Articles
doi.org/10.1002/batt.202100031

Chemistry Europe
European Chemical Societies Publishing

Special Collection

Supervised Machine Learning-Based Classification of Li–S Battery Electrolytes

Steffen Jeschke*[a] and Patrik Johansson[a, b]

Machine learning (ML) approaches have the potential to create a paradigm shift in science, especially for multi-variable problems at different levels. Modern battery R&D is an area intrinsically dependent on proper understanding of many different molecular level phenomena and processes alongside evaluation of application level performance: energy, power, efficiency, life-length, etc. One very promising battery technology is Li–S batteries, but the polysulfide solubility in the electrolyte must be managed. Today, many different electrolyte compositions and concepts are evaluated, but often in a more or less trial-and-error fashion. Herein, we show how supervised ML can be applied to accurately classify different Li–S battery electrolytes a priori based on predicting polysulfide solubility. The developed framework is a combined density functional theory (DFT) and statistical mechanics (COSMO-RS) based quantitative structure-property relationship (QSPR) model which easily can be extended to other battery technologies and electrolyte properties.

1. Introduction

Artificial intelligence (AI) has been around for quite some time, even if the term was not coined until 1956.[1] AI is a collection of technologies combining *e.g.* data handling, machine learning (ML) algorithms including artificial neural networks (ANN), and computational power.[2] Today, AI is making significant contributions to our day-to-day life, but furthermore also addresses the global challenges of climate change, limited natural resources, poverty&health, and accessible and clean energy for all. With respect to the latter, advances in multi-scale modelling including AI and ML/ANN have been used in modern battery R&D, most notably for lithium-ion batteries (LIBs).[3] ML is currently heavily applied to very fundamental LIB materials research[4–6] for example within the MGI,[7] BIG-MAP,[8] and larger concerted efforts such as Battery 2030+[9] and the Faraday Institution,[10] as well as to more practical problems, such as optimizing the production processes[11,12] and the balance between usage flexibility and life-length for electrification of vehicles.[13,14]

The power and versatility of the application of ML to LIB research was recently exemplified by the first ever AI written book.[15] However, AI and ML might play an even more pertinent role when it comes to R&D for next generation batteries (NGBs). Today, the expensive trial-and-error experimental development of NGB materials is supported and fast-tracked by ML-driven computational chemistry (*de novo* design) and related quantitative structure-activity/property relationship (QSAR/QSPR).

QSAR/QSPR was developed to understand mechanisms by combining experimental data and/or computed molecular descriptors to explain interactions and proved to be an effective predictive tool, controversially,[16] even in combination with statistical approaches unique for "Big Data". The accuracy as well as the interpretability is largely driven by the molecular descriptors used, which are hence subject to dedicated research focusing on their development and selection.[17]

The Li–S battery is a NGB with a large built-in complexity,[18] and thus many descriptors, at the molecular level; the elemental sulfur contained in the composite C/S cathode dissolves into the electrolyte and is converted to various polysulfides during battery operation. Hence the very composition of the electrolyte varies as a function of the battery state-of-charge (SOC). To predict and control the solubilities and understand the speciation is an at large unresolved problem whatever the Li–S battery electrolyte concept employed and there are many concurrent lines of development.[19–22]

For this kind of systems the conductor-like screening model for real solvents (COSMO-RS)[23–27] approach is very suited; it combines a density functional theory (DFT) based continuum solvation approach with statistical thermodynamics to obtain multiple molecular descriptors of physical meaning, all related to molecular interactions and electronic properties. Herein, we combine different supervised ML-approaches with COSMO-RS-based descriptors to first build a polysulfide solubility regression model and second, make a classification of different Li–S electrolytes as salt-in-solvent, solvent-in-salt, or solvated ionic

[a] *Dr. S. Jeschke, Prof. P. Johansson*
Department of Physics
Chalmers University of Technology
412 96 Gothenburg, Sweden
E-mail: steffen.jeschke@rmit.edu.au
Patrik.johansson@chalmers.se
[b] *Prof. P. Johansson*
Alistore-ERI European Research Institute
CNRS FR
3104, Hub de l'Energie 15 Rue Baudelocque, 80039 Amiens, France

Supporting information for this article is available on the WWW under https://doi.org/10.1002/batt.202100031

Special Collection *An invited contribution to a Special Collection on Artificial Intelligence in Electrochemical Energy Storage*

liquid (SIL) electrolytes (Figure 1a–c). The solubility data of polysulfides to build regression models is hardly available due to the complex nature of such analytical experiments. A classification of different electrolytes regarding polysulfide solubility, however, is a conceptual approach and potentially suitable for a first assessment. The predictability of both models is discussed and compared.

2. Results and Discussion

2.1. $\mu(\sigma)$ as Electrolyte Descriptor

In the first step of the QSPR/QSAR model development proper descriptors for the electrolytes are obtained from the COSMO-RS framework. The COSMO-RS model approach provides access to several molecular descriptors and parameters, some of physical relevance and others without. The core framework is the calculation of molecular surface screening charges, σ. These are often illustrated in the form of a σ-profile, depicting the probability $p(\sigma)$ to find a molecular surface segment of value σ on the surface of a molecule. Screening charge densities $> \pm 0.01$ eÅ^{-2} are considered to participate in hydrogen-bonds (Brønsted-theory) or electron-donor-acceptor interactions (Lewis-theory). Additionally, the σ-potential, $\mu(\sigma)$, quantifies the affinity towards a σ of a specific value. In Figure 1d the physical meaning of the interplay between $\mu(\sigma)$ and $p(\sigma)$ is illustrated for a SIL electrolyte, [Li(G4)]TFSI, and a polysulfide solute, Li_2S_8.

For the $[Li(G4)]^+$ complexes with different degrees of G4 ether oxygen atom coordination to Li^+, $n = 1$–5, the $\mu(\sigma)$ for $\sigma = -0.03$ increases, indicating that the ability of the electrolyte to interact with additional Li^+, with the screening charge σ

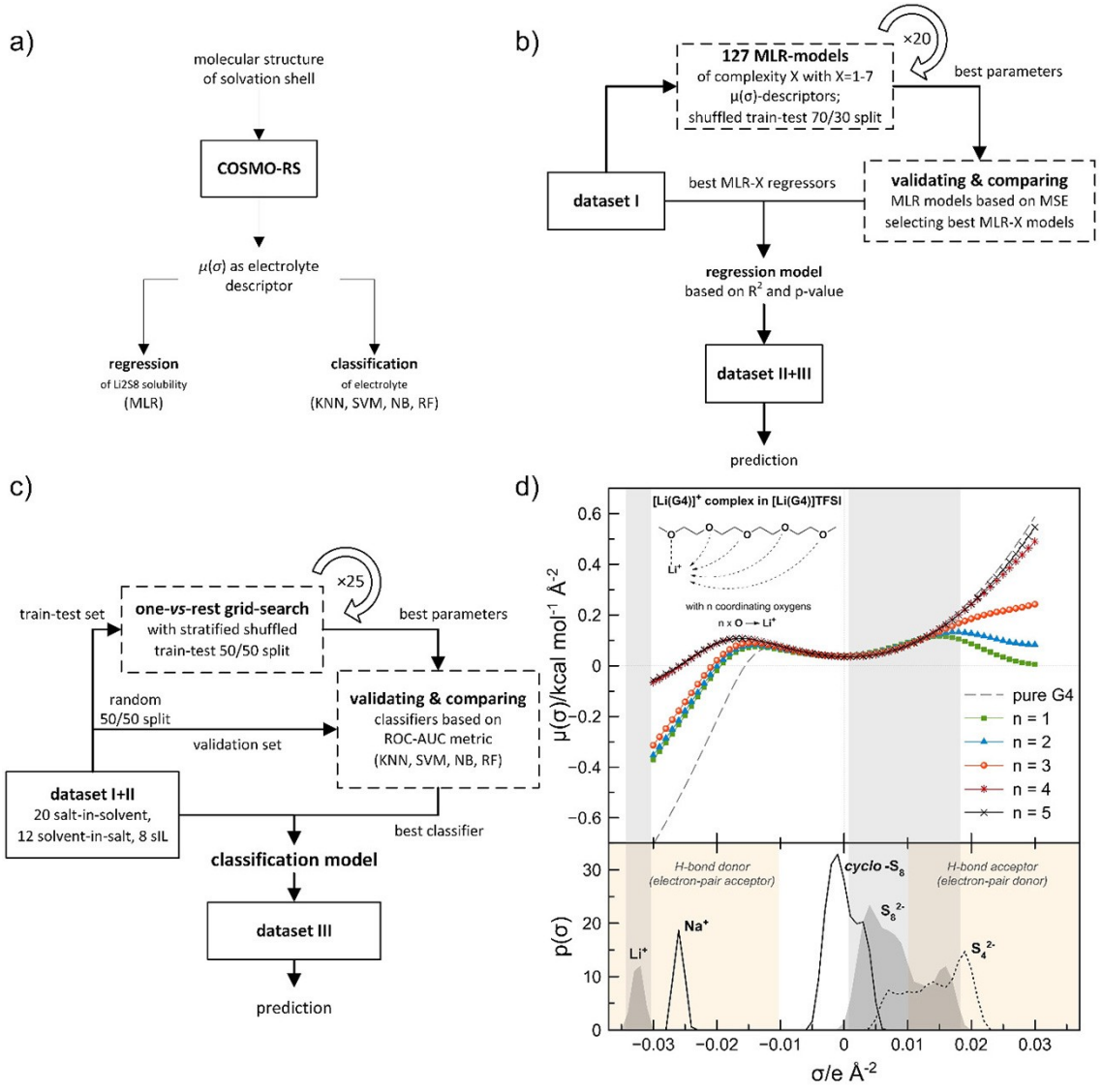

Figure 1. a–c) Our conceptual approach of using COSMO-RS to obtain descriptors, develop models and validate them for classification of Li–S battery electrolytes, and d) examples of σ-potentials for the case of a SIL electrolyte; [Li(G4)]TFSI: $[Li(G4)]^+$ complexes with different degrees of coordination ($n = 1$–5) by the G4 ether oxygen atoms (top) and σ-profiles for elemental sulfur, Li^+ and Na^+ cations and polysulfide anions (bottom).

distributed between −0.03 and −0.035 eÅ^{-2}, declines. The $\mu(\sigma)$ of the electrolytes are nearly identical for interactions with σ between 0 and 0.02 eÅ^{-2}, the σ boundaries of the S_8^{2-} polysulfide anion. Accordingly, the ability of the electrolyte (solvent) to interact with Li^+ determines the polysulfide solubility, here using as proxy Li_2S_8,[27] while the interaction with the anion seems much less important, in agreement with the common-ion effect[28] and the "excess-glyme-theory".[29,30] Hence, $\mu(\sigma)$ is a useful molecular level descriptor to directly correlate the ability of an electrolyte to solvate polysulfides formed during the cycling of a Li–S battery, *i.e.* varying as function of SOC.

By default, the computed $\mu(\sigma)$ σ-potential is based on 61 datapoints between ±0.03 eÅ^{-2}. As several are very similar (Figure 1d) the dimensionality was here henceforth reduced to only seven (7) datapoints: $\sigma = \pm 0.03$, ±0.02, ±0.01, and 0.0 eÅ^{-2} as $\mu(\sigma)$-descriptors.

2.2. Regression of Polysulfide Solubility

The development of a regression model is based on electrolytes for which experimental data for Li_2S_8 solubility and the Li^+ coordination details are known. First, we validate the method by screening 127 possible combinations of the $\mu(\sigma)$-descriptors to build MLR-X models of increasing complexity X using one to seven $\mu(\sigma)$-descriptors: $X = 1$–7; $X = 1$: 7 combinations, $X = 2$: 21, $X = 3$: 35, $X = 4$: 35, $X = 5$: 21, $X = 6$: 7, $X = 7$: 1) (Figure 1b). Due to the small size of dataset **I**, a shuffled dataset split into training and test set at a ratio of 70/30, respectively, was chosen to improve the generalizability of the obtained MLR-models. The MLR-X models of the same complexity X were evaluated using the mean squared error (MSE) for the test splits of the 30 random dataset splits, MSE_{test}, which indicates that an accurate regression model has been obtained from the training data, and ΔMSE, the difference between the MSE_{train} and MSE_{test}, which indicates that a MLR model built with a training split is generalizable for the corresponding test split as well. By studying MSE as function of X a validation for the MLR-X approach is obtained. From this the MSE_{train} decreases with increasing complexity, while the MSE_{test} passes a shallow minimum for MLR-4 (Figure 2a, Table S1).

This behaviour reflects the variance-bias trade-off, indicating that the models MLR-6 and 7 are overfitted to the training data, whereas MLR-1 to 3 are underfitted. Accordingly, the $\mu(\sigma)$-descriptor combination of the MLR-4 model with $\sigma = < 0.02$, 0.0, 0.01 and 0.03 shows the best regression of Li_2S_8 solubility (Table S1).

Second, the final predictive models for all possible MLR-*X* models were built using the entire dataset **I** and the significance of each descriptor was assessed by its p-value (Table S2). Here, as part of a hypothesis test, the *null* hypothesis (*no relationship between descriptor and Li_2S_8 solubility*) is rejected if its related p-value shows a probability of less than 5%, accepting the *alternative* hypothesis (*some relationship between descriptor and Li_2S_8 solubility*). In the MLR-1, 2 and 4 models, all descriptors have p-values < 5% or even < 1%. For the MLR-3, and 5 to 7 models, however, p-values > 5% were obtained, indicating that the respective descriptor is irrelevant to the corresponding model. As all $\mu(\sigma)$-descriptors originate from the same σ-potential curve, the increasing impact of collinearity between the descriptors in MLR-*X* models of higher complexity is also anticipated. This also indicates that these MLR-models are potentially overfitted independently of the sample size. Surprisingly, rather than $\mu(-0.03)$, $\mu(-0.02)$ is the descriptor that all MLR-*X* models have in common-reflecting the electrolyte's ability to interact with additional Li^+. For MLR-2 and MLR-3, the descriptors μ(0.01) and μ(0.03) are introduced, respectively, describing the affinity towards electron-pair donors. Overall, the MLR-4 model [Figure 2a, c, Eq. (1)]:

$$\log_{10}(s_{predicted}) = -8.53 \times \mu(-0.02) - 47.72 \times \mu(0.0) \\ 119.40 \times \mu(0.01) - 1.04 \times \mu(0.03) + 13.16 \quad (1)$$

and an $R^2 = 0.99$, is identified as the optimal regression model (followed by MLR-2 ($R^2 = 0.93$) and MLR-1 ($R^2 = 0.86$)).

2.3. Classification of Li–S Electrolytes

As noted above, the Li–S electrolytes used are classified as either salt-in-solvent (A), solvent-in-salt (B) or SIL (C) electrolytes. Similar to the development of regression models, classification models of different complexity were scanned but performed poorly. As the pair-plots (Figure S2) of the collected $\mu(\sigma)$-descriptors illustrate, the clearest separation between the classes is evident for $\mu(\pm 0.03)$, showing only a partial overlap for A and B. Hence, only these $\mu(\sigma)$-descriptors were utilized for the further systematic development of a classification model testing several classifier methods, including K-nearest neighbours (KNN), support vector machine (SVM) with linear or radial basis function (RBF) kernel, Gaussian Naïve Bayes (NB), and random forest (RF). A *one-vs.-rest* (OvR) approach was used to account for the multiple classes, representing each class by a classifier and comparing it to the remaining two classifiers/classes by receiver-operating-characteristics (ROC)-curves including the corresponding area-under-curve (AUC) (Figure 2d).

As expected from the clearly separated C grouping in Figure S2, most classifiers identify these SIL electrolytes with high reliability. For electrolytes of class A and B, however, a few false classifications are observed for all classifiers. Further, to also account for the statistical imbalance between the classes, *i.e.* 20 entities for A, 12 for B and 8 for C, the overall performance of each classification method is evaluated by the weighted- and micro-averaged AUC-score (Table 1).

Accordingly, KNN is the best-performing classifier. Its decision boundaries clearly show false classification of three class B electrolytes as class A (Figure S3). A final classification model was built using datasets **I** + **II** and was subsequently applied to dataset **III** (Figure 2b).

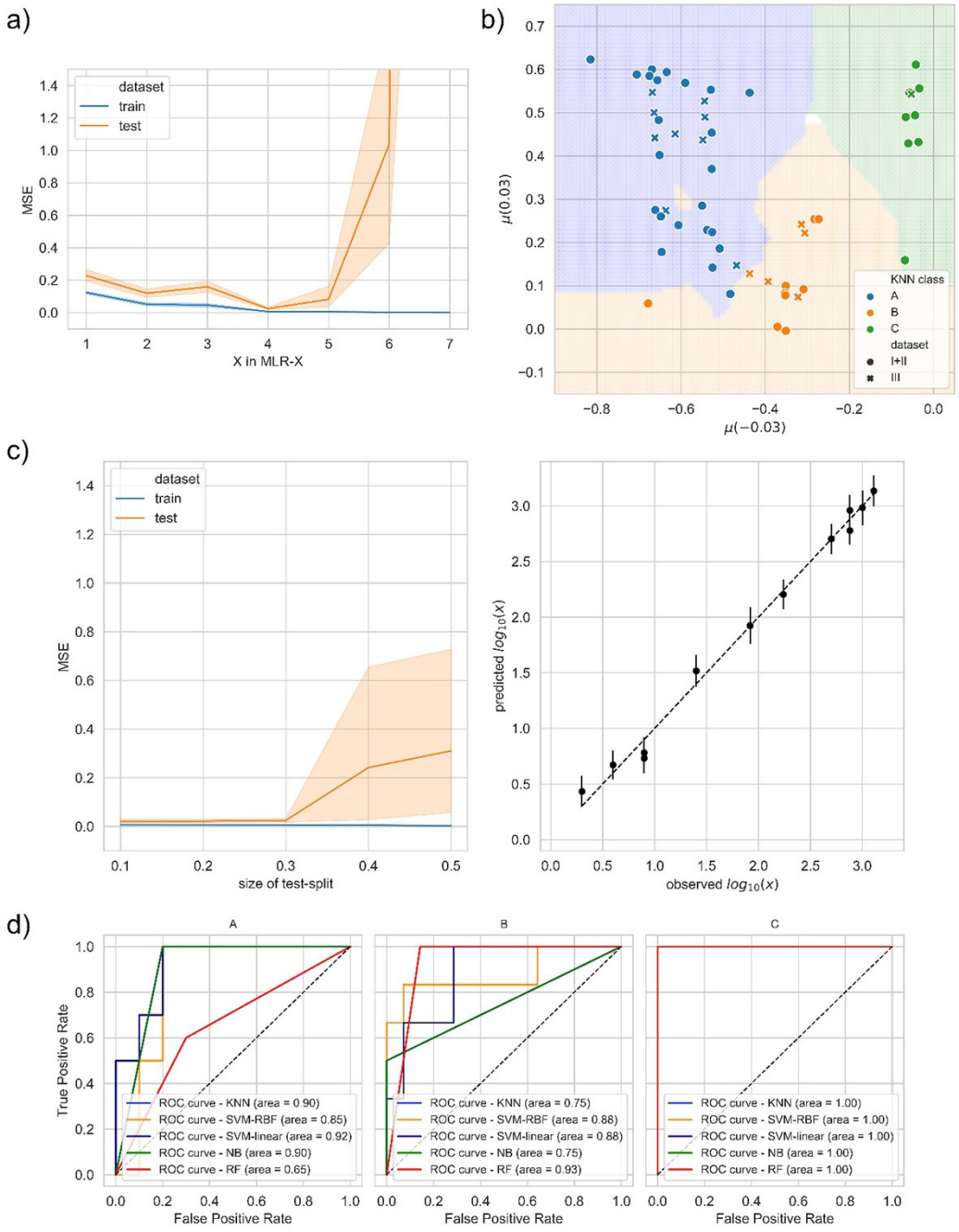

Figure 2. a) Validation of the MLR-X approach at 70/30 train/test data split, b) decision boundaries of the KNN classifier with blue, orange, green areas: classification as A, B or C respectively; white area: no or indistinct classification, c) learning curve for the MLR-4 $\mu(\sigma)$-descriptor combination (left) and observed *vs.* predicted Li_2S_8 solubility (right), and d) ROC-curves comparing the performance of the classifiers for each class where the vertical dotted line corresponds to an AUC of 0.5 and represents a randomly guessing classifier. KNN = K-nearest neighbours, SVM = support vector machine, NB = Gaussian Naïve Bayes, RF = random forest.

Table 1. Ranking of classifiers based on weighted- and micro-averaged ROC-AUC-scores with the macro-average provided for comparison.

Rank	Classifier	OvR-macro	OvR-weighted	OvR-micro
1	KNN	0.97	0.97	0.98
2	NB	0.96	0.95	0.98
3	SVM-linear	0.93	0.92	0.95
4	RF	0.94	0.92	0.95
5	SVM-RBF	0.91	0.89	0.94

2.4 Lessons Learnt

The created MLR-4 regression and KNN classification model was applied to the entire dataset of **I** + **II** + **III**-in total 55 electrolytes (Table 2).

Table 2. Electrolytes, their classification and Li_2S_8 solubility-experimental/literature and predicted (n = degree of coordination).

Electrolyte	Dataset	Class	KNN pred. class	Exp. log10[s]	MLR-4 log10[s]	Ref.
DME:DOL 1:1 (v:v)	I	A	A	2.88	2.96	[32]
1 M LiTDI DME:DOL	I	A	A	1.92	1.92	[33]
1 M LiTFSI DME:DOL	I	A	A	2.70	2.70	[33]
$[Li(G3_{n=3})_1]OTf$	I	B	A	3.11	3.13	[34]
$[Li(G3_{n=4})_1]TFSI$	I	C	C	0.60	0.67	[35,36]
$[Li(G4_{n=5})_1]TFSI$	I	C	C	0.90	0.73	[35,36]
$[Li(G4_{n=5})_1]BETI$	I	C	C	0.30	0.43	[34]
$[Li(THF)_2]TFSI \times 2$ THF	I	B	A	2.24	2.20	[35,36]
THF	I	A	A	3.00	2.98	[32]
1 M LiTFSI $G4_{n=3}$	I	A	A	2.88	2.78	[37]
$[Li(G1)_1]TFSI$ x G1	I	B	B	1.4	1.52	[35,36]
$[Li(G2)_{4/3}]TFSI$	I	C	C	0.9	0.78	[35,36]
DME	II	A	A		3.09	
DOL	II	A	A		2.74	
2 M LiTFSI DME:DOL	II	A	A		2.46	[31]
6 M LiTFSI DME:DOL	II	B	B		1.34	[31]
7 M LiTFSI DME:DOL	II	B	B		1.25	[31]
$[Li(G3_{n=3})_1]TFSI$	II	B	B		0.99	
$[Li(G3_{n=3})_1]BETI$	II	B	B		0.89	
$[Li(G3_{n=4})_1]BETI$	II	C	C		0.42	
$[Li(G4_{n=4})_1]TFSI$	II	C	C		0.68	
$[Li(G4_{n=4})_1]BETI$	II	C	C		0.39	
$[Li(G4_{n=1})_1]TFSI$	II	B	B		0.46	
$[Li(G4_{n=2})_1]TFSI$	II	B	B		0.90	
G4	II	A	A		3.16	
$[Li(G3_{n=1})_1]TFSI$	II	B	B		0.40	
$[Li(G3_{n=2})_1]TFSI$	II	B	B		0.85	
G3	II	A	A		2.92	
$[Li(THF)_1]TFSI$ x 3 THF	II	B	B		1.84	
$[Li(THF)_4]TFSI$	II	C	C		-0.04	[36]
1 M LiTFSI DOL:$G4_{n=1}$	II	A	A		2.37	
1 M LiTFSI DOL:$G4_{n=2}$	II	A	A		2.57	
1 M LiTFSI DOL:$G4_{n=3}$	II	A	A		2.64	
1 M LiTFSI DOL:$G4_{n=4}$	II	A	A		2.66	
1 M LiTFSI DOL:$G4_{n=5}$	II	A	A		2.60	
1 M LiTFSI $G4_{n=1}$	II	A	A		2.64	
1 M LiTFSI $G4_{n=2}$	II	A	A		2.69	
1 M LiTFSI $G4_{n=4}$	II	A	A		2.83	
1 M LiTFSI $G4_{n=5}$	II	A	A		2.79	
G2	II	A	A		2.60	
3 M LiTFSI DME:DOL	III		A		2.01	[31]
4 M LiTFSI DME:DOL	III		B		1.89	[31]
5 M LiTFSI DME:DOL	III		B		1.62	[31]
$[Li(G3_{n=3})_1]TDI$	III		A		2.12	
$[Li(G3_{n=4})_1]OTf$	III		A		3.53	
$[Li(G3_{n=4})_1]TDI$	III		A		2.48	
$[Li(G4_{n=4})_1]OTf$	III		A		3.42	
$[Li(G4_{n=4})_1]TDI$	III		A		2.52	
$[Li(G4_{n=5})_1]OTf$	III		A		3.63	
$[Li(G4_{n=5})_1]TDI$	III		A		2.62	
$[Li(G4_{n=3})_1]TFSI$	III		B		1.17	
$[Li(THF)_3]TFSI$ x THF	III		A		2.16	
1 M LiTFSI DOL:$TTE_{n=1}$	III		B		1.91	
1 M LiTFSI DOL:$TTE_{n=2}$	III		B		2.02	
$[Li(G1)_2]TFSI$	III		C		1.05	

From Table 2 we can extract some more detailed information for several sub-sets of electrolytes-both as function of salt type and concentration as well as the solvent applied.

Starting with the role of salt concentration, the LiTFSI in DME:DOL electrolytes have experimentally[31] been revealed to have a transition from the salt-in-solvent (A) to the solvent-in-salt (B) class upon increasing the LiTFSI concentration from 1 to 7 M and this is concomitant with a decreased polysulfide solubility. From the KNN-classification the transition is found to occur between 3–4 M-in agreement with the experimental findings, and additionally the MLR-4 regression model confirms the gradually decreased Li_2S_8 solubility-which is consistent with the common-ion effect.[28]

For G4 and DOL:G4 solvent matrices at a constant salt concentration of 1 M LiTFSI, however, the details in the molecular description of the Li^+-solvation *i.e.* using $[Li(G4)]^+$ complexes of different degrees of ether oxygen atom coordination are irrelevant as they all result in class A electrolytes. These also have similar Li_2S_8 solubilities, confirming that the "excess

glyme"[29,30] dominates the electrolyte properties rather than the nature of the $[Li(G4)]^+$ complexes.

An even clearer effect is observed when replacing DME by TTE, a fluorinated ether, as here exemplified by 1 M LiTFSI in DOL:TTE. The KNN-classification show this to be a class B electrolyte indicating a poor affinity towards additional Li^+. This is in, agreement with the poor solubility of Li^+ controlling the polysulfide solubility and ultimately improving the Li–S cell cycle-life.[28]

Moving to electrolytes originally designed to be SILs, *i.e.* class C, we can for example look at the $[Li(THF)_{4-n}]TFSI \times n$ THF electrolyte system. Experimental results indicate that the Li^+ first solvation shell undergoes frequent ligand exchange, a behaviour rather indicative of a class B electrolyte.[35,36] Here, the KNN-classification model reveals that for $n=0$ this system can indeed be considered a class C electrolyte, while for $n=1$ or 2 these electrolytes get falsely classified as A instead of B, the latter expected based on the (very) high salt concentrations applied.

Finally, we turn to an electrolyte design feature seldom targeted for Li–S batteries – the role of the Li–salt anion. Here four anions TFSI, BETI, OTf and TDI have been investigated. Out of these, the TFSI and BETI anions are both sulfonyl-imides and structurally very similar, and accordingly their Li_2S_8 solubilities and furthermore, they both enable class C electrolytes in equimolar mixtures with the glymes G3 or G4. In contrast, however, the anions OTf and TDI show a much higher affinity towards Li^+ as some negatively charged heteroatoms are poorly shielded (Figure S4). The tendency of LiTDI, despite TDI being known as a weakly coordinating anion in equimolar ratios with a glyme such as G3 or G4, to form ionic aggregates rather than forming an SIL was recently shown by molecular dynamics simulations.[38] The KNN-classification model results in class A rather than C electrolyte, even though that based on the high salt concentration a class B-type would have been expected.

3. Conclusions

We demonstrate that the σ-potential, $\mu(\sigma)$, derived within the COSMO-RS framework can be used to quantify the affinity of a binary or ternary electrolyte to interact with additional ions. The crucial step is found to be the molecular description of the electrolyte and in particular the first solvation shell of the Li^+ cation.

For Li–S battery electrolytes, this supervised ML approach can ultimately be used to predict the solubility of polysulfides and thereby classify the electrolytes more or less automatically. However, as the dataset used herein is rather limited, its current predictive power is rather qualitative than quantitative. The approach can be used to explain how different electrolyte design choices such as salt type and concentration as well as the solvent(s) applied affects $\mu(\sigma)$, which apart from the here targeted polysulfide solubility, has the promise to be a generic descriptor for future use in combination with other types of electrolyte and other physicochemical properties such as ionic conductivity, viscosity, *etc.*

Experimental Section

Computational Details

Structures were built in the graphical user interface (GUI) of TmoleX 4.1 and the quantum chemical calculations were performed using the TURBOMOLE[39,40] V7.0 program package. Geometries were optimized using the BP86-functional[41,42] and TZVP basis set[43] in gas phase and for the perfect conductor (COSMO: $\varepsilon=\infty$). Additionally, a single point calculation at BP86/def2-TZVP//BP86/def2-TZVPD level of theory was performed for gas phase and COSMO phase geometries to generate a fine grid cavity surface (FINE) for the molecules, which subsequently were saved in cosmo-files. The COSMO-RS calculations were performed using COSMOtherm and the BP_TZVPD_FINE_C30_1701 parametrization via the COSMOthermX GUI[44] to obtain the σ-potentials (μ(σ)). The (mole fraction) composition of each electrolyte was determined based on reported solvent-salt ratios.[26]

Datasets

The full dataset of 55 electrolytes contained binary and ternary mixtures of solvents and different dissolved Li–Salts and was divided into three sub-sets: (**I**) 12 electrolytes with experimental references for Li_2S_8 solubility and electrolyte classification, (**II**) 28 electrolytes of anticipated class but unknown Li_2S_8 solubility, and (**III**) 15 electrolytes of unknown class and Li_2S_8 solubility. A classification model was built using the combined sub-sets **I**+**II**, while the regression model (see below) for prediction of Li_2S_8 solubility was built using only sub-set **I**. Pure solvents and solvent mixtures, i.e. no Li–Salt contained, were considered as a limiting case within the class of salt-in-solvent electrolytes.

Regression Model Development and Validation

Models were built using supervised machine learning techniques as implemented in the open-source scikit-learn library (v0.22.1) for the Python programming language.[45] For the prediction, a multiple linear regression (MLR) technique using an ordinary least squares (OLS) method was applied to develop the QSAR-model correlating the COSMO-RS computed $\mu(\sigma)$ with the $\log_{10}$ of the experimental Li_2S_8-solubility s (in mM). In total, 127 MLR regression models were built and compared, each using a different complexity X and combination of $\mu(\sigma)$-descriptors (see section Results and discussion, *II. Regression*). MLR models were validated using a cross-validation algorithm over 20 random dataset splits into a training and test set at a 70/30 ratio. The fitted models were statistically evaluated via their respective mean-square error (MSE) according to Eq. (2)

$$\mathrm{MSE}=\frac{1}{N}\sum_{i=1}^{N}\left(\log_{10}\left(s^{\mathrm{calc}}\right)-\log_{10}\left(s^{\mathrm{exp}}\right)\right)^2 \quad (2)$$

and the correlation coefficient (R^2). The p-value analysis was conducted with the Python module *statsmodels* (v0.11.0).[46]

Classification Model Development and Validation

The dataset I+II contains 40 electrolytes: 20 salt-in-solvent, 12 solvent-in-salt, and 8 SILs. The dataset was randomly split at a ratio 1:1 in a training-test set and a validation set. The former was split again at a 1:1 ratio 25-times using a stratified shuffle algorithm to maintain the imbalanced class population in a training and test set throughout the cross-validation of one-vs-rest (OvR) hyper-parameter grid-search. The optimised classifiers *k*-nearest neighbours (KNN), support vector machine (SVM), and random forest (RF) have all been compared based on the area-under-curve (AUC) of their respective receiver-operator-characteristics (ROC) curves. The best performing classifier was selected to build a predictive classification model using the entire dataset **I+II**, subsequently employed on dataset III to create the final prediction.

Acknowledgements

This research received funding through the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 666221, "High-Energy Lithium Sulfur Cells and Batteries" (HELIS) and from VINNOVA (Sweden's Innovation Agency) through an Applied AI grant.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: electrolyte design · lithium-sulfur batteries · solubility · polysulfide · supervised machine learning

[1] S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson Education, Essex, **2016**.
[2] "White Paper on Artificial Intelligence: a European approach to excellence and trust," can be found under https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, **2020**.
[3] A. A. Franco, A. Rucci, D. Brandell, C. Frayret, M. Gaberscek, P. Jankowski, P. Johansson, *Chem. Rev.* **2019**, *119*, 4569–4627.
[4] Y. Liu, B. Guo, X. Zou, Y. Li, S. Shi, *Energy Storage Mater.* **2020**, *31*, 434–450.
[5] B. Liu, J. Yang, H. Yang, C. Ye, Y. Mao, J. Wang, S. Shi, J. Yang, W. Zhang, *J. Mater. Chem. A* **2019**, *7*, 19961–19969.
[6] K. Min, B. Choi, K. Park, E. Cho, *Sci. Rep.* **2018**, *8*, 1–7.
[7] J. J. de Pablo, N. E. Jackson, M. A. Webb, L.-Q. Chen, J. E. Moore, D. Morgan, R. Jacobs, T. Pollock, D. G. Schlom, E. S. Toberer, J. Analytis, I. Dabo, D. M. DeLongchamp, G. A. Fiete, G. M. Grason, G. Hautier, Y. Mo, K. Rajan, E. J. Reed, E. Rodriguez, V. Stevanovic, J. Suntivich, K. Thornton, J.-C. Zhao, *npj Comput. Mater.* **2019**, *5*, 41.
[8] "BIG-MAP: Battery Interface Genome - Materials Acceleration Platform," can be found under https://www.big-map.eu/, **n.d**.
[9] "Battery 2030+," can be found under https://battery2030.eu/, **n.d**.
[10] "Faraday Institution," can be found under faraday.ac.uk, **n.d**.
[11] "ARTISTIC," can be found under https://www.u-picardie.fr/erc-artistic/, **n.d**.
[12] R. P. Cunha, T. Lombardo, E. N. Primo, A. A. Franco, *Batteries & Supercaps* **2020**, *3*, 60–67; Supercaps **2020**, 3, 60–67.
[13] H. Johansson, *A Neural Network Approach to Absolute State-of-Health Estimation in Electric Vehicles Battery Degradation Study Based on Fleet Data*, Chalmers University of Technology, **2018**.
[14] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, R. D. Braatz, *Nat. Energy* **2019**, *4*, 383–391.
[15] B. Writer, *Lithium-Ion Batteries*, Springer International Publishing, Cham, **2019**.
[16] T. Fujita, D. A. Winkler, *J. Chem. Inf. Model.* **2016**, *56*, 269–274.
[17] Y. Liu, J. Wu, M. Avdeev, S. Shi, *Adv. Theory Simulations* **2020**, *3*, 1900215.
[18] Q. Pang, X. Liang, C. Y. Kwok, L. F. Nazar, *Nat. Energy* **2016**, *1*, 16132.
[19] J. Scheers, S. Fantini, P. Johansson, *J. Power Sources* **2014**, *255*, 204–218.
[20] E. Josef, Y. Yan, M. C. Stan, J. Wellmann, A. Vizintin, M. Winter, P. Johansson, R. Dominko, R. Guterman, *Isr. J. Chem.* **2019**, *59*, 832–842.
[21] Q. Pang, A. Shyamsunder, B. Narayanan, C. Y. Kwok, L. A. Curtiss, L. F. Nazar, *Nat. Energy* **2018**, *3*, 783–791.
[22] S. Drvarič Talian, G. Kapun, J. Moškon, A. Vizintin, A. Randon-Vitanova, R. Dominko, M. Gaberšček, *Chem. Mater.* **2019**, *31*, 9012–9023.
[23] A. Klamt, *J. Phys. Chem.* **1995**, *99*, 2224–2235.
[24] A. Klamt, F. Eckert, *Fluid Phase Equilib.* **2000**, *172*, 43–72.
[25] A. Klamt, *Rev. Comput. Mol. Sci*, Wiley Interdiscip. **2011**, 1, 699–709.
[26] S. Jeschke, P. Johansson, *Chem. A Eur. J.* **2017**, *23*, 9130–9136.
[27] S. Drvarič Talian, S. Jeschke, A. Vizintin, K. Pirnat, I. Arčon, G. Aquilanti, P. Johansson, R. Dominko, *Chem. Mater.* **2017**, *29*, 10037–10044.
[28] E. S. Shin, K. Kim, S. H. Oh, W. I. Cho, *Chem. Commun.* **2013**, *49*, 2004–2006.
[29] K. Ueno, K. Yoshida, M. Tsuchiya, N. Tachikawa, K. Dokko, M. Watanabe, *J. Phys. Chem. B* **2012**, *116*, 11323–11331.
[30] S. Tsuzuki, W. Shinoda, S. Seki, Y. Umebayashi, K. Yoshida, K. Dokko, M. Watanabe, *ChemPhysChem* **2013**, *14*, 1993–2001.
[31] L. Suo, Y. S. Hu, H. Li, M. Armand, L. Chen, *Nat. Commun.* **2013**, *4*, 1481.
[32] H. Pan, X. Wei, W. a. Henderson, Y. Shao, J. Chen, P. Bhattacharya, J. Xiao, J. Liu, *Adv. Energy Mater.* **2015**, *5*, 1500113.
[33] J. Chen, K. S. Han, W. A. Henderson, K. C. Lau, M. Vijayakumar, T. Dzwiniel, H. Pan, L. A. Curtiss, J. Xiao, K. T. Mueller, Y. Shao, J. Liu, *Adv. Energy Mater.* **2016**, *6*, 1600160.
[34] K. Ueno, J.-W. Park, A. Yamazaki, T. Mandai, N. Tachikawa, K. Dokko, M. Watanabe, *J. Phys. Chem. C* **2013**, *117*, 20509–20516.
[35] C. Zhang, K. Ueno, A. Yamazaki, K. Yoshida, H. Moon, T. Mandai, Y. Umebayashi, K. Dokko, M. Watanabe, *J. Phys. Chem. B* **2014**, *118*, 5144–5153.
[36] C. Zhang, A. Yamazaki, J. Murai, J. W. Park, T. Mandai, K. Ueno, K. Dokko, M. Watanabe, *J. Phys. Chem. C* **2014**, *118*, 17362–17373.
[37] J.-W. Park, K. Yamauchi, E. Takashima, N. Tachikawa, K. Ueno, K. Dokko, M. Watanabe, *J. Phys. Chem. C* **2013**, *117*, 4431–4440.
[38] P. Jankowski, M. Dranka, W. Wieczorek, P. Johansson, *J. Phys. Chem. Lett.* **2017**, *8*, 3678–3682.
[39] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165–169.
[40] TURBOMOLE V7.0 **2015**, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007, available from http://www.turbomole.com.
[41] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.
[42] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822–8824.
[43] A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* **1994**, *100*, 5829.
[44] COSMOthermX version C30_1701, 2016, COSMOlogic GmbH & Co. KG, Leverkusen, Germany.
[45] F. Pedregosa, V. Gaël, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *JMLR* **2011**, *12*, 2825–2830.
[46] S. Seabold, J. Perktold, *Proc. 9th Python Sci. Conf.* **2010**.

Manuscript received: February 1, 2021
Revised manuscript received: March 12, 2021
Accepted manuscript online: March 17, 2021
Version of record online: May 4, 2021

