Identifying the Impact of Chemical Functional Groups on Ionic Liquid Conductivity

By: J. E. Umaña, N. A. Zawicki, Matthew A. Gebbie, Victor M. Zavala, and Rose K. Cersonsky*

Department of Chemical and Biological Engineering, University of Wisconsin—Madison, Madison, WI 53706, United States of America

*Corresponding Author

Regression Results

Table S1. Train set coefficients of determination (R^2) and root mean squared errors (RMSE) for linear ridge regression, RBF kernel regression, random forest regression, and multi-layer perceptron neural network regression models employed for conductivity at 298 K using the SMARTS[1, 2] representation we developed (SMARTS), MACCS Keys[3] (MACCS), and a 36-dimension MACCS principal component representation (PCA MACCS).

Model	SMARTS R^2	SMARTS RMSE	MACCS R^2	MACCS RMSE	PCA MACCS R^2	PCA MACCS RMSE
Ridge Regression	0.711	0.430	0.664	0.464	0.579	0.519
RBF Kernel Regression	0.97	0.139	0.981	0.112	0.950	0.179
Random Forest Regression	0.935	0.203	0.934	0.206	0.912	0.237
MLP Neural Network	0.980	0.113	0.982	0.107	0.988	0.088

Table S2. Train set coefficients of determination (R^2) and root mean squared errors (RMSE) for linear ridge regression, RBF kernel regression, random forest regression, and multi-layer perceptron neural network regression models employed for conductivity for all temperatures using the SMARTS[1, 2] representation we developed (SMARTS), MACCS Keys[3] (MACCS), and a 36-dimension MACCS principal component representation (PCA MACCS).

Model	SMARTS R^2	SMARTS RMSE	MACCS R^2	MACCS RMSE	PCA MACCS R^2	PCA MACCS RMSE
Ridge Regression	0.721	1.019	0.768	0.930	0.739	0.986
RBF Kernel Regression	0.979	0.278	0.995	0.143	0.988	0.209
Random Forest Regression	0.966	0.356	0.992	0.174	0.994	0.154
MLP Neural Network	0.973	0.318	0.978	0.288	0.985	0.235

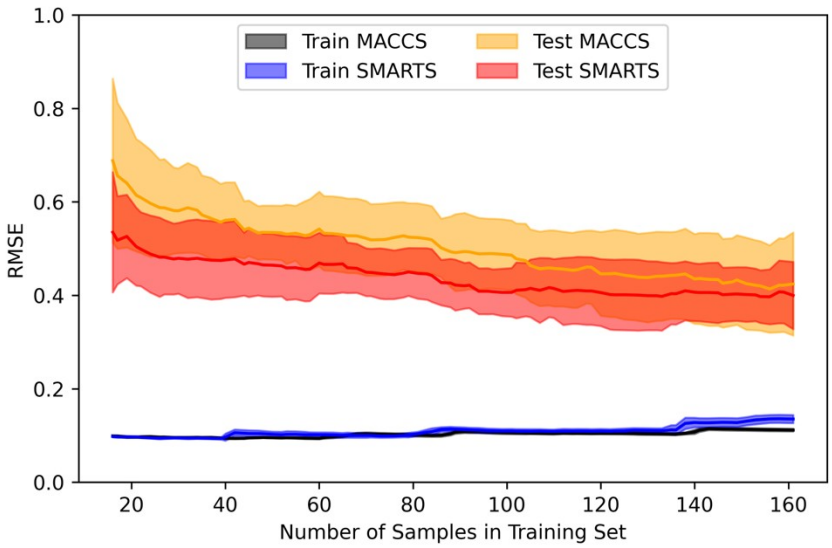

Figure S1. Kernel SVM learning curves using SMARTS and MACCS Keys representation for conductivity at 298 K. Plots shown are means and standard deviations of training and test sets for 10-fold cross validations.

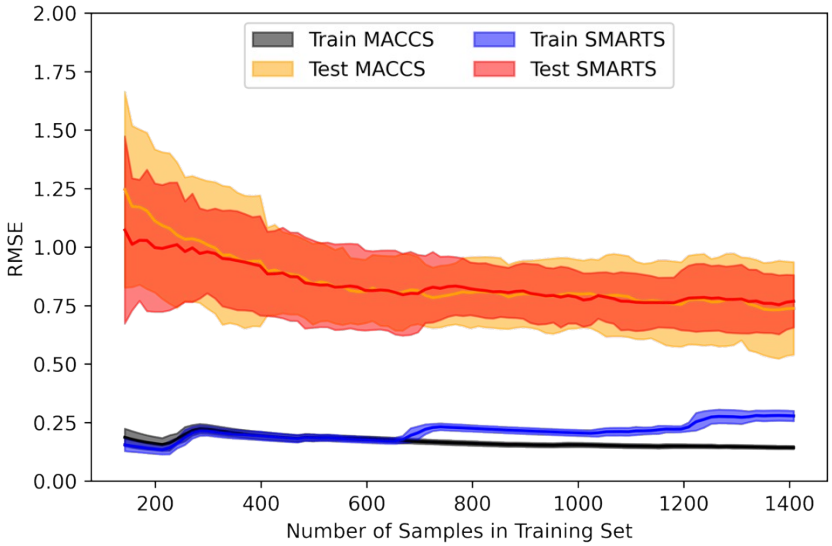

Figure S2 Kernel SVM learning curves using SMARTS and MACCS Keys representation for temperature-dependent molar conductivity. Plots shown are means and standard deviations of training and test sets for 10-fold cross validations.

Principal Covariates Regression Results

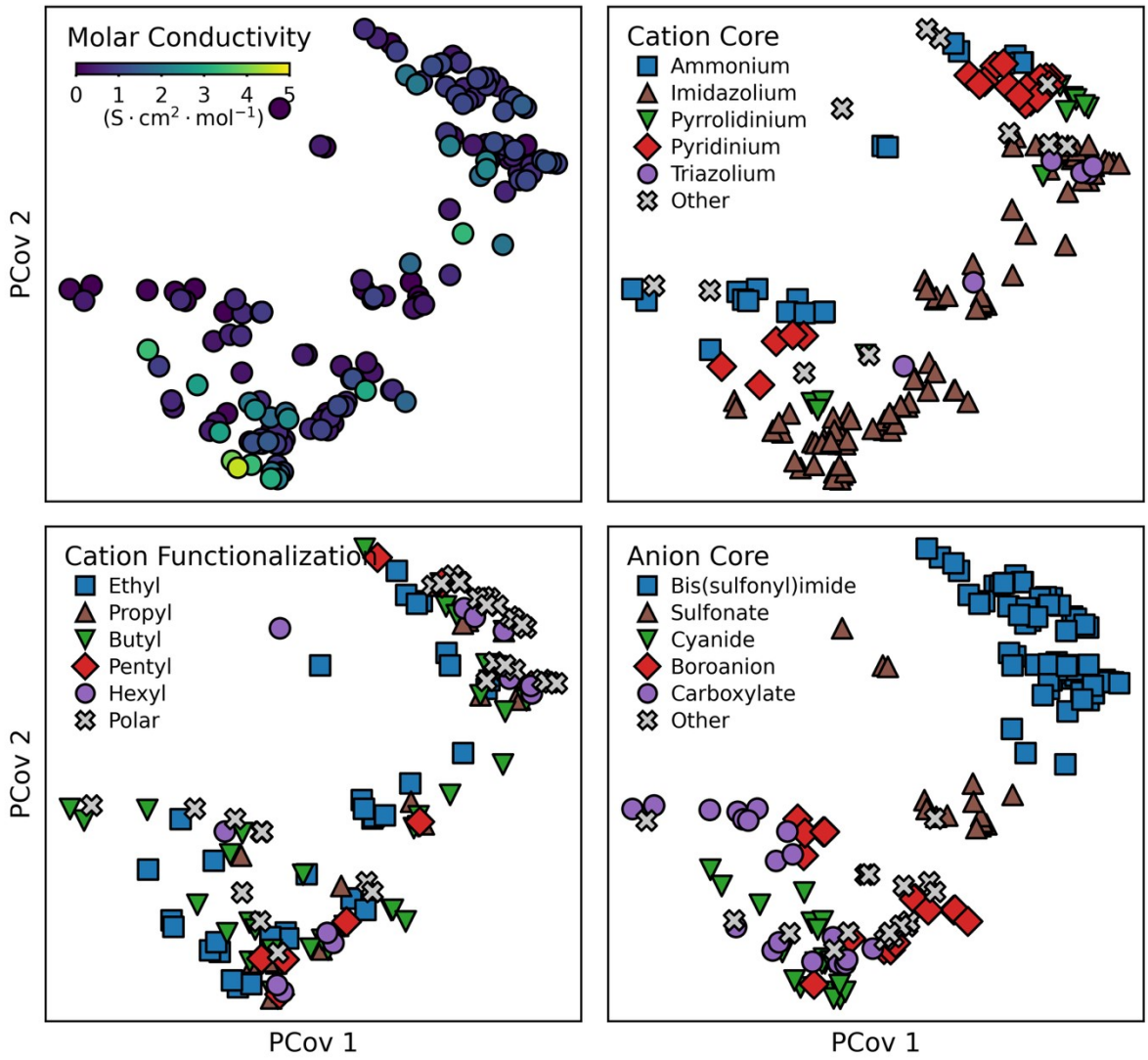

Figure S3: *2-dimensional principal covariate data projections of 182 ionic liquids represented by MACCS Keys overlayed with molar conductivity at 298 K (Top Left), cation core identity (Top Right), cation functionalization (Bottom Left), and anion core identity (Bottom Right).*

Table S3. SMARTS fragments weights for PCovR[4] 1 and 2. [c]Cationic Structure. [a]Anionic Structure.

Name:	PCovR 1:	PCovR 2:
Ammonium 0H[c]	0.0562	-0.0208
Ammonium 3H[c]	0.2213	0.4437
Pyrrolidinium[c]	0.1365	0.2618
Imidazolium[c]	0.2653	0.3955
Pyridinium[c]	0.1372	0.2383
Triazolium[c]	-0.0915	-0.2551
Ether[c]	-0.0451	-0.2656
Nitrile[c]	-0.2235	-0.5156
Propyl[c]	0.1346	-0.3857
Butyl[c]	0.0097	-0.4972
Pentyl[c]	0.0252	-0.1880
Hexyl[c]	-0.0204	-0.2428
Hydroxyl[c]	-0.0828	-0.2016
Secondary Amine	-0.1790	-0.4430
3-Unit Chain[c]	0.1889	-0.2671
4-Unit Chain[c]	0.2639	-0.0858
Nitrile[a]	0.3075	0.7243
Methyl[a]	-0.1735	-0.4162
Ethyl[a]	-0.0437	-0.1134
Propyl[a]	-0.1088	-0.2545
Fluoromethyl[a]	0.1416	0.2402
e- Fluoromethyl[a]	0.0290	-0.3287
Fluoroethyl[a]	-0.0727	-0.2444
Fluoropropyl[a]	-0.0714	-0.2042
Fluorobutyl[a]	-0.0691	-0.1799
Hydroxyl[a]	-0.0159	-0.0806
2-Unit Chain[a]	0.3358	0.1883
3-Unit Chain[a]	0.3358	0.1883
4-Unit Chain[a]	0.3738	0.3864
Bis(sulfonyl)imide[a]	0.2698	0.2904
Sulfonate[a]	0.0495	0.0623
Cyanide[a]	0.0345	0.1053
Carboxylate[a]	0.0930	0.1267
Phosphate[a]	0.1253	0.2376
Borate[a]	0.1580	0.2657
Carboxyl[a]	0.0319	0.0569

Methodology

Table S4. SMARTS definitions used for substructure matching ionic liquid molecular structures. Structures which were excluded due to overrepresentation (present in more than 179 of 182 ionic liquid structures) or underrepresented (present in less than 4 of 182 ionic liquid structures) are with an asterisk. *Structure was not included, **Structure was included for cations but not for anions, ***Structure was included for anions but not for cations. [c]Cationic Structure. [a]Anionic Structure.

Name:	SMARTS:	Other representations:
Ammonium 0H[c]	[#7+;D4;H0;!R]	
Ammonium 1H[c,]*	[#7+;D3;H1;!R]	
Ammonium 2H[c,]*	[#7+;D2;H2;!R]	
Ammonium 3H[c]	[#7+;D1;H3;!R]	
Pyrrolidinium[c]	[N+;R]1(CCCC1)	
Imidazolium[c]	[n;H0]1cc[n+;H0]c1	[n+;H0]1cc[n;H0]c1
Trialkylimidazolium[c,]*	n1**[n+]c1~c	[n+]1**nc1~c
Piperidinium[c,]*	[N+]1(CCCCC1)	
Pyridinium[c]	c1cc[n+](cc1)	
Tetrahydrothiophenium[c,]*	C1CC[S+]C1	
Phosphorus Cation[c,]*	[P+;D4]	
Iminium[c,]*	NC(=[N+])	
Triazolium[c]	n1~c~[n+]~c~n~1	[n+]1~c~n~c~n~1
Protic Imidazolium[c,]*	[n;H1]1**[n+]c1	N1**[n+;H1]c1
Neutral Aromatic Ring: 5-membered*	[*;+0]1[*;+0][*;+0][*;+0][*;+0]1	
Neutral Aromatic Ring: 6-membered*	[*;+0]1[*;+0][*;+0][*;+0][*;+0][*;+0]1	
Ether**	C~O~C	
Nitrile	*~[C+0]#[N+0]	
Carboxyl***	[O+0]~C(=O)	
Methyl***	*~[CH3;!R]	
Ethyl	*~[CH2;!R]~*	
Propyl	*~[CH2;!R]~[CH2;!R]~*	
Butyl**	*~[CH2;!R]~[CH2;!R]~[CH2;!R]~*	
Pentyl**	*~[CH2;!R]~[CH2;!R]~[CH2;!R]~[CH2;!R]~*	
Hexyl**	*~[CH2;!R]~[CH2;!R]~[CH2;!R]~[CH2;!R]~[CH2;!R]~*	
Fluoromethyl***	C(~F)(~F)(~F)~C(~[*+0])(~[*+0])	
e- Fluoromethyl[a]	C(~F)(~F)(~F)~C(~[*-])(~[*+0])	C(~F)(~F)(~F)~S(~[*-

		])(~[*+0])(~[*+0])
Fluoroethyl***	*~C(~F)(~F)~C(~F)(~F)(~F)	
Fluoropropyl***	*~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)(~F)	
Fluorobutyl***	*~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)(~F)	
Fluoropentyl*	*~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)(~F)	
Fluorohexyl*	*~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)~C(~F)(~F)(~F)	
Hydroxyl	*~[OH1]	
Primary Amine*	[#7+0;H2;R0;D1]	
Secondary Amine**	[#7+0;H1;R0;D2]	
Tertiary Amine*	[#7+0;H0;R0;D3]	
2-Unit Chain***	[A;!R]~[A;!R]	
3-Unit Chain	[A;!R]~[A;!R]~[A;!R]	
4-Unit Chain	[A;!R]~[A;!R]~[A;!R]~[A;!R]	
Bis(sulfonyl)imide[a]	*~S(=O)(=O)[N-]S(=O)(=O)~*	*~S(=O)([O-])=NS(=O)(=O)~*
Sulfonate[a]	*~[S-](~O)(~O)~O	*~S(~[O-])(~O)~O
Cyanide[a]	[*-;!X4]~C#N	[*;!X4]=C=[N-]
Carboxylate[a]	*~C(=O)[O-]	
Azide[a,*]	[N-]=[N+]=[N-]	[N--]-[N+]#N
Nitrate[a,*]	[N+](~O)(~[O-])(~[O-])	
Perrhenate[a,*]	[Re](~[O-])(~O)(~O)(~O)	[Re-](~O)(~O)(~O)(~O)
Phosphate[a]	[P-;D6]	[P;D4]~[O-]
Borate[a]	[B-;D4]	
Diazole Anion[a,*]	c1~c~n~c~[n-]1	c1~c~[n-]~c~n~1

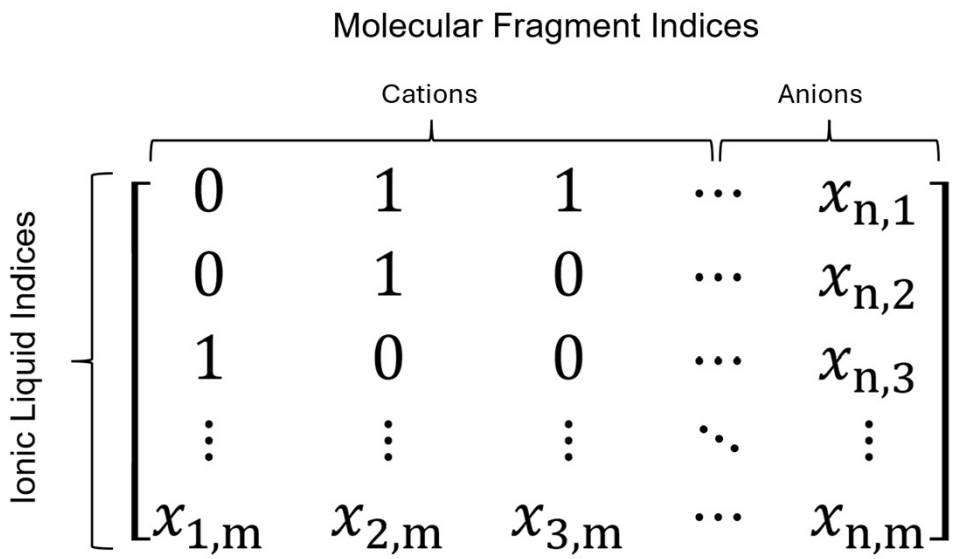

Figure S4. Visualization of an $n \times m$ isothermal bit matrix created for machine learning analysis, where n represents the number of molecular fragment strings and m denotes the number of ionic liquids in the dataset. Each x_{ij} element represents a binary state.

Parity Plots

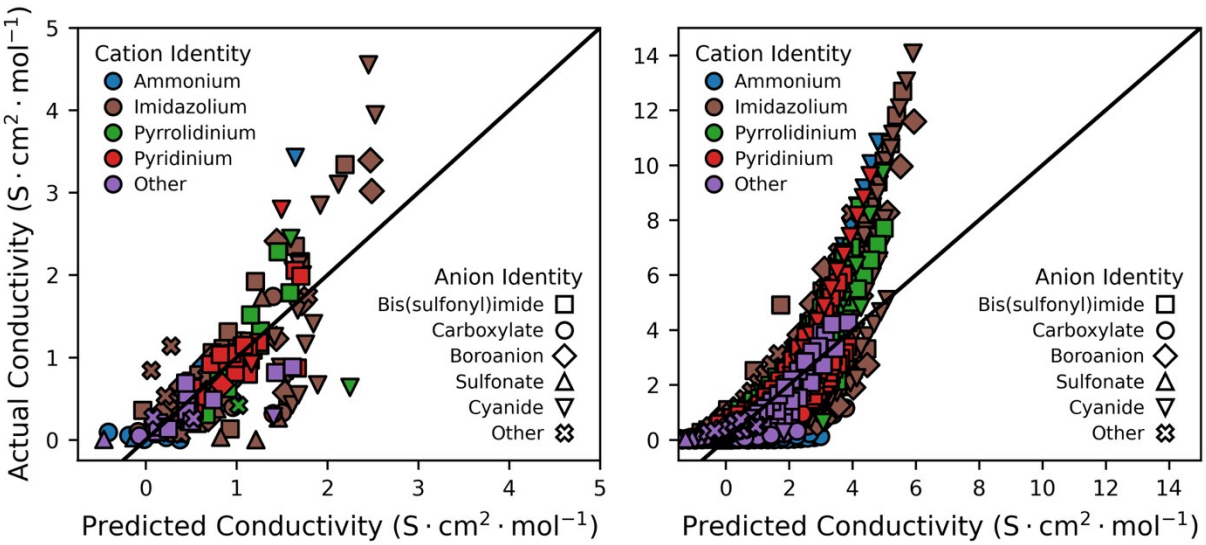

Figure S5. *Linear ridge regression test set parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

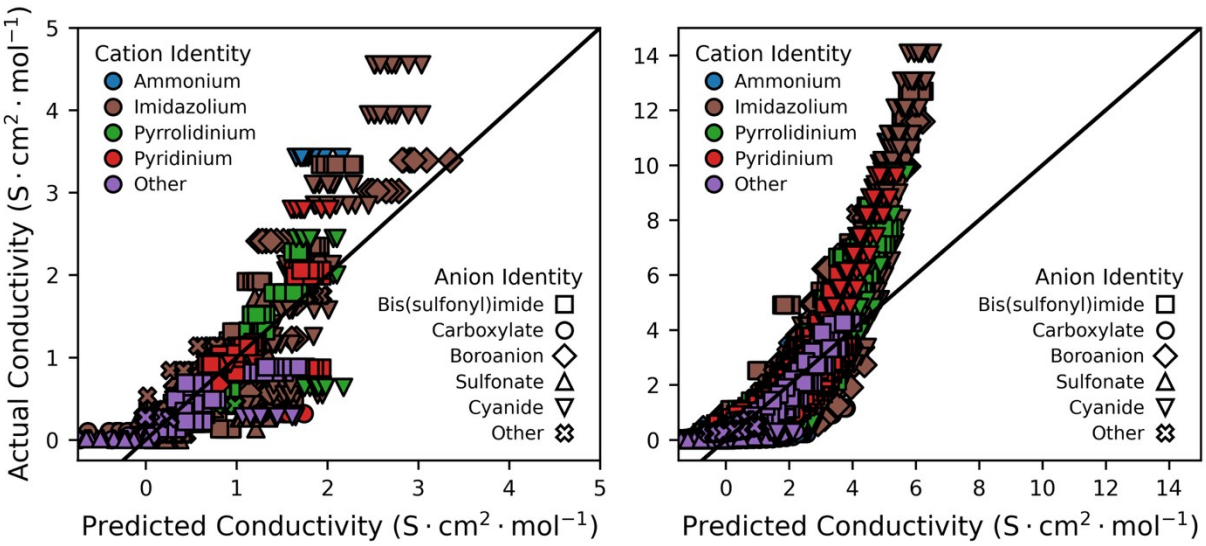

Figure S6. *Linear ridge regression train set parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

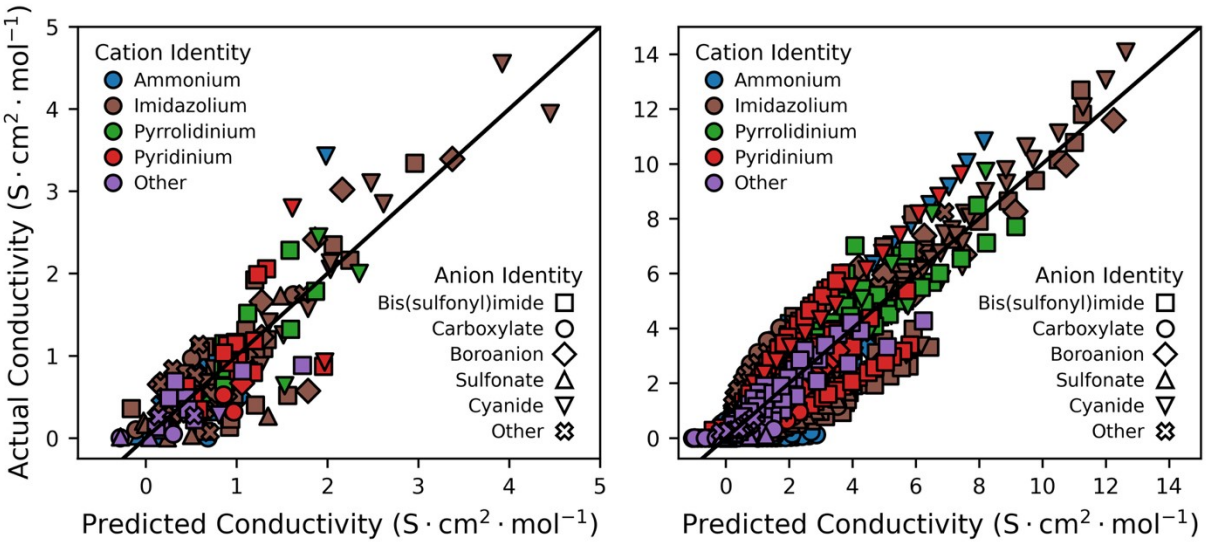

Figure S7. *RBF kernel ridge regression test set parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

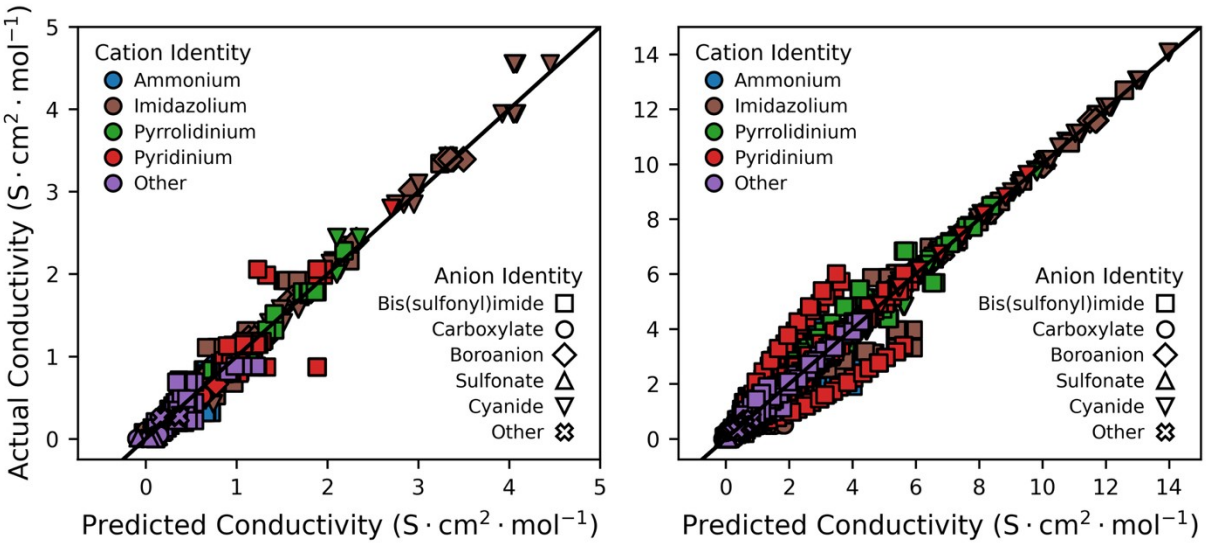

Figure S8. *RBF kernel ridge regression test train parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

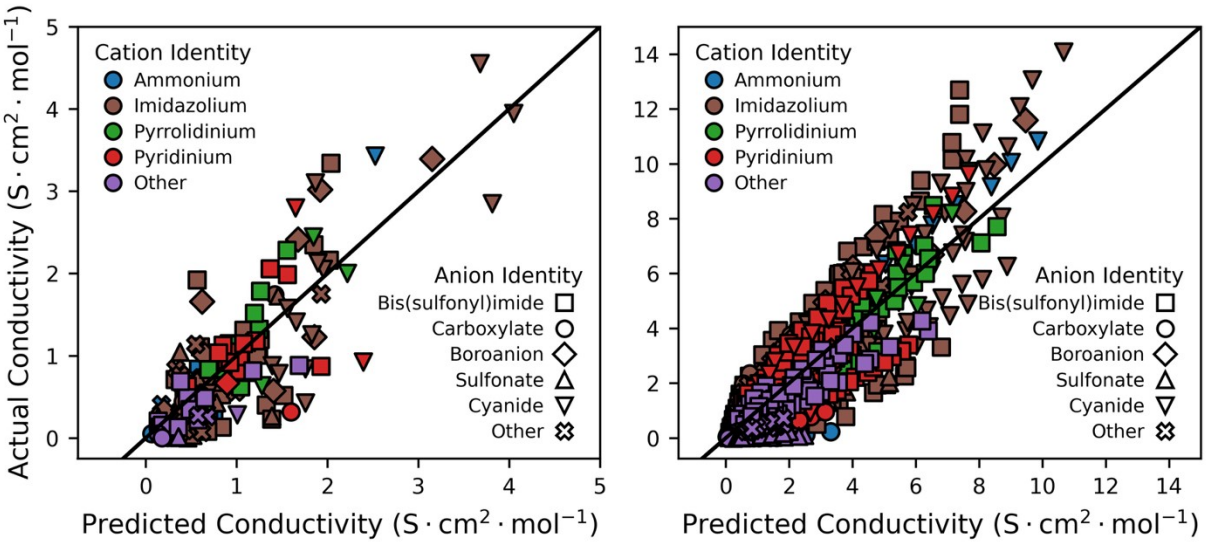

Figure S9. *Random forest regression test set parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

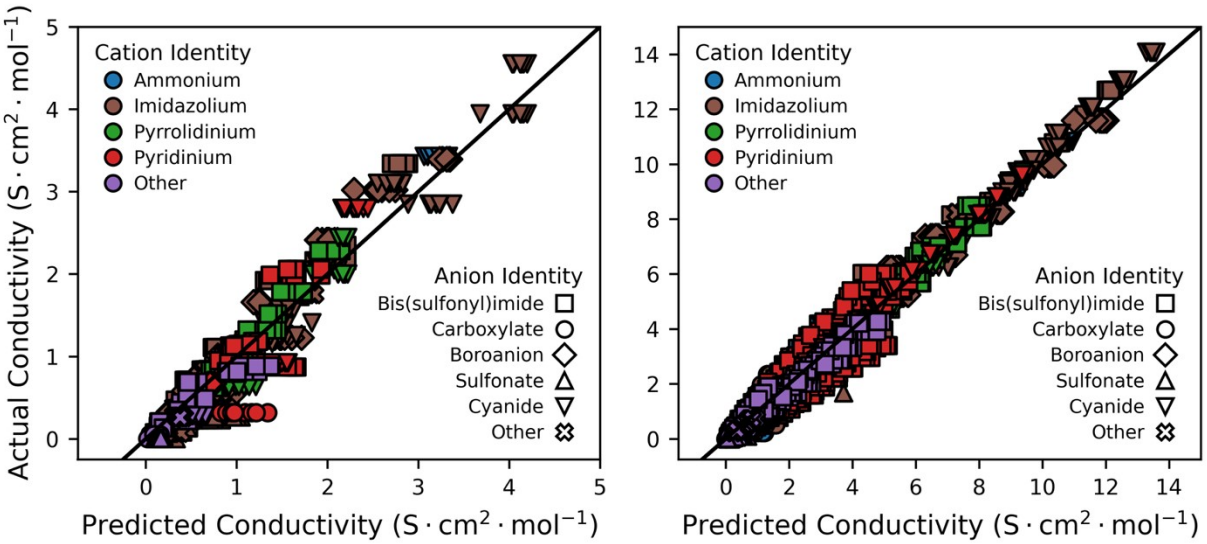

Figure S10. *Random forest regression train set parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

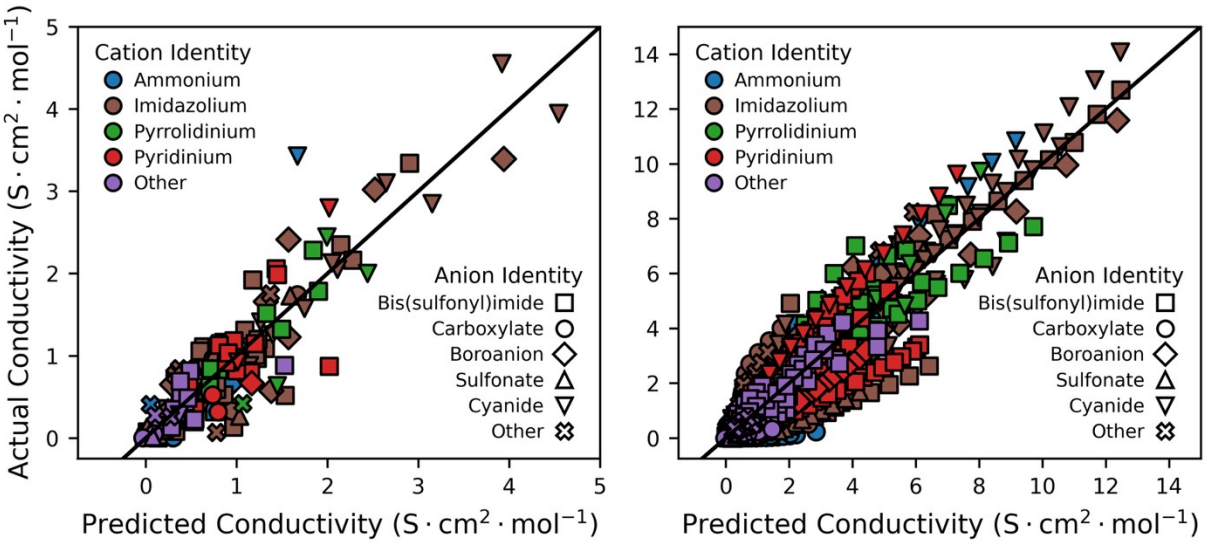

Figure S11. *Neural network regression test set parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

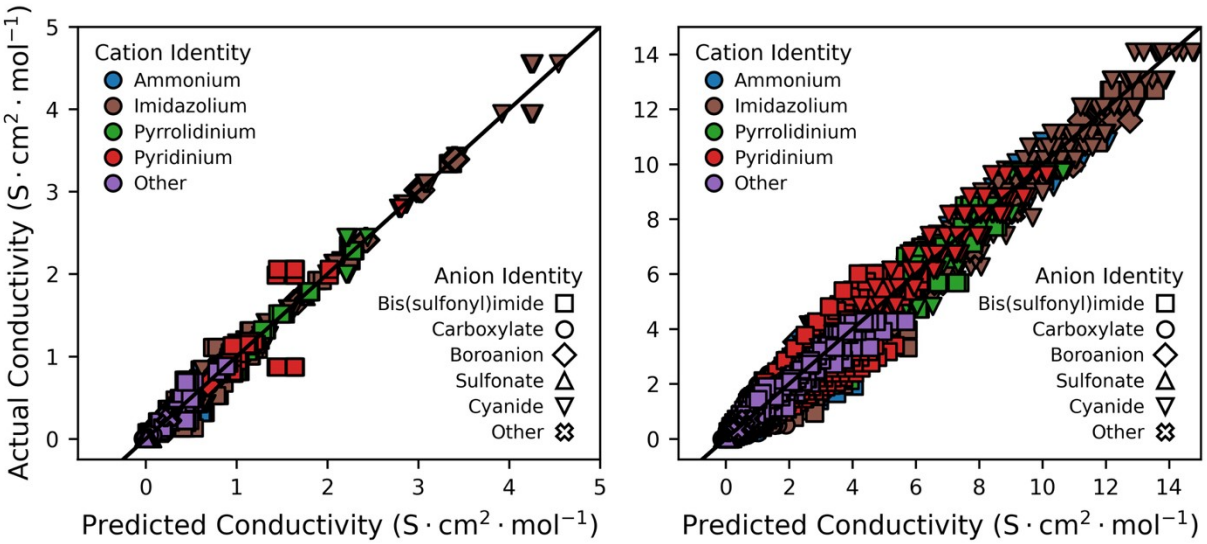

Figure S12. *Neural network regression train set parity plot using our custom SMARTS representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

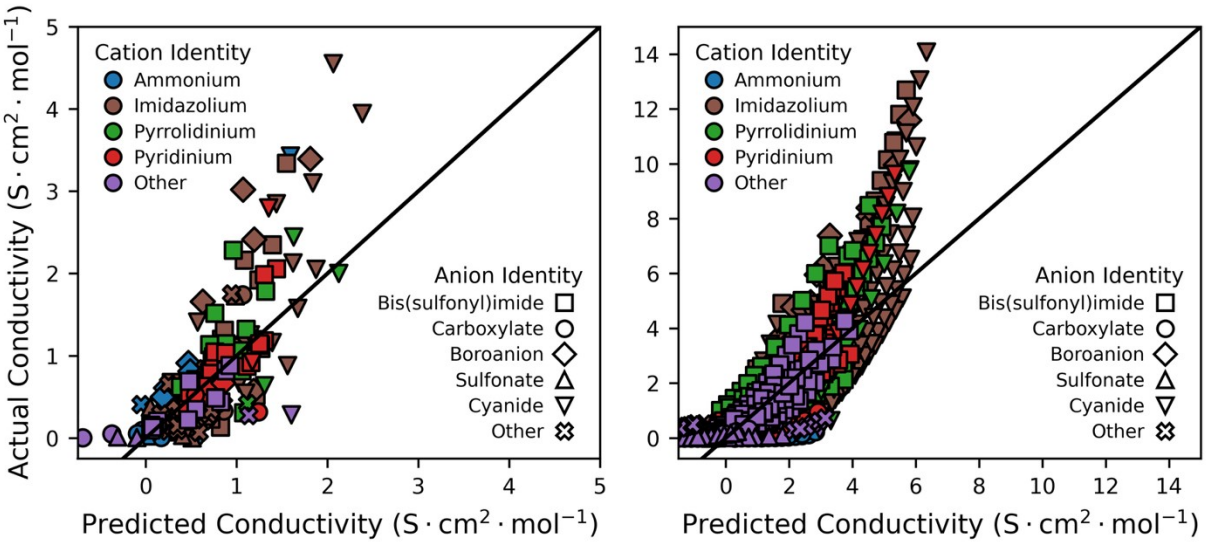

Figure S13. *Linear ridge regression test set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right)*

.

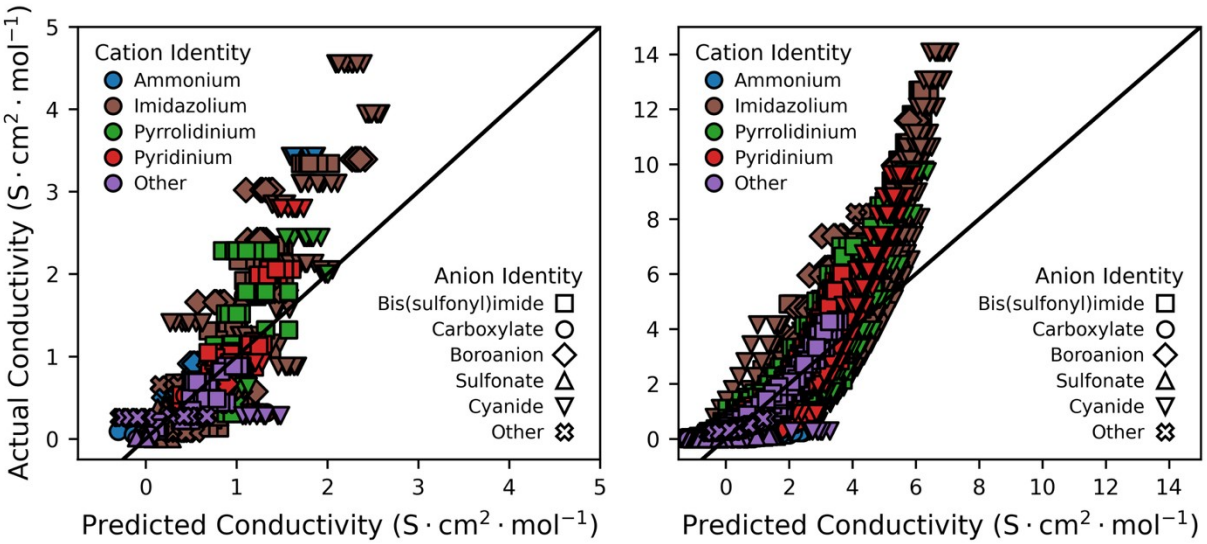

Figure S14. *Linear ridge regression train set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

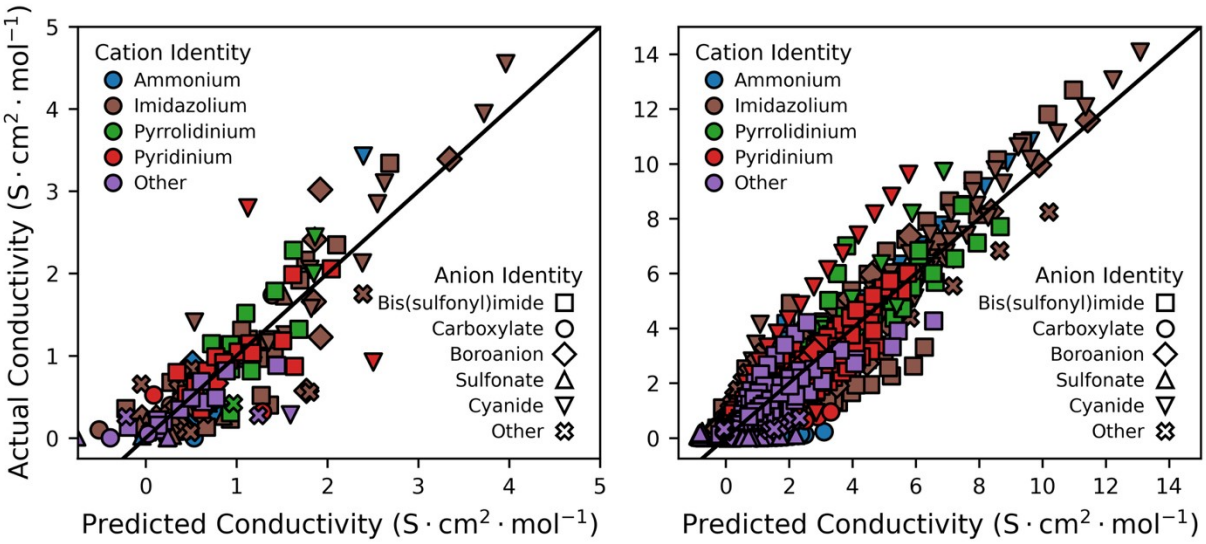

Figure S15. *RBF kernel ridge regression test set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

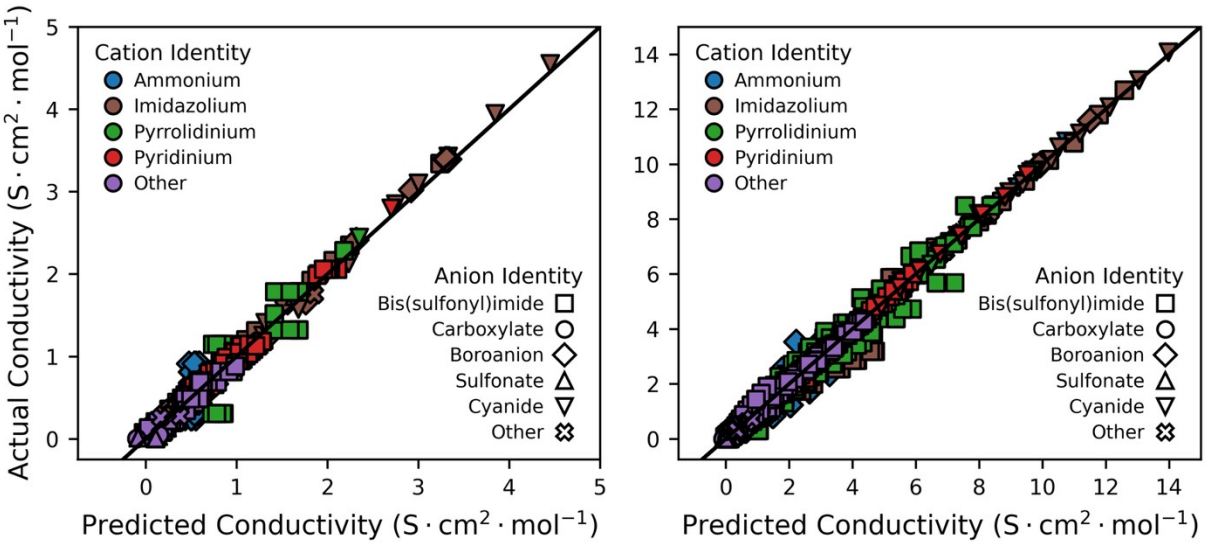

Figure S16. *RBF kernel ridge regression train set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

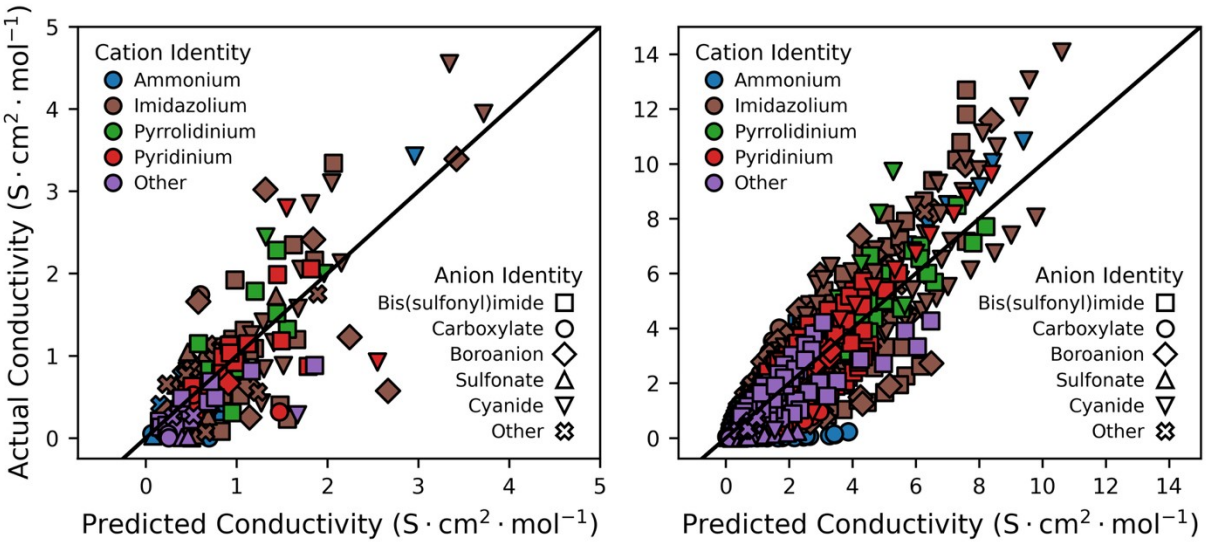

Figure S17. *Random forest regression test set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

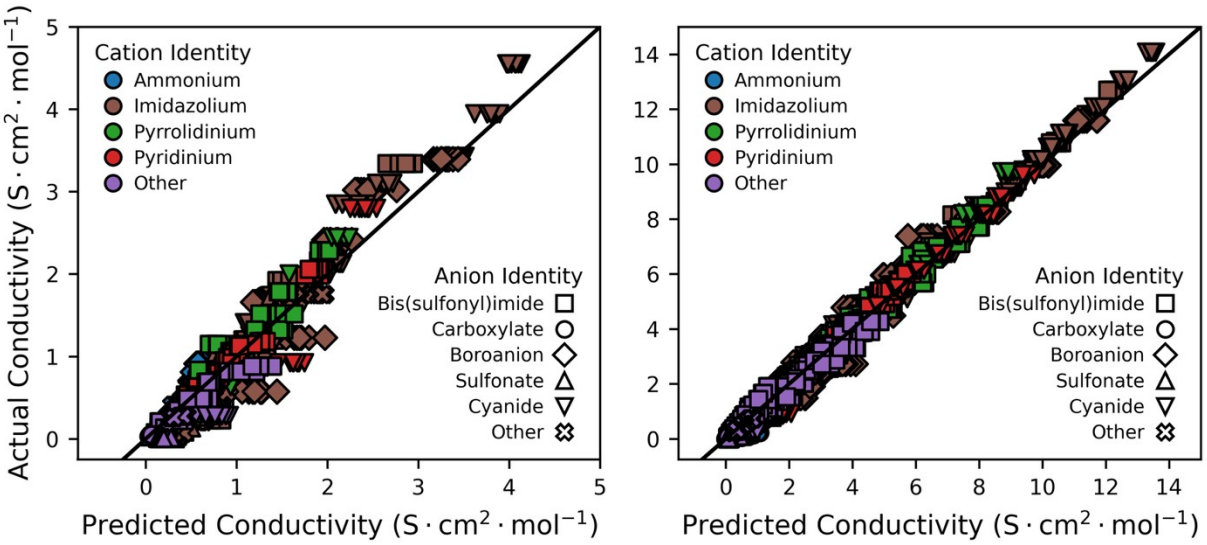

Figure S18. *Random forest regression train set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

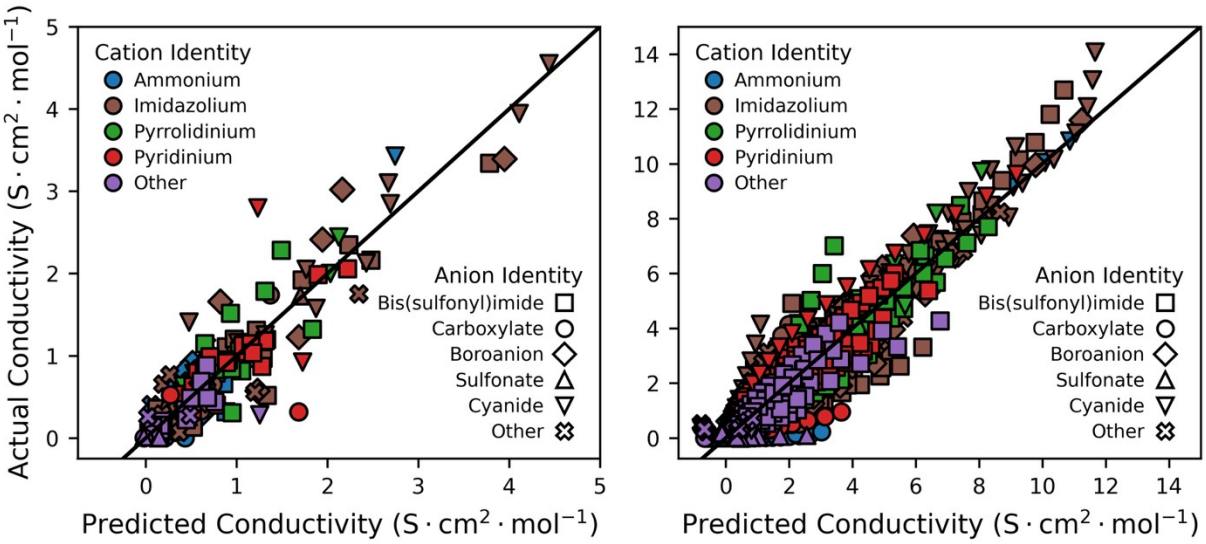

Figure S19. *Neural network regression test set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

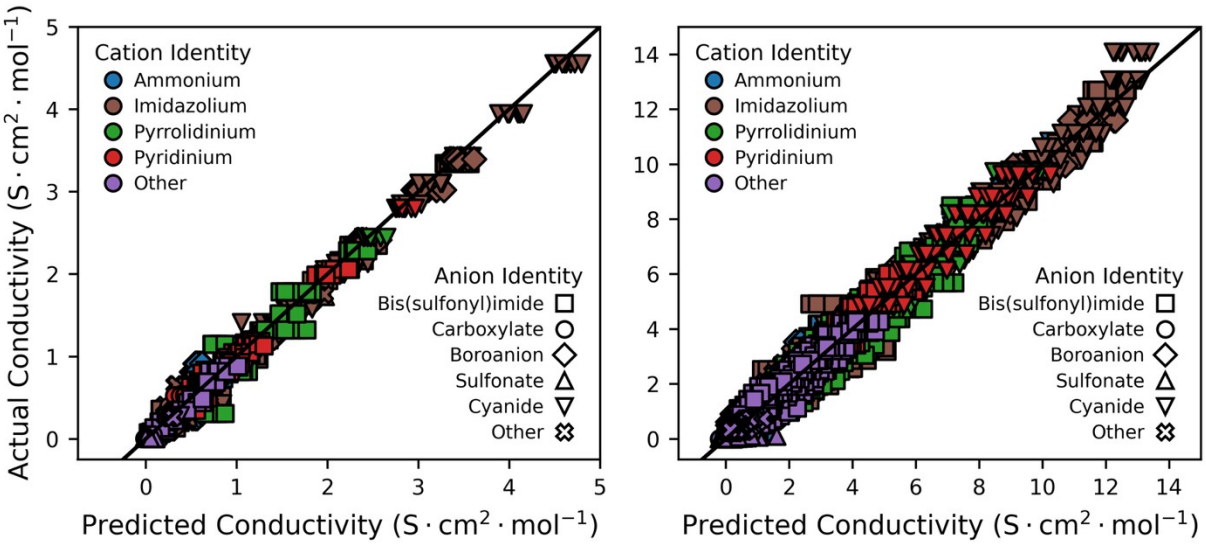

Figure S20. *Neural network regression train set parity plot using MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

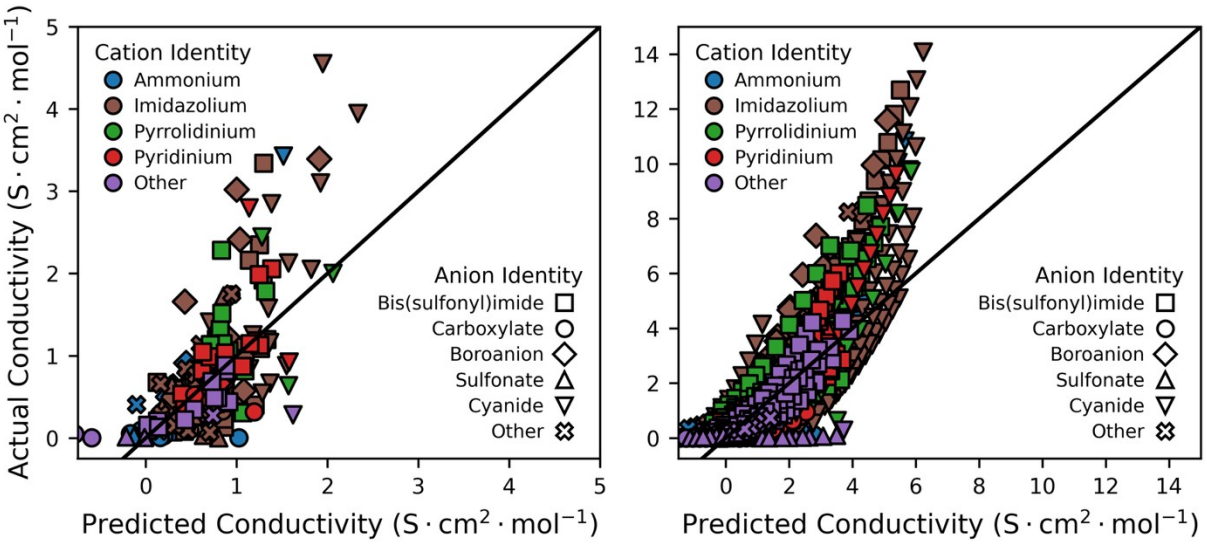

Figure S21. *Linear ridge regression test set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

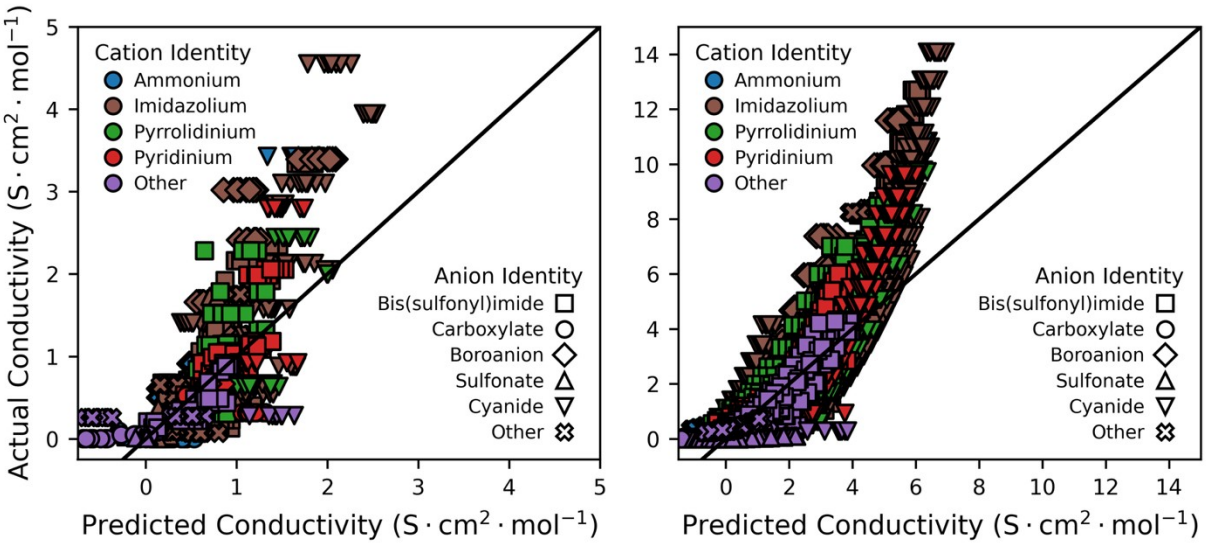

Figure S22. *Linear ridge regression train set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

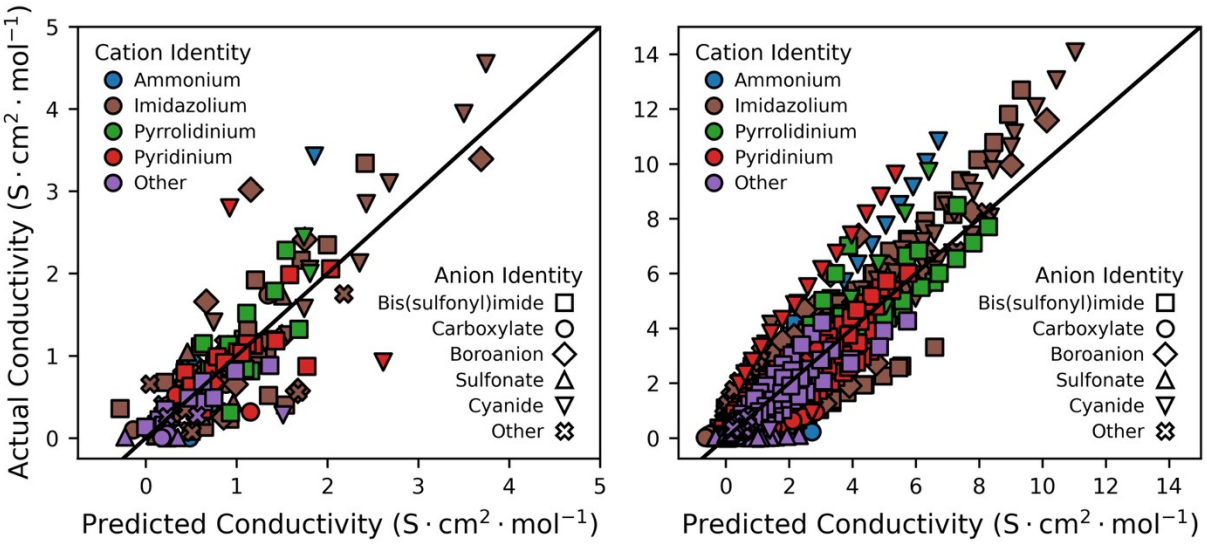

Figure S23. *RBF kernel ridge regression test set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

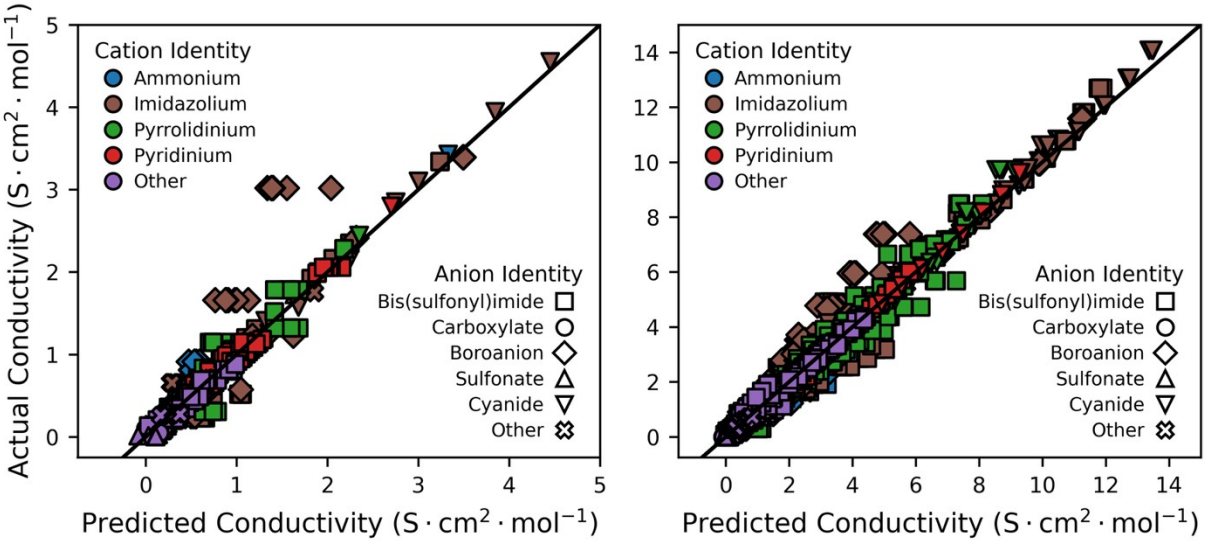

Figure S24. *RBF kernel ridge regression train set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

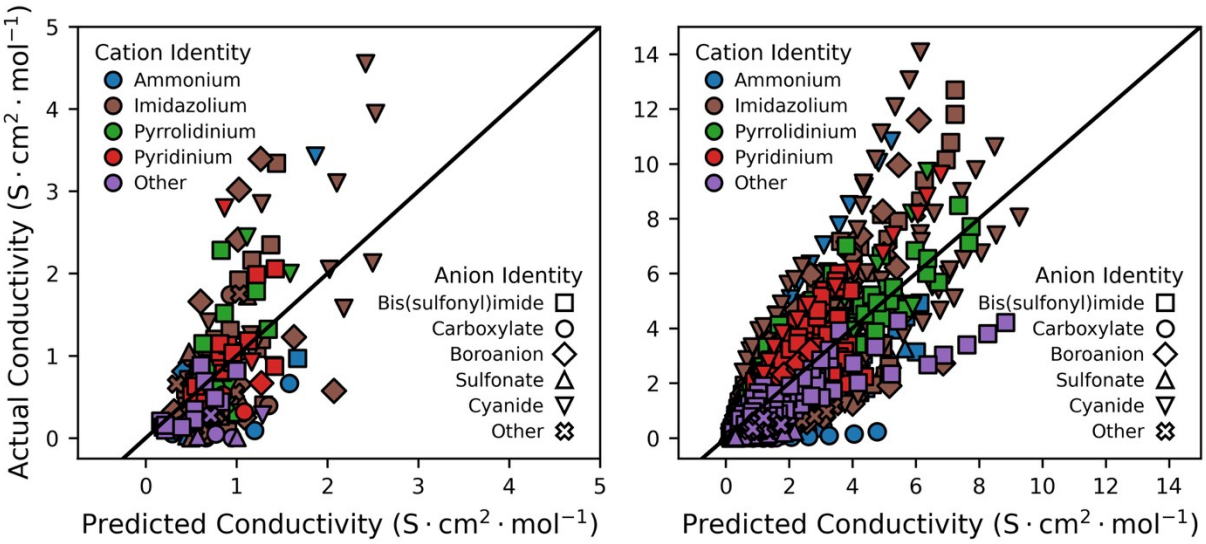

Figure S25. *Random forest regression test set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

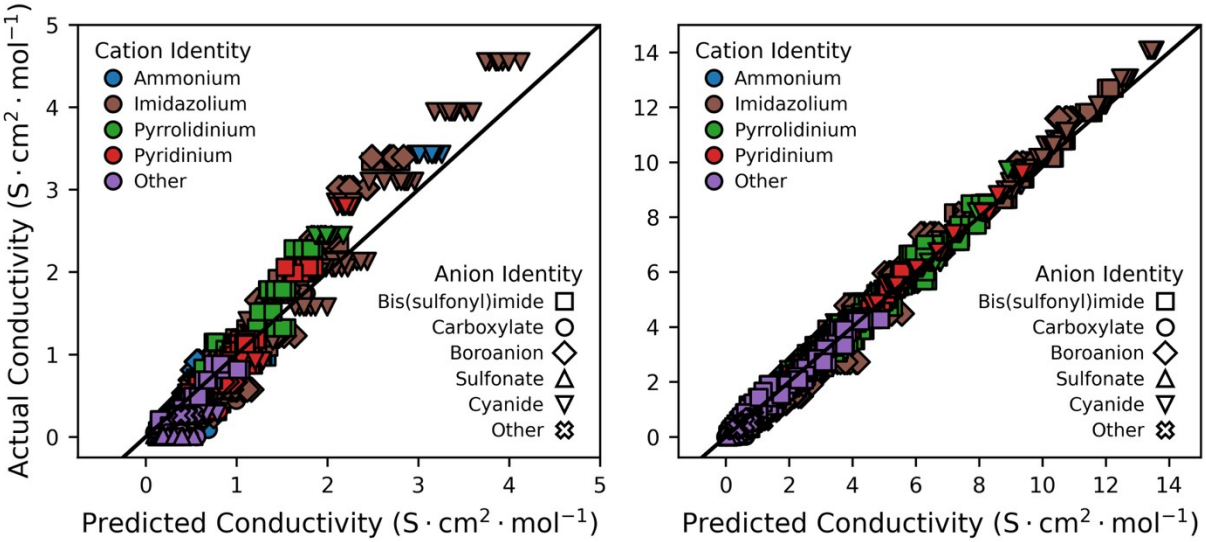

Figure S26. *RBF kernel ridge regression train set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

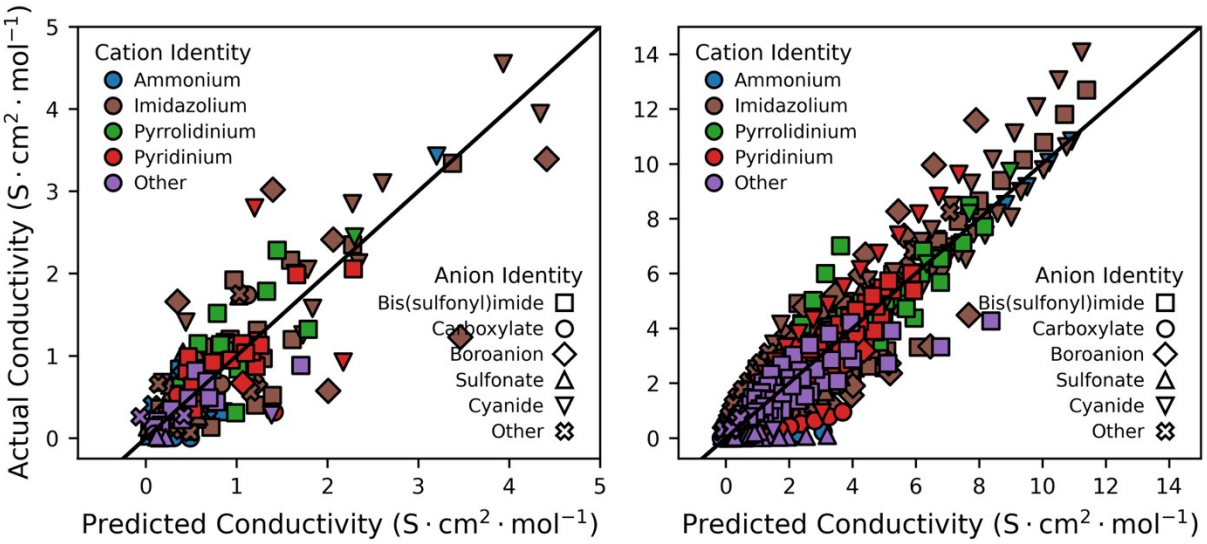

Figure S27. *Neural network regression test set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

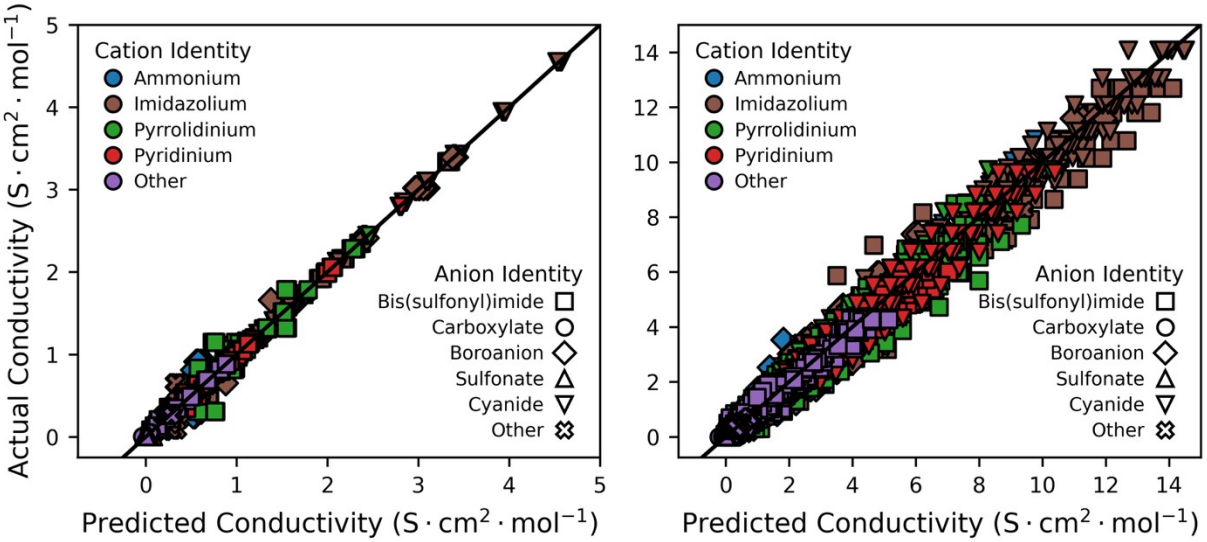

Figure S28. *Neural network regression train set parity plot using 36-dimension PCA MACCS Keys representation inputs colored by cations and shaped by anions for 298 K (left) and at all temperatures (right).*

Bibliography

1. SMARTS: A Language for Describing Molecular Patterns, http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.).
2. A. Goscinski, V. Principe, G. Fraux, S. Kliavinek, B. Helfrecht, P. Loche, M. Ceriotti and R. Cersonsky, *Open Research Europe*, 2023, **3**.
3. RDKit: Open-Source Cheminformatics Software, https://www.rdkit.org/, (accessed 8/8/2022, 2022).
4. B. A. Helfrecht, R. K. Cersonsky, G. Fraux and M. Ceriotti, *Mach Learn-Sci Techn*, 2020, **1**.