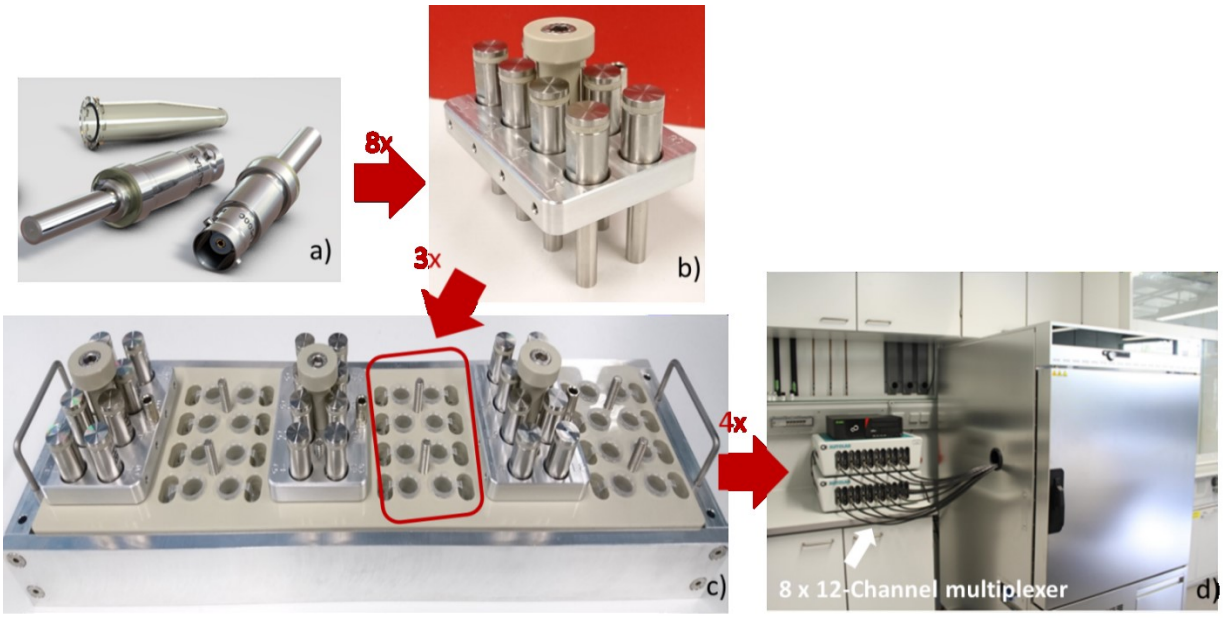
Supplementary Information

Learning the laws of lithium-ion electrolyte transport using symbolic regression

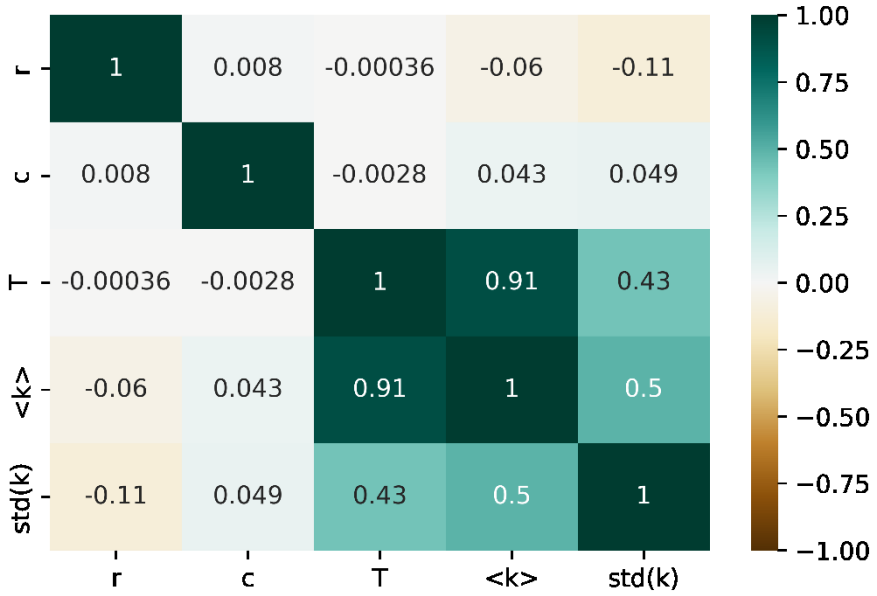
Eibar Flores[1*], Christian Wölke[2], Peng Yan[2], Martin Winter[2,3], Tejs Vegge[1], Isidora Cekic-Laskovic[2] and Arghya Bhowmik[1*]

[1]: Department of Energy Conversion and Storage, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
[2]: Helmholtz-Institute Münster (IEK-12), Forschungszentrum Jülich GmbH, Corrensstraße 46, 48149 Münster, Germany.
[3]: MEET Battery Research Center, University of Münster, Corrensstrasse 46, 48149 Münster, Germany.

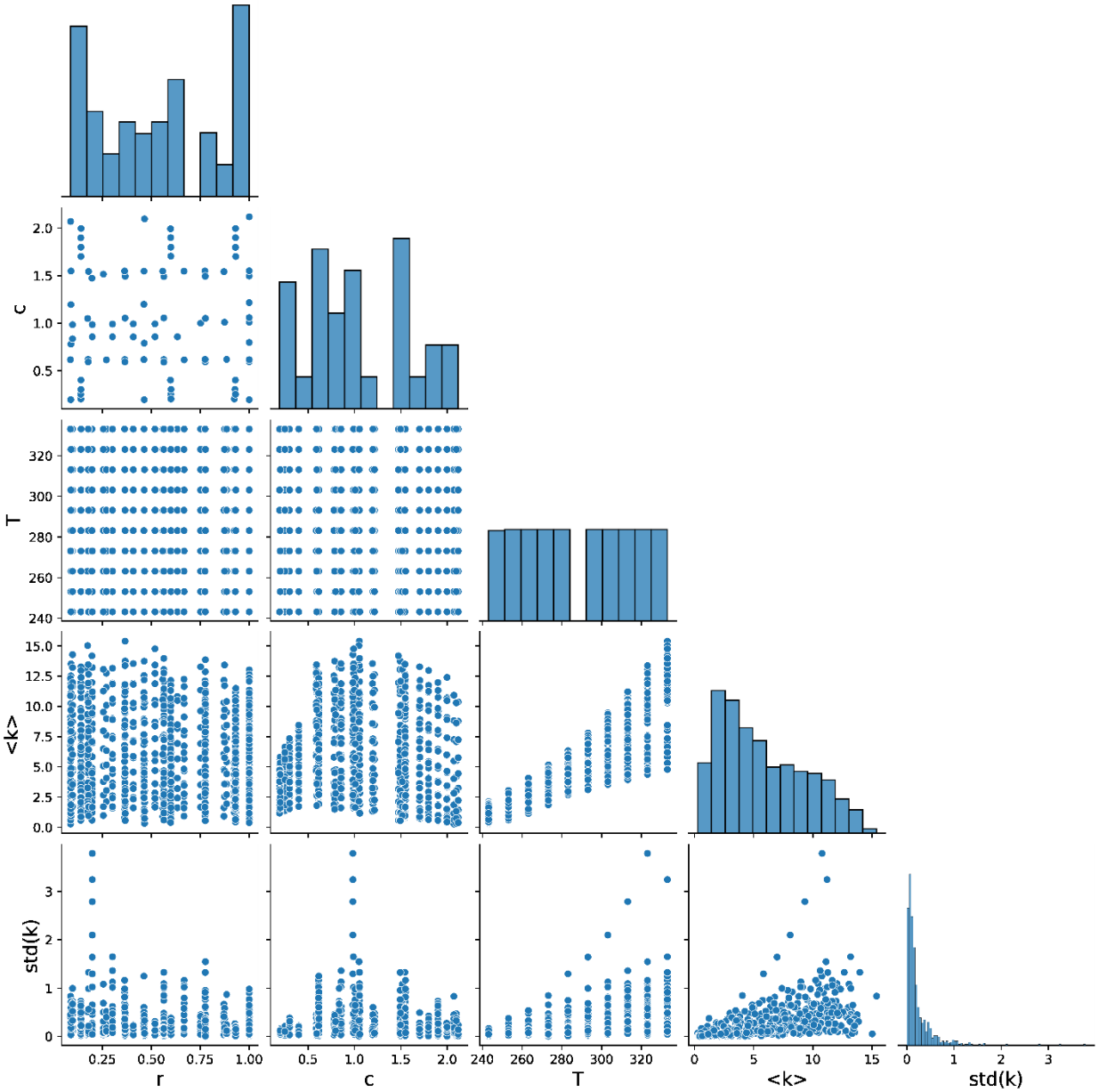
Corresponding authors
Eibar Flores: eibfl@dtu.dk
Arghya Bhowmik: arbh@dtu.dk



Supplementary Figure 1. High-throughput conductivity module comprising: a) in-house developed impedance electrode and Eppendorf tube b) small rack containing 8 electrodes c) big rack with 24 conductivity cells and d) potentiostat/galvanostat with 8 x 12-channel multiplexer and temperature chamber.



Supplementary Figure 2. Pearson correlation plots of the predictors: PC ratio *r*, conducting salt concentration *c*, temperature *T*, mean conductivities <*k*> and their corresponding standard deviations *std(k)*.



Supplementary Figure 3. Representation of the pairwise relationships in the data: PC ratio *r*, conducting salt concentration *c*, temperature *T*, mean conductivities <*k*> and their corresponding standard deviations *std(k)*. Off-diagonal subplots show trends between initial predictors and target. Diagonal elements show the univariate distribution of each quantity in the dataset.

Model	Feature engineering steps	Feature selection steps	Units	Operators set
Autofeat_1	6	5	T [1/K], c [mol/kg]	$x^{1/2}$
Autofeat_2	6			x^{-1}
Autofeat_3	6			x^2, x^3
Autofeat_4	6			e^x, log(x)
Autofeat_5	6			$x^{1/2}$, x^{-1}
Autofeat_6	5			x^2, x^3, e^x, log(x)

Supplementary Table 1. Hyperparameters of the feature generation step implemented via the AutoFeat Python package considering 5 feature selection steps and temperature and concentration in [1/K] and [mol/kg] units, respectively.

Model name	Functional form	Expression
Linear	Linear	$\kappa_0 + \beta_1 T + \beta_2 c + \beta_3 r$
Polynomial	Polynomial	$\kappa_0 + \beta_1 T + \beta_2 c + \beta_3 r + \beta_4 Tc + \beta_5 cr + \beta_6 rT + \beta_7 crT + \beta_8 T^2 + \beta_9 c^2 + \beta_{10} r^2 + \beta_{11} T^2 c + \beta_{12} Tc^2 + \beta_{13} c^2 r + \beta_{14} cr^2 + \beta_{15} T^2 r + \beta_{16} Tr^2 + \beta_{17} T^3 + \beta_{18} c^3 + \beta_{19} r^3$
Arrhenius	Exponential	$\kappa_0 \exp(\beta_1 T + \beta_2 c + \beta_3 r)$
Arrhenius Poly	Exponential / polynomial	$\kappa_0 \exp(\beta_1 T + \beta_2 c + \beta_3 r + \beta_4 Tc + \beta_5 cr + \beta_6 rT + \beta_7 crT + \beta_8 T^2 + \beta_9 c^2 + \beta_{10} r^2 + \beta_{11} T^2 c + \beta_{12} Tc^2 + \beta_{13} c^2 r + \beta_{14} cr^2 + \beta_{15} T^2 r + \beta_{16} Tr^2 + \beta_{17} T^3 + \beta_{18} c^3 + \beta_{19} r^3)$

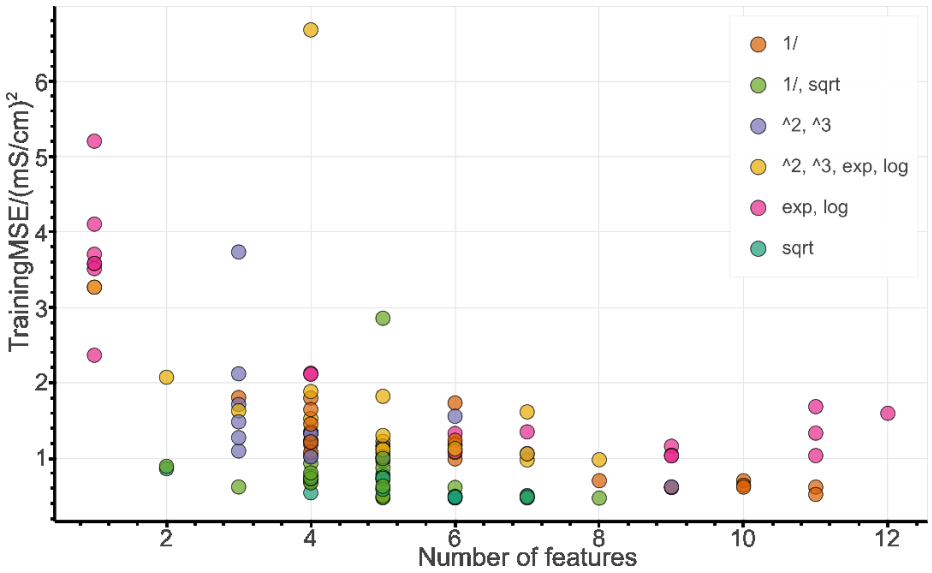
Supplementary Table 2. Functional expression of benchmark models. β_i represents the i^{th} fitting coefficient.

Hyperparameter	Value
N-fold cross-validation	10
Range of alphas	10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100, 10^3, 10^4, 10^5, 10^6
Maximum number of iterations	10^5

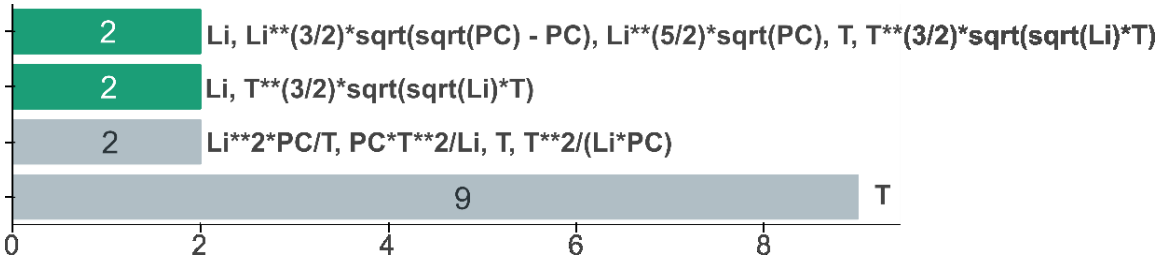
Supplementary Table 3. Hyperparameters of the feature selection step implemented via a Cross-Validated Lasso Regressor as implemented in the scikit-learn python package.

Expression	Val. MSE	Val. r^2	N. features	Trans form. set
$0.00028T^{\frac{3}{2}}\sqrt{T\sqrt{c}} - 0.00479Tc\sqrt{cr} - 0.0407T - 1.79c^{\frac{3}{2}}\sqrt{\sqrt{r}-r} - 3.8c + 1.88\sqrt{cr\left(\sqrt{r}-r\right)}$	0.399	0.968	6	sqrt
$0.000254T^{\frac{3}{2}}\sqrt{T\sqrt{c}} - 0.0291T - \frac{2.66\cdot10^{-5}T}{c\sqrt{\frac{c}{T}}} - 0.113c^{\frac{5}{2}}\sqrt{r} - 0.251c^{\frac{3}{2}}\sqrt{\sqrt{r}-r} - 5.59c - 0.474r$	0.430	0.966	7	1/, sqrt
$0.000256T^{\frac{3}{2}}\sqrt{T\sqrt{c}} + \frac{9.87\cdot10^{-5}T^{\frac{3}{2}}}{\sqrt[4]{r}} - 0.0337T - \frac{2.14\cdot10^{-5}T}{c\sqrt{\frac{c}{T}}} - 0.193c^{\frac{5}{2}}\sqrt{r} - \frac{0.0232c^{\frac{5}{2}}}{\sqrt{r}} - 5.33c - \frac{0.00193}{\sqrt{r}-r-\frac{1}{r+\frac{1}{r}}}$	0.432	0.966	8	1/, sqrt
$0.000254T^{\frac{3}{2}}\sqrt{T\sqrt{c}} + \frac{0.000105T^{\frac{3}{2}}}{\sqrt[4]{r}} - 0.0317T - \frac{2.95\cdot10^{-5}T}{c\sqrt{\frac{c}{T}}} - 6.0c - \frac{0.00164}{\sqrt{r}-r-\frac{1}{r+\frac{1}{r}}} - \frac{0.00188}{r^{\frac{3}{2}}-\frac{1}{\sqrt{r}+\frac{1}{r}}}$	0.434	0.966	7	1/, sqrt
$0.000192T^{\frac{3}{2}}\sqrt{T\sqrt{c}} - 0.0072T - \frac{5.25\cdot10^{-5}T}{c\sqrt{\frac{c}{T}}} - 0.173c^{\frac{5}{2}}\sqrt{r} - 0.527c^{\frac{3}{2}}\sqrt{\sqrt{r}-r} - 3.67c - 0.514r - \frac{13600.0\sqrt{\frac{\sqrt{c}}{T}}}{T}$	0.446	0.965	8	1/, sqrt
$0.000267T^{\frac{3}{2}}\sqrt{T\sqrt{c}} + 0.000124T^{\frac{3}{2}}\sqrt{\sqrt{r}-r} - 0.00551T\sqrt{c}\left(-\sqrt{r}+r\right) - 0.0381T - 0.373c^{\frac{5}{2}}\sqrt{r} - 1.28c^{\frac{3}{2}}\sqrt{\sqrt{r}-r} - 4.36c$	0.450	0.964	7	sqrt
$0.000258T^{\frac{3}{2}}\sqrt{T\sqrt{c}} - 0.0307T - \frac{2.4\cdot10^{-5}T}{c\sqrt{\frac{c}{T}}} - 0.181c^{\frac{5}{2}}\sqrt{r} - \frac{0.0381c^{\frac{5}{2}}}{\sqrt{r}} - 5.47c - 0.456r$	0.451	0.964	7	1/, sqrt
$0.000268T^{\frac{3}{2}}\sqrt{T\sqrt{c}} + 0.000278T^{\frac{3}{2}}\sqrt{\sqrt{r}-r} - 0.0391T - 0.444c^{\frac{5}{2}}\sqrt{r} - 1.19c^{\frac{3}{2}}\sqrt{\sqrt{r}-r} - 4.2c$	0.453	0.964	6	sqrt
$0.00026T^{\frac{3}{2}}\sqrt{T\sqrt{c}} - 0.0319T - \frac{1.87\cdot10^{-5}T}{c\sqrt{\frac{c}{T}}} - 0.215c^{\frac{5}{2}}\sqrt{r} - \frac{0.0349c^{\frac{5}{2}}}{\sqrt{r}} - 5.3c - 0.437r$	0.456	0.964	7	1/, sqrt

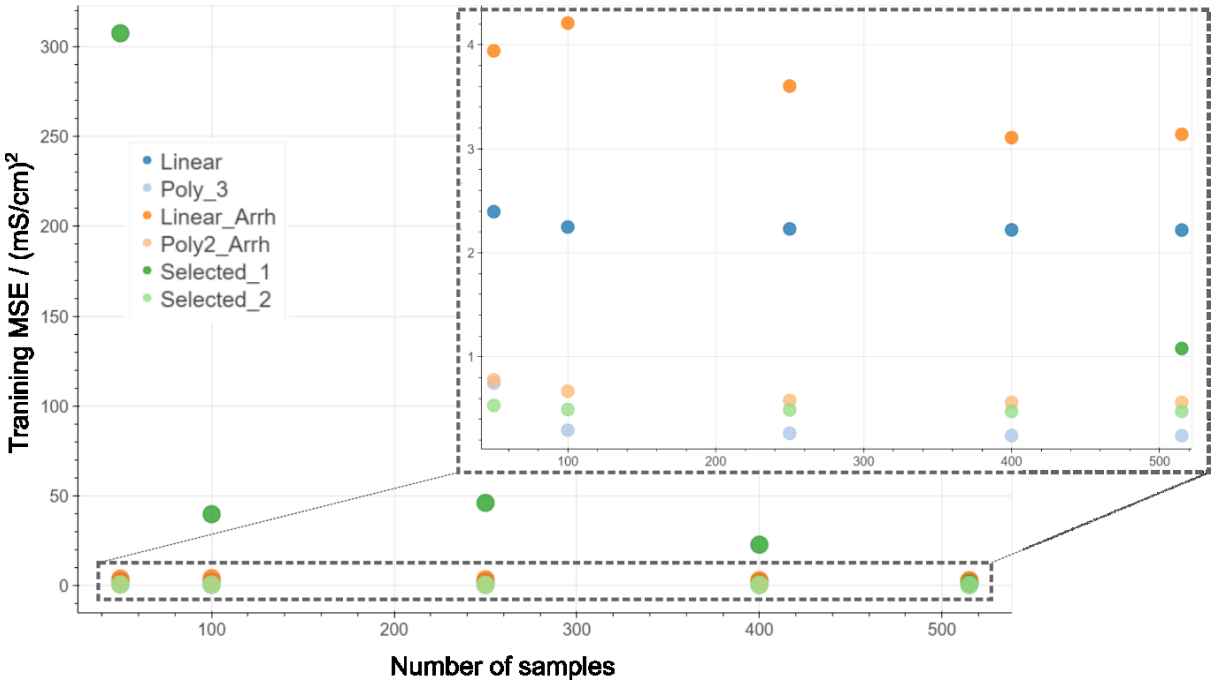
Supplementary Table 4. Discovered expressions with the lowest validation MSE, with the intercept = 0 constraint. Each of these expressions are found only once across subsamples of the training set.



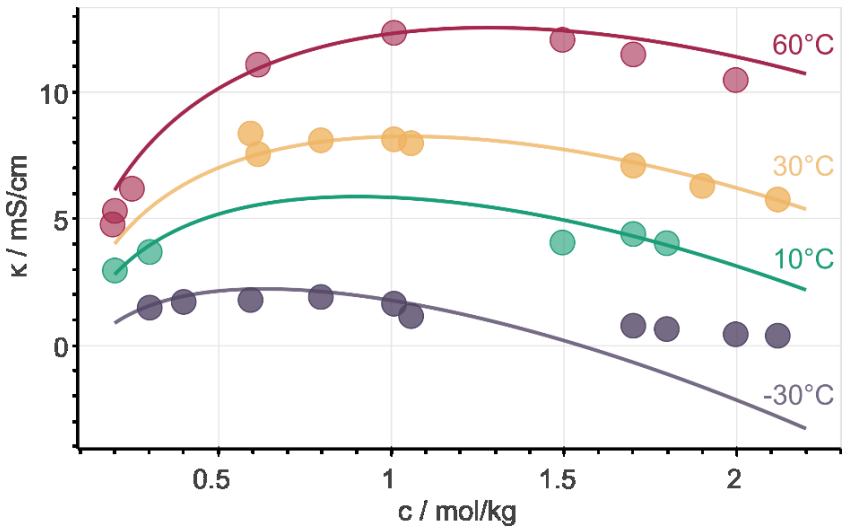
Supplementary Figure 4. Accuracy vs parsimony of unconstrained models (i.e. y0 is allowed to vary freely). Each data point represents an expression, whose color indicates its parent transformation set.



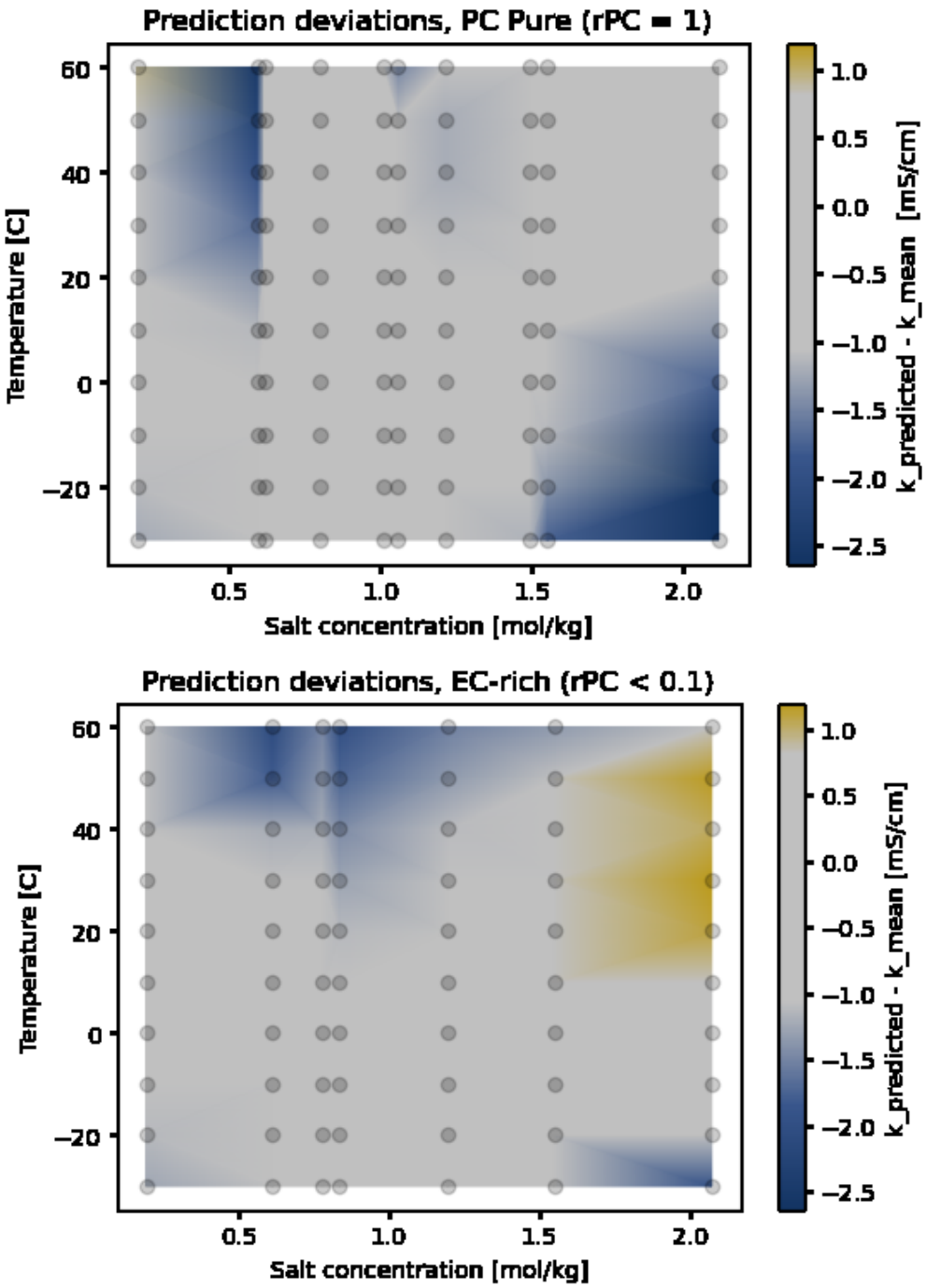
Supplementary Figure 5. Most frequent unconstrained models throughout the 20 training sessions. The only consistent model is dependent only on temperature and so it is not viable as it fails to capture the effects of salt concentration and solvent mixture.



Supplementary Figure 6. Learning curves. Selected model 1 (dark green) was trained with the y=0 constrain, while Selected model 2 (light green) without it. The figure inset zooms into the low-MSE region of the plot.



Supplementary Figure 7. Fit of the selected model, trained with free intercept on the withheld (validation and test) set at r=1.0.



Supplementary Figure 8. Deviations between the predicted and measured conductivities as a function of salt concentration and temperature. Top: deviations in PC-pure electrolytes. Bottom: deviations in EC-rich electrolytes. The colour code indicates zones where the model predicts conductivities that are i) higher than the measured conductivity, in yellow, ii) lower than the measured conductivity, in blue, and, iii) comparable to the measurement error, in grey. The magnitude of the measurement error is ca. 0.8 mS/cm (quantile 0.95 from the measurement deviations in Figure 3a in the main manuscript). The circular markers indicate locations where measurements are available in the dataset.