Supplementary Information

Data-Driven Discovery of Carbonyl Organic Electrode Molecules: Machine Leaning and Experiment

Jiayi Du,[‡a,b] Jun Guo,[‡a,c] Qiqi Sun,[a,d] Wei Liu,[a] Tong Liu,*[a,b] Gang Huang[a,b] and Xinbo Zhang*[a,b]

[a] State Key Laboratory of Rare Earth Resource Utilization, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, PR China.

[b] School of Applied Chemistry and Engineering, University of Science and Technology of China, Hefei 230026, PR China.

[c] School of Materials Science and Engineering, Changchun University of Science and Technology, Changchun 130022, PR China.

[d] National & Local United Engineering Laboratory for Power Battery, Department of Chemistry, Northeast Normal University, Changchun, 130024, PR China

‡ These authors contributed equally to this work.

* Corresponding authors and E-mail: tongliu@ciac.ac.cn (TL), xbzhang@ciac.ac.cn (XZ)

Supplementary Note 1: Extracting Database

PubChem[1] is a freely accessible chemical database including 116 million organic molecules. PubChemPy allows for the download of molecule information from PubChem using the CID number, which serves as the identification key for each molecule within the database. In this study, molecules with CID values ranging from 1 to 1,000,000 were retrieved and presented in the form of simplified molecular input line entry specification (SMILES).

Supplementary Note 2: The calculating method of theoretical capacity

In the selected carbonyl molecules, one carbonyl bond is regarded as an active site. The theoretical capacity (TC, mAh g^{-1}) of OEMs is calculated as[2] :

$$TC = \frac{nF}{3.6M_w} \qquad \text{Equation S1}$$

where n is transferred-electron number of one organic molecule, F is the Faraday constant (C mol^{-1}), and M_w is the molar mass (g mol^{-1}) of investigated organic molecule.

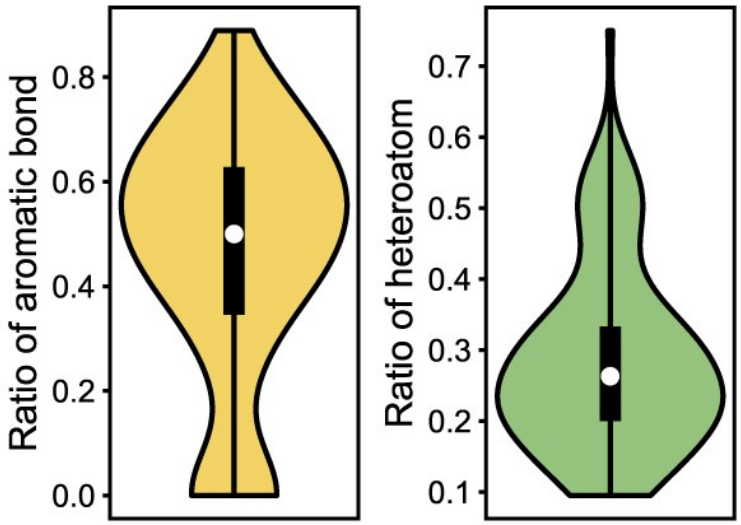

Figure S1. The distribution of aromatic bond ratio and heteroatom ratio in CMDS.

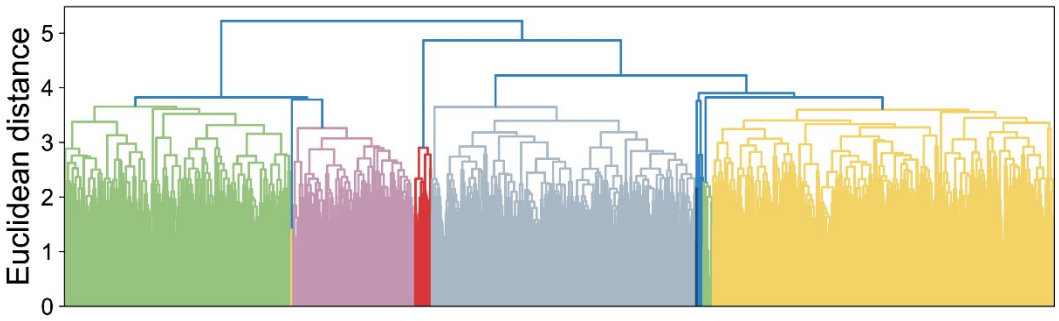

Figure S2. The complete dendrogram of hierarchical clustering for CMDB.

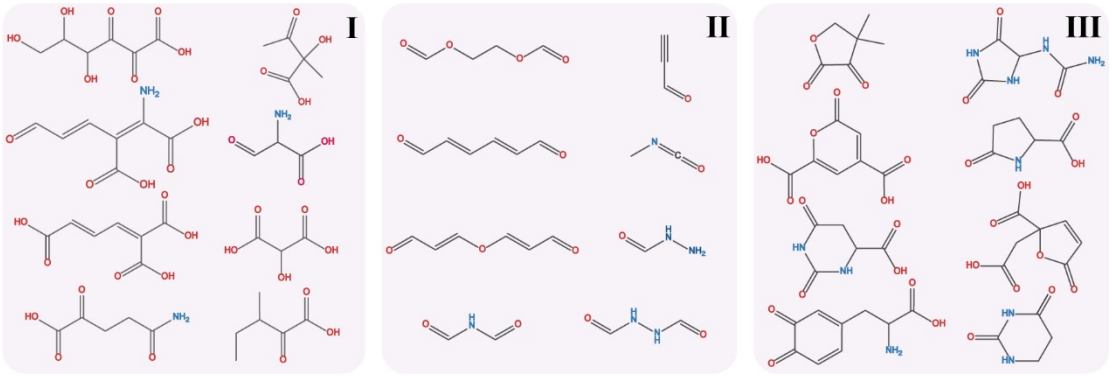

Figure S3. Partial molecules in each of three clusters.

Supplementary Note 3: High-throughput calculations

In our work, R, defined as the organic electrode molecule, was calculated in four different states, for example, R in gas phase (R(gas)), R in solvation phase (R(sol)), R with one negative charge in gas phase (R^-(gas)) and R with one negative charge in solvation phase (R^-(sol)). Figure S4 shows the thermodynamic cycle[3]. The reduction free energy of R(sol) was calculated by Equation S2. The reduction potential (E^{red}) was decided by Equation S3, where *F* represents the Faraday constant and *n* denotes the number of electrons obtained by R(sol) during the reduction process. A constant 1.4 V is derived from the absolute potential of hydrogen electrode (-4.44 V)[4] and the potential of Li/Li+, -3.04 V vs. SHE[5].

The calculations of reduction potential for 1524 molecules were performed by Gaussian 16 software[6]. To strike a balance between computational accuracy and time cost, each molecule/anion was optimized at the wb97xd/6-31+(d, p) level of theory[7]. Meanwhile, the solvation effect was taken into consideration using SMD solvation model.

$$\Delta G^{red}(R, sol) = \Delta G^{red}(R, gas) + \Delta G^{sol}(R^-) - \Delta G^{sol}(R) \qquad \text{Equation S2}$$

$$E^{red} = -\frac{\Delta G^{red}(R,sol)}{nF} - 1.4\ V \qquad \text{Equation S3}$$

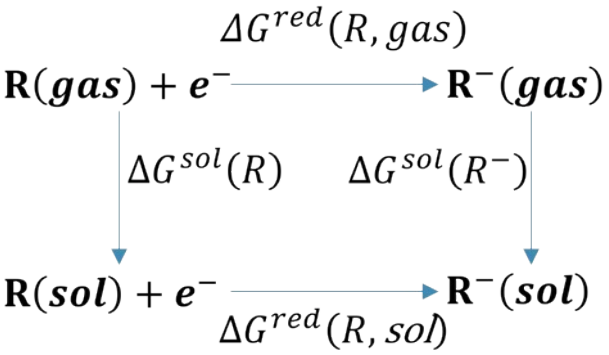

Figure S4. Thermodynamic cycle utilized to calculate the equilibrium reduction potential[3].

Table S1. The detailed description of SubFPC2, SubFPC3, SubFPC137, SubFPC143, SubFPC274, and SubFPC287[8].

SubFingerprintCount	SMARTS Code	Description
SubFPC2	[CX4H2]([#6])[#6]	Secondary carbon
SubFPC3	[CX4H1]([#6])([#6])[#6]	Tertiary carbon
SubFPC137	[#6X3](=[OX1])[#6X3]=,:[#6X3][#6;!$(C=[O,N,S])]	Vinylogous ester
SubFPC143	[!#6][#6X3](=[!#6])[!#6]	Carbonic acid derivatives
SubFPC274	a	Aromatic
SubFPC287	*=*[*]=,#,:[*]	Conjugated double bond

Table S2. The selected descriptors.

Descriptors	Bit	Package
SubFPC	22	Padel
SubFP	22	Padel
ETA	43	Padel
Topological charge	21	Padel
BCUT	6	Padel
MolMR	1	RDKit
TPSA	1	RDKit
FractionCSP3	1	RDKit
HallKierAlpha (HKA)	1	RDKit
Number of valence electrons (NVE)	1	RDKit
Max partial charge (MaxPC)	1	RDKit
Min partial charge (MinPC)	1	RDKit

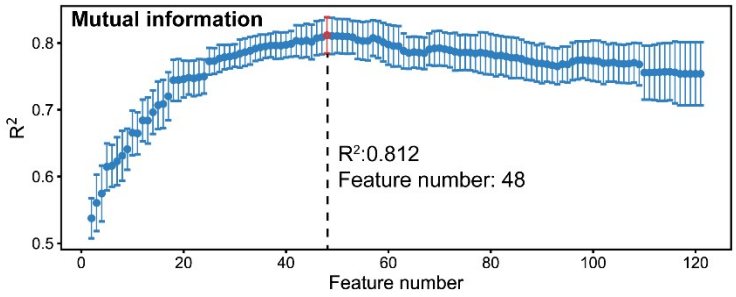

Figure S5. The feature selection based on the mutual information.

Table S3. The finally descriptors after the selection of the mutual information.

Descriptor	Sub-descriptor	MIV
SubFPC	SubFPC135	0.075783291
	SubFPC137	0.277716782
	SubFPC274	0.100764289
	SubFPC275	0.12300611
	SubFPC287	0.485312071
	SubFPC295	0.071011697
	SubFPC300	0.128301604
	SubFPC301	0.129026096
	SubFPC303	0.233345082
	SubFPC307	0.107008915
SubFP	SubFP137	0.191818529
	SubFP274	0.081706438
	SubFP287	0.353126043
	SubFP303	0.086648261
ETA	ETA_AlphaP	0.110638753
	ETA_dAlpha_B	0.109555779
	ETA_Epsilon_1	0.098342864
	ETA_Epsilon_2	0.143255146
	ETA_Epsilon_4	0.076268085
	ETA_dEpsilon_A	0.134939987
	ETA_dEpsilon_B	0.267286989
	ETA_dEpsilon_C	0.091440932
	ETA_Psi_1	0.140622332
	ETA_dPsi_A	0.138031357
	ETA_Shape_P	0.057991984
	ETA_Beta	0.173095892
	ETA_BetaP	0.230217858
	ETA_Beta_ns	0.215775755
	ETA_BetaP_ns	0.39029535
	ETA_dBeta	0.256327942
	ETA_dBetaP	0.401276238
	ETA_Eta_F	0.082658742
	ETA_EtaP_F	0.152589706
	ETA_EtaP_L	0.170670794
	ETA_Eta_F_L	0.170169227
	ETA_EtaP_F_L	0.177467244
Topology Charge	JGI1	0.070457046
	JGT	0.07115601
BCUT	BCUTw-1l	0.105666535
	BCUTc-1l	0.259742361

	BCUTc-1h	0.158678313
MolMR	mr	0.077901658
TPSA	tpsa	0.068758274
FractionCSP3	fracsp3	0.230833695
HallKierAlpha	hall_kier_alpha (HKA)	0.231417502
Number of valence electrons	number_valence_electrons (NVE)	0.109861894
Max partial charge	max_partial_charge (MaxPC)	0.265835814
Min partial charge	min_partial_charge (MinPC)	0.223576882

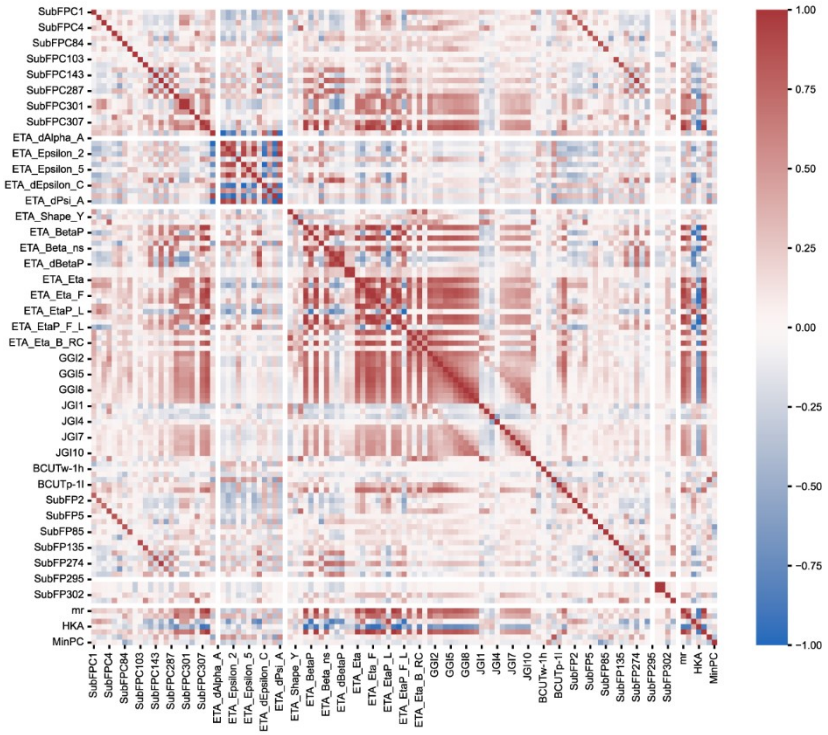

Figure S6. The heat map of Pearson correlation coefficient matrix among the 121 descriptors for OEMs.

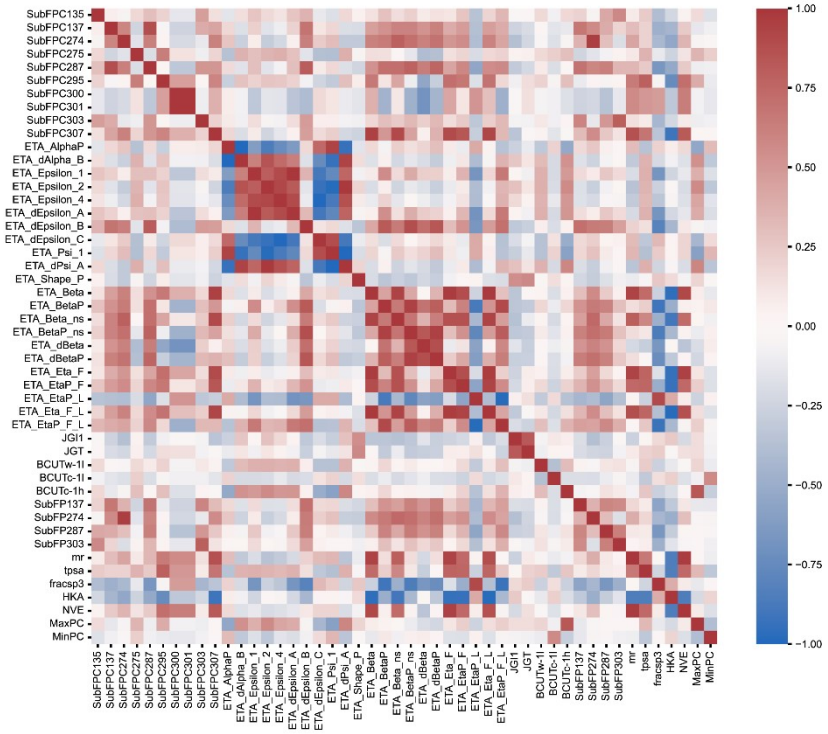

Figure S7. The heat map of Pearson correlation coefficient matrix among the 48 descriptors for OEMs.

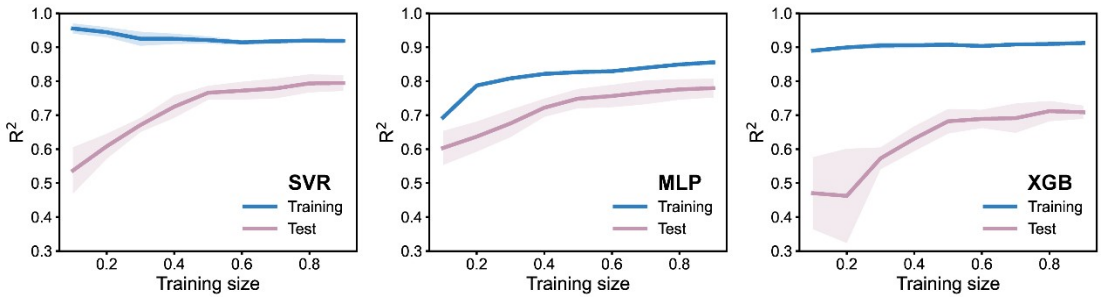

Figure S8. The learning curves of SVR, MLP, and XGB models.

Table S4. The mean absolute error (MAE) and root mean square error (RMSE).

Model	Task	MAE	RMSE
SVM	Training	0.131	0.277
	Test	0.243	0.334
MLP	Training	0.244	0.354
	Test	0.277	0.383
XGB	Training	0.197	0.272
	Test	0.307	0.415

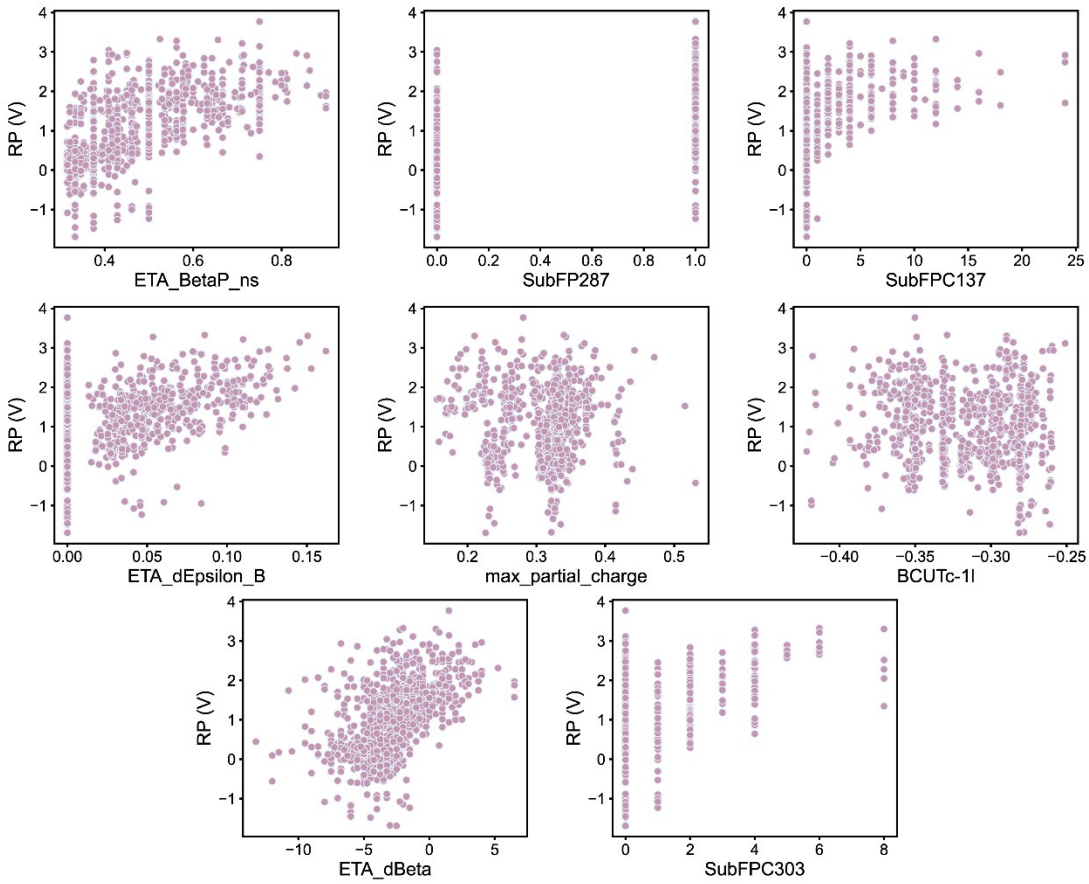

Figure S9. The relationship of RP and other 8 descriptors. The 8 descriptors are ETA_BetaP_ns, SubFPC287, SubFPC137, ETA_dEpsilon_B, max_partial_charge, BCUT-1l, ETA_dBeta, and SubFPC303.

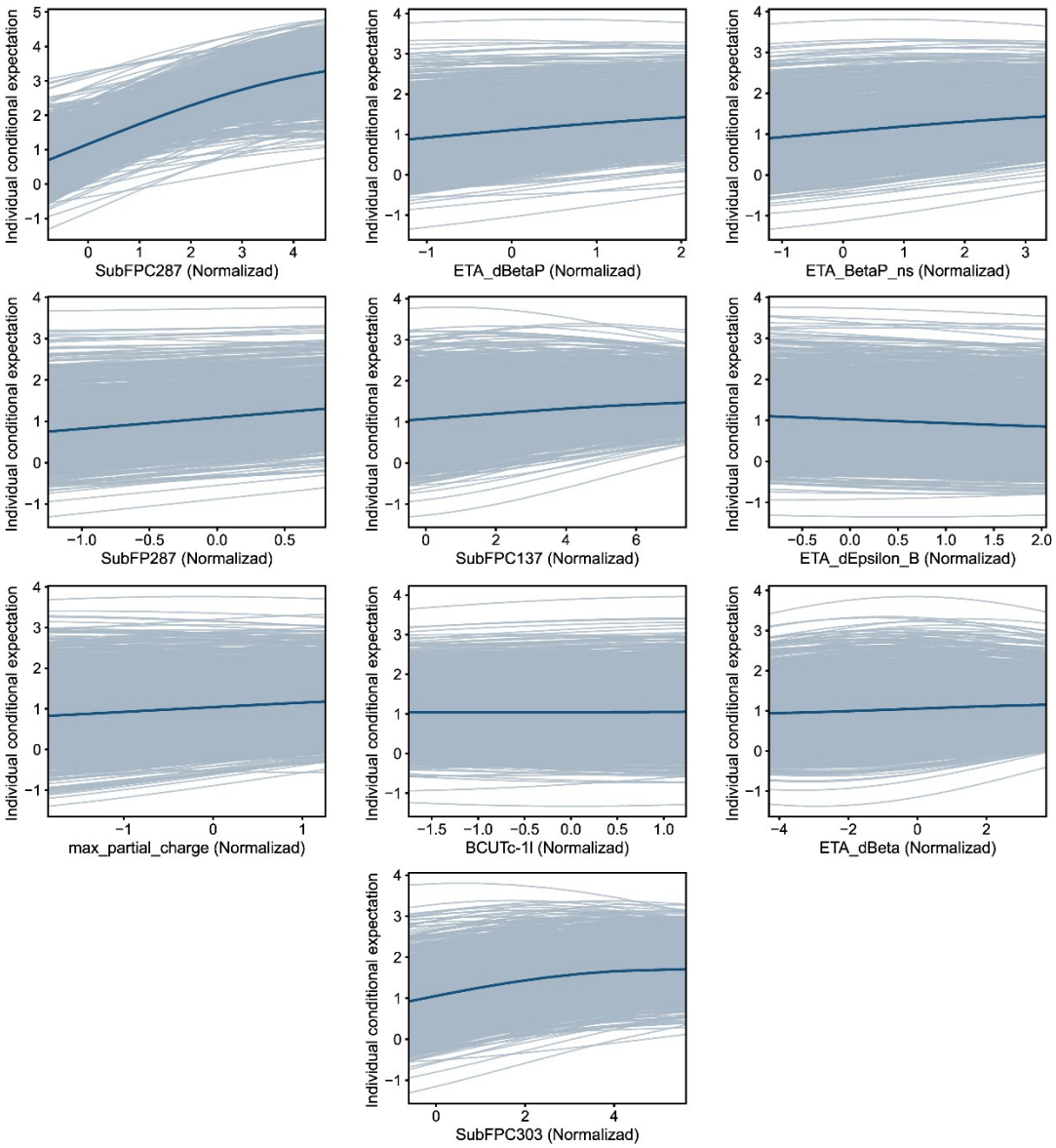

Figure S10. The individual conditional expectation between RP and the top 10 MIV of descriptors.

Supplementary Note 5: Experiments

Synthesis of 5,8-dihydroxynaphthalene-1,4-dione (DND)

The anhydrous $AlCl_3$ (1 mmol) and NaCl (0.5 mmol) were added into a Schlenk flask at room temperature and then the mixture was heated till molten. After this, the furan-2,5-dione (0.1 mmol) and hydroquinone (0.1 mmol) were added into the mixture, and the temperature was increased to 180 °C and maintained for 300 min. When cooling down to 0 °C via the ice-water bath, 10% HCl was added, and the mixture was stirred for 30 min and followed by refluxing for 60 min. Finally, the product, DND, was extracted by ethyl acetate and purified by silica gel column chromatography.[9]

Characterizations

NMR were performed on a Bruker AVANCE III 400 WB spectrometer. XPS measurement was carried out with a VG Scientific ESCALAB MKII X-ray photoelectron spectrometer. FTIR measurement were carried out on a Bruker IFS 66V/S spectrometer.

Electrochemical measurements

The electrode was prepared with DND[10], KB conductive additive, and polytetrafluoroethylene (PTFE) at a weight ratio of 5: 4: 1 by anhydrous ethanol. The mixed slurry was fully ground by ball milling for 20 min, and then rolled into film and pressed on aluminum mesh current collector to obtain the DND cathodes. After this, the DND electrodes were dried overnight in a vacuum oven at 50 °C. The mass loading of DND was controlled at around 0.8 mg cm^{-2}. The electrochemical performance of the PNT//Li metal batteries were tested in 2025-type coin cells, and the electrolyte was 1

M $LiPF_6$ in EC: DMC=1:1 Wt%. Galvanostatic charge–discharge cycling test was performed with a Land CT2001A battery testing system in the voltage range of 1.5-4.2 V (vs. Li/Li+) at room temperature. Electrochemical impedance spectroscopy (EIS) test was carried out by a Biologic VMP3 electrochemical station with coin cells at room temperature.

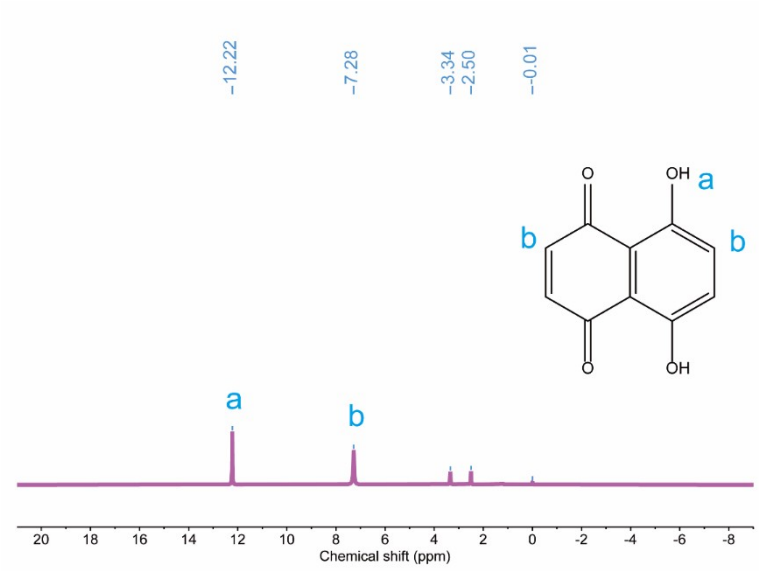

Figure S11. ^{1}H NMR (DMSO-d6) spectrum of DND.

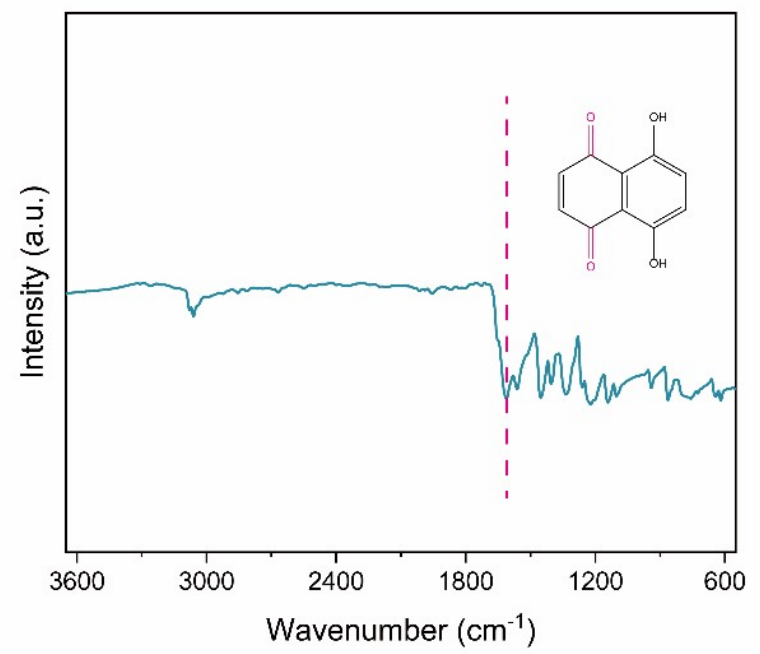

Figure S12. The FITR spectra of DND.

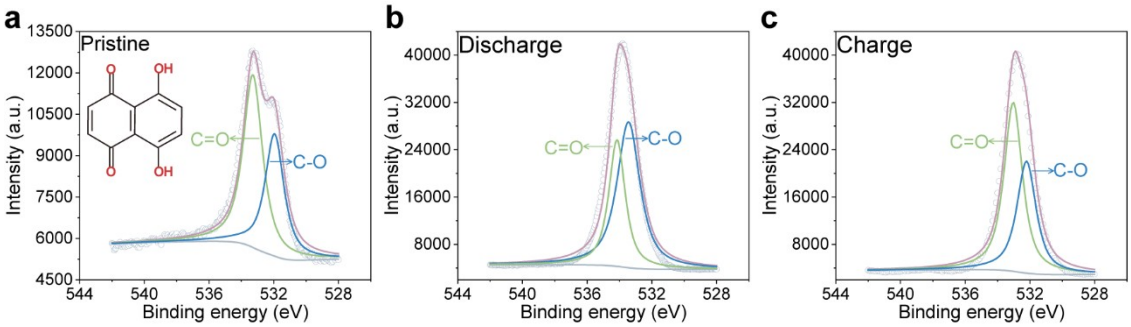

Figure S13. The O 1s XPS spectra of NT cathode. a) The XPS spectra of pristine of NT cathode. b) The XPS spectra of NT cathode after discharging. c) The XPS spectra of NT cathode after charging.

Table S5. The performance comparison of NT with the reported organic carbonyl cathode materials.

Name	Energy density (Wh/kg)	Capacity (mAh/g)	Voltage (V)	Cycle number
C6O6[11]	1533.4	902	1.7	100
DHCQ[12]	460.5	921	0.5	400
Li2TP[13]	346.4	433	0.8	1000
C4Q[14]	1158.05	437	2.65	100
C6Q[15]	1012	440	2.3	300
BBQB[16]	917.5	367	2.5	100
Li-lawsone[17]	644	280	2.3	1000
TPB[18]	471.5	230	2.05	100
Li2BZL[19]	150	200	0.75	70
IDT[20]	565	226	2.5	1000
NTCDI[21]	342.7	149	2.3	23
DAAP[22]	570	285	2	400
2PTCDA[23]	300	125	2.4	200
DBTOP[24]	351.396	90.8	3.87	100
LiPHB[25]	305.3	86	3.55	500
H2PDI[26]	207	92	2.25	1000
PDI[27]	143	55	2.6	300
This work	**1277.5**	**511**	**2.5**	**2500**

Reference

1. S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *Nucleic Acids Res.*, 2021, **49**, D1388-D1395.
2. Y. Lu, Q. Zhang, L. Li, Z. Niu and J. Chen, *Chem*, 2018, **4**, 2786-2813.
3. Q. Zhang, A. Khetan, E. Sorkun, F. Niu, A. Loss, I. Pucher and S. Er, *Energy Stor. Mater.*, 2022, **47**, 167-177.
4. S. Trasatti, *J. Electroanal. Chem. Interfacial Electrochem.*, 1986, **58**, 955-966.
5. S. J. Yang, X. Y. Qin, R. He, W. Shen, M. Li and L. B. Zhao, *Phys. Chem. Chem. Phys.*, 2017, **19**, 12480-12489.
6. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, ***Gaussian 16 Rev. C.01***, 2016.
7. G. Agarwal, H. A. Doan, L. A. Robertson, L. Zhang and R. S. Assary, *Chem. Mater.*, 2021, **33**, 8133-8144.
8. C. W. Yap, *J. Comput. Chem.*, 2011, **32**, 1466-1474.
9. J. Zhang, Y. Liu, D. Shi, G. Hu, B. Zhang, X. Li, R. Liu, X. Han, X. Yao and J. Fang, *Eur. J. Med. Chem.*, 2017, **140**, 435-447.
10. V. P. Papageorgiou, A. N. Assimopoulou, E. A. Couladouros, D. Hepworth and K. C. Nicolaou, *Angew. Chem., Int. Ed.*, 1999, **38**, 270-301.
11. Y. Lu, X. Hou, L. Miao, L. Li, R. Shi, L. Liu and J. Chen, *Angew. Chem., Int. Ed.*, 2019, **58**, 7020-7024.
12. L. Chen, S. Liu, L. Zhao and Y. Zhao, *Electrochim. Acta*, 2017, **258**, 677-683.
13. S. Zhang, S. Ren, D. Han, M. Xiao, S. Wang and Y. Meng, *J. Power Sources*, 2019, **438**.
14. W. Huang, M. Zhang, H. Cui, B. Yan, Y. Liu and Q. Zhang, *Chem. - Asian J.*, 2019, **14**, 4164-4168.
15. W. Huang, X. Zhang, S. Zheng, W. Zhou, J. Xie, Z. Yang and Q. Zhang, *Sci. China Mater.*, 2019, **63**, 339-346.
16. J. Yang, P. Xiong, Y. Shi, P. Sun, Z. Wang, Z. Chen and Y. Xu, *Adv. Funct. Mater.*, 2020, **30**, 109597.
17. J. Lee and M. J. Park, *Adv. Energy Mater.*, 2017, **7**, 1602279.
18. Z. Luo, L. Liu, Q. Zhao, F. Li and J. Chen, *Angew. Chem., Int. Ed.*, 2017, **56**, 12561-12565.

19. V. Medabalmi and K. Ramanujam, *J. Electrochem. Soc.*, 2017, **164**, A1720-A1725.
20. H. Zhang, R. Zhang, F. Ding, C. Shi and N. Zhao, *Energy Stor. Mater.*, 2022, **51**, 172-180.
21. S. More, N. Khupse, M. Bhosale, J. Ambekar, M. Kulkarni and B. Kale, *ChemistrySelect*, 2020, **5**, 2157-2163.
22. H. Peng, P. Chen, X. Yang, Z. Xue, S. Wang, J. Na, J. Yu and Y. Yamauchi, *J. Mater. Chem. A*, 2020, **8**, 11521-11528.
23. Y. Hong, J. Hu, W. Tang, B. Wei, M. Guo, S. Jia and C. Fan, *Energy Stor. Mater.*, 2022, **52**, 61-68.
24. Y. Zheng, H. Ji, J. Liu, Z. Wang, J. Zhou, T. Qian and C. Yan, *Nano Lett.*, 2022, **22**, 3473-3479.
25. M. Rajesh, F. Dolhem, C. Davoisne and M. Becuwe, *ChemSusChem*, 2020, **13**, 2364-2370.
26. V. Medabalmi and K. Ramanujam, *ChemistrySelect*, 2018, **3**, 10657-10662.
27. J. Aher, A. Graefenstein, G. Deshmukh, K. Subramani, B. Krueger, M. Haensch, J. Schwenzel, K. Krishnamoorthy and G. Wittstock, *ChemElectroChem*, 2020, **7**, 1160-1165.