, *ACS Energy Lett.*, 2017, **2**, 1156; W. Weng, J. Huang, I. A. Shkrob, L. Zhang and Z. Zhang, Redox Shuttles with Axisymmetric Scaffold for Overcharge Protection of Lithium-Ion Batteries, *Adv. Energy Mater.*, 2016, **6**, 1600795; J. Huang, B. Pan, W. Duan, X. Wei, R. S. Assary, L. Cheng, F. R. Brushett, M. S. Ferrandon, C. Liao, Z. Zhang, et al., The Lightest Organic Radical Cation for Charge Storage in Redox Flow Batteries, *Sci. Rep.*, 2016, **6**, 32102; S. V. Rosokha and J. K. Kochi, Continuum of Outer- and Inner-Sphere Mechanisms for Organic Electron Transfer. Steric Modulation of the Precursor Complex in Paramagnetic (Ion-Radical) Self-Exchanges, *J. Am. Chem. Soc.* 2007, **129**, 12, 3683.
(17) E. M. Cramer, Some Symmetric, Invariant Measures of Multivariate Association, *Psychometrica*, 1979, **44**, 43.
(18) R. Tibshirani, Regression Shrinkage and Selection via the Lasso, *J. R. Statist. Soc. B*, 1996, **58**, 267.
(19) H. Zou and T. Hastie, Regularization and Variable Selection via the Elastic Net, *J. R. Statist. Soc. B*, 2005, **67, Part 2**, 301.
(20) D. Broadhurst, R. Goodacre, A. Jones, J. J. Rowland and D. B. Kelp, Genetic Algorithms as a Method for Variable Selection in Multiple Linear Regression and Partial Least Squares Regression, with Applications to Pyrolysis Mass Spectrometry, *Anal. Chim. Acta*, 1997, **348**, 71; A. Yasri and D. Hartsough, Toward an Optimal Procedure for Variable Selection and QSAR Model Building, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1218; J. K. Wegner and A. Zell, Prediction of Aqueous Solubility and Partition Coefficient Optimized by a Genetic Algorithm Based Descriptor Selection Method, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1077.