Machine Learning Solvation Environments in Conductive Polymers: Application to ProDOT-2Hex with Solvent Swelling

Ioan B. Magdău* and Thomas F. Miller III*

Division of Chemistry & Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States

E-mail: imagdau@caltech.edu; tfm@caltech.edu

Abstract

Automated identification and classification of ion solvation sites in diverse chemical systems will improve the understanding and design of polymer electrolytes for battery applications. We introduce a machine learning approach to classify and characterize ion solvation environments based on feature vectors extracted from all-atom simulations. This approach is demonstrated in poly(3,4-propylenedioxythiophene), which is a promising candidate polymer binder for Li-ion batteries. In the dry polymer, four distinct Li^+ solvation environments are identified close to the backbone of the polymer. Upon swelling of the polymer with propylene carbonate solvent, the nature of Li^+ solvation changes dramatically, featuring a rapid diversification of solvation environments. This application of machine learning can be generalized to other polymer condensed-phase systems to elucidate the molecular mechanisms underlying ion solvation.

Introduction

Polymer binders such as polyvinylidene fluoride (PVDF) are conventionally used to confer chemical and mechanical stability to Li-ion batteries. Traditionally, these materials are neither good electronic nor ionic conductors, and their main role is to preserve the structural integrity of the electrodes.[1] New efforts to explore the use of conjugated polymers as binders[2–10] are underway, promising to substantially improve the efficiency of Li-ion batteries. These conjugated polymers, when doped, exhibit high electronic conductivities, however their ionic conductivities are substantially lower.[11] Poly(3-hexylthiophene) (P3HT), for instance, has good electronic conductivity owing to its π-stackings, and new studies show that functionalization of its side chains could improve its ionic conductivity as well.[12,13] Another candidate is poly(3,4-propylenedioxythiophene) (ProDOT-2Hex), which is more stable than P3HT and a better ionic conductor.[14,15]

A molecular-level explanation for the good ionic conductivity of ProDOT-2Hex remains elusive. Clarification of the structure-function relationship through computational simulation will help guide the synthesis of more efficient ProDOT derivatives as well as the exploration of other polymer chemistries and architectures. However, the simulation of ion transport in polymers using conventional all-atom molecular dynamics (MD) is hindered by the large system sizes needed and the slow timescales of ion diffusion,[16] requiring the development of more efficient strategies.

Extensive computational work towards understanding Li^+ transport has been performed on solid polymer electrolytes, of which polyethylene oxide (PEO) is the canonical example.[17–23]

Previous MD studies have shown that Li^+ diffusion in PEO takes place on multiple time scales,[24–35] featuring multiple modes of motion: intra-chain mobility, co-translation with the polymer and inter-chain hoping. Long-timescale MD trajectories enable the assessment of the relative importance of these motions.[24,25,36] Alternative approaches to investigate the ion transport mechanisms include the dynamical bond percolation (DBP) model,[37–40] the chemically specific dynamical bond percolation (CSDBP),[36,41] and trajectory-extending kinetic Monte Carlo (TEKMC),[42–44] which employ various strategies to access long timescales through stochastic hopping among solvation sites. Key to the applicability of these methods is a good understanding of the Li^+ ion solvation environment (SE) in the polymers of interest.[36] While in PEO and PEO-derivatives, the molecular mechanisms underpinning solvation are well understood, in most other systems these mechanisms are non-obvious and generally unknown.

The current work introduces an efficient machine learning (ML) strategy to achieve SE classification (SEC) in ProDOT-2Hex. We focus on the use of ML to automate the analysis of complex ion SEs, revealing in great detail the underlying molecular mechanisms. This approach can be extended to other polymer architectures and ionic charge carriers.

There are many strategies to represent chemical environments for the regression of the potential energy surfaces,[45–51] or for predicting solvation energies by supervised learning.[52–55] Only recently, it has been recognized that the unsupervised classification of chemical environments, in itself, can be useful to map out and visualize complex molecular and crystal structure spaces[56] or analyze solvation shell structures in simple liquids.[57] The current work takes the approach of mapping chemical environments based on simple molecular descriptors to address the problem of classifying SEs in condensed phase systems like polymers, where the richness of local structure highlights the true benefits of the ML analysis. This is demonstrated in ProDOT-2Hex, both dry and swollen (realistic battery environment), revealing in detail the molecular mechanisms of Li^+ solvation.

Methods

The SEC approach aims to identify and characterize ionic SEs in polymers with diverse chemistries by unsupervised learning from all atom molecular dynamics (MD) simulations. The method uses input from equilibrium MD trajectories of the ion in the solvent environment, where each trajectory is split into multiple windows. In each window, the molecular environment of the ion is characterized by a radial distribution function (RDF) specific to each atom type. These RDFs are then concatenated into feature vectors which serve as input to ML classifiers that cluster and label the SEs explored by the ion. The interpretation of each SE cluster is achieved by inspecting the cumulative distribution function (CDF), the distribution of Li^+ binding energy (which is not employed in learning) and the visual molecular representation (VMR). This visual representation is constructed by aligning all trajectory windows corresponding to an SE with respect to the atoms coordinating the Li^+ ion. A fully open-source version of the SEC code is available at `https://github.com/imagdau/SEC`.

Molecular Dynamics

This section explains how the MD data is generated. The approach to SEC is first demonstrated for Li^+ solvation in neat ProDOT-2Hex in the crystalline phase (Figure 1A), and then extended to ProDOT-2Hex swollen with propylene carbonate (PC), as well as to pure PC liquid solvent. Seven independent sets of production runs are performed, exploring the influence of swelling and polymer morphology on the nature of Li^+ solvation (details in Table 1). Trajectories are obtained in the dilute-ion regime, with single Li^+ ions countered by a background charge.

All MD simulations are performed using the LAMMPS simulation software,[58] employing the OPLS-AA force field with CM1A charges.[59–61] The equations of motion are propagated with a

Table 1: Details for the independent simulation sets. Simulations 1, 2, 3 are performed with both crystalline (c) and amorphous (a) phases. Simulation 4 corresponds to the pure PC solvent (liquid) simulation. The solvent concentration is $C_{solv} = N_{solv}/(N_{solv} + N_{mon})$.

	1(c),5(a)	2(c),6(a)	3(c),7(a)	4(l)
N_{pol}	2×4	2×4	2×4	–
N_{mon}	80	80	80	–
N_{solv}	0	16	40	320
N_{atms}	4176	4384	4696	4160
C_{solv}	0%	17%	33%	100%
N_{dim}	271	315	315	227

time step of 1 fs. For each condition detailed in Table 1, the simulation box is first equilibrated in the absence of Li^+ for 5 ns in the NPT ensemble at room temperature and ambient pressure. Subsequently, for each simulation condition and starting from the equilibrated configuration of the neat or swollen polymer, initial positions for Li^+ are drawn at random (Figure 1A), ensuring a spherical exclusion volume of radius 2.3 Å. An independent, 1 ns long MD simulation is then carried out for every initial position of Li^+ ion; the procedure being repeated for a total of $N_{traj} = 100$ initial positions per simulation condition. Given that these trajectories are shorted than the timescale of Li^+ ion diffusion, it is not expected that the resulting distribution of SEs correspond to the Boltzmann distribution.

Construction of Feature Vectors

The feature vectors for the SEC approach are obtained by concatenating the type-specific Li^+ RDFs into vectors (Figure 1C). The atoms are automatically divided into seven different types (Figure 1B) based on their Lennard-Jones (LJ) σ and ε parameters and each RDF is truncated to a cutoff radius $R_{max,j}$ (Table 2) that depends on the LJ parameters.

It is found that a cutoff radius $R_{max,j}$ that is slightly larger than the distance of maximal LJ attraction between the atom pairs yields robust classification, specifically, $R_{max,j} = \alpha \times R_{eq,j}$,

Table 2: Lennard-Jones parameters, as well as RDF cutoffs and binning for each atomic type.

	σ_j (Å)	$\varepsilon_j \left(\frac{\text{Kcal}}{\text{mol}}\right)$	$R_{max,j}$ (Å)	$N_{bin,j}$
H_{cap}	2.42	0.030	3.38	34
H	2.50	0.030	3.52	35
C_4	3.50	0.066	5.19	52
C_3	3.55	0.070	5.27	53
O_2	2.90	0.140	4.29	43
O_1	2.96	0.210	4.39	44
S	3.60	0.355	5.36	54

where

$$R_{eq,j} = \sqrt[6]{\frac{2\varepsilon_{Li}\sigma_{Li}^{12} + 2\varepsilon_j\sigma_j^{12}}{\varepsilon_{Li}\sigma_{Li}^{6} + \varepsilon_j\sigma_j^{6}}}, \qquad (1)$$

$\varepsilon_{Li} = 0.40$ Kcal/mol, $\sigma_{Li} = 1.40$ Å and $\alpha \approx 1.33$. The same binning resolution is used for all atom types, and therefore the number of bins that each RDF contributes to the combined feature vector depends on $R_{max,j}$ as summarized in Table 2. The combined feature vector is of length $N_{dim} = \sum_j^{N_{type}} N_{bin,j}$, which differs for each simulation set (Table 1) based on the atom types involved (Table 2).

Calculation of each RDF is performed by subdividing each single-ion 1 ns MD trajectory into $N_{win} = 50$ windows of length 20 ps and by averaging the RDFs within each window. The 20 ps time length was chosen to be shorter than the residence time of the Li^+ ion within each solvation site, but long enough to provide convergence of the RDFs. The total number of RDF feature vectors calculated for each simulation condition is $N_{win} \times N_{traj} = 5000$.

ML Classification of SEs

Classification of the Li^+ SEs is performed in two stages, as illustrated in Figure 2. First, the nonlinear dimensionality reduction algorithm UMAP[62] is used to embed the high-dimensional RDF feature representation into a low-dimensional latent space. Second, the clustering algorithm HDBSCAN[63] is employed to classify the embedded data into specific SEs.

Each bin in the RDF vector can be considered as a separate dimension, and the list of

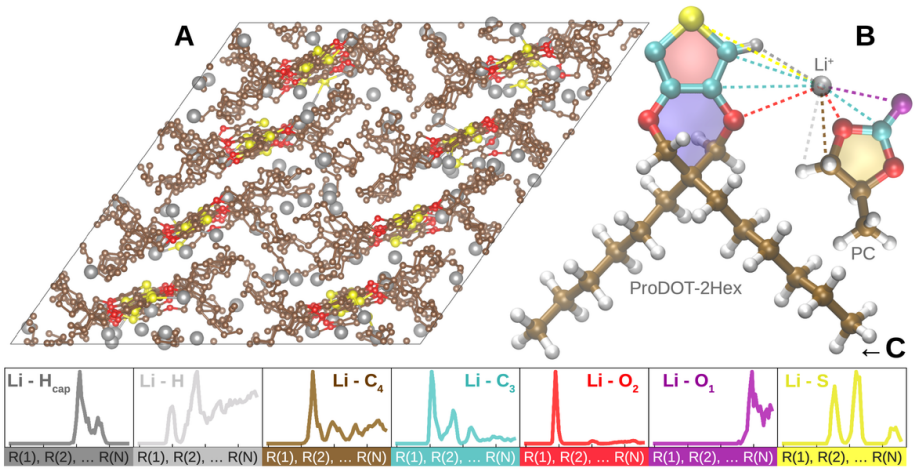

Figure 1: Construction of the feature vectors. Panel A: Li^+ starting positions within the equilibrated neat crystalline polymer. Each simulation was performed independently with a single ion. Here, we compiled all Li^+ initial coordinates into the same initial snapshot to demonstrate the uniform sampling of the simulation cell. Panel B: solvation environments are characterized by type-specific RDFs computed from the reference Li^+ to all other atom types. Panel C: RDFs are concatenated into a global RDF feature vector which is the input to SEC.

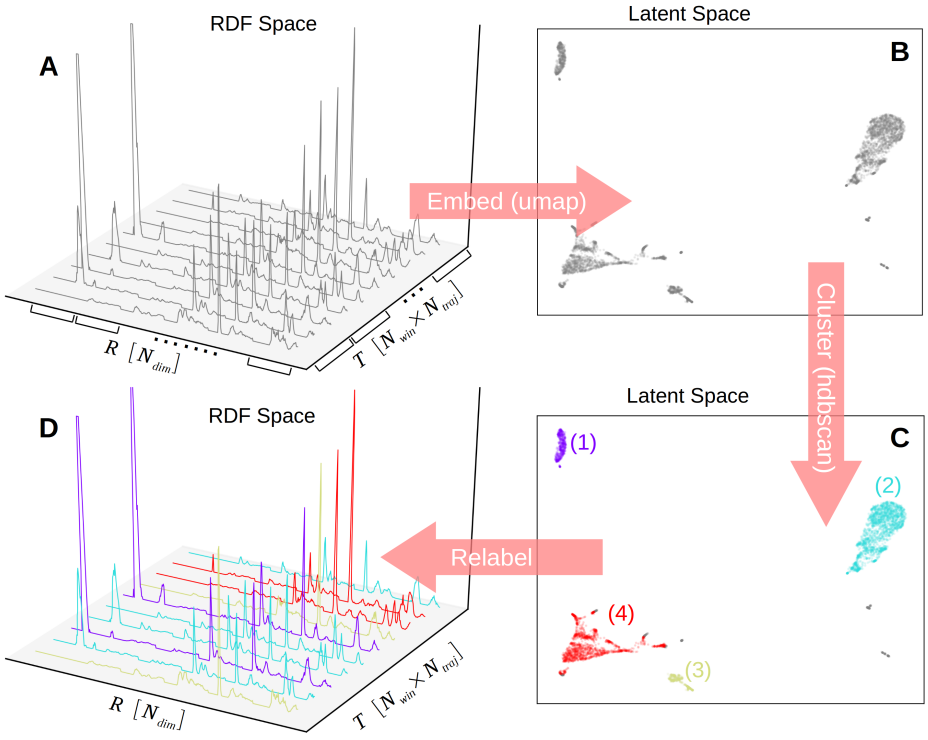

Figure 2: Schematic illustration of the SEC approach. Panel A shows the high-dimensional feature space comprising of $N_{win} \times N_{traj}$ RDF feature vectors of length N_{dim}. Panel B shows the unlabeled low-dimensional latent space obtained by dimensionality reduction. Panel C shows the clustering of the latent space into separate types of solvation environments. Panel D: back in high-dimensional space, each trajectory window is assigned to a SE.

all RDF values in the order of bin index correspond to coordinates in a high-dimensional N_{dim}-space. UMAP reduces the dimensionality from N_{dim} and embeds the data in a low, two-

dimensional latent space. Each RDF vector becomes a point in this new representation (Figure 2B). The proximity between these points indicates the similarity of the corresponding RDF feature vectors and, by extension, the solvation sites represented by these vectors. The natural clustering of sites means there is a finite number of general solvation environments, and our SEC approach can successfully distinguish between them. The role of the clustering algorithm is to label these clusters in latent space and, in effect, classify the solvation sites into specific SEs (Figure 2C).

Table 3: Clustering parameters for UMAP and HDBSCAN for each simulation set.

	N_{nbs}	D_{min}	N_{clst}	P_{cut}
1(c)	400	0.1	200	0.35
2(c)	400	0.1	100	0.5
3(c)	350	0.1	122	0.5
4(l)	400	0.01	50	0.5
5(a)	750	0.1	250	0.5
6(a)	400	0.1	100	0.35
7(a)	400	0.1	150	0.35

Both the UMAP and HDBSCAN algorithms can be tuned to improve the clustering and distinguishability of SEs (Table 3). The main parameters for UMAP are N_{nbs} and D_{min}. N_{nbs} determines the trade-off between local and global neighborhoods in the data structure, while D_{min} sets a lower bound on the distance between points in the embedded space. The main parameter of HDBSCAN is N_{clst} which sets the minimum size for a labeled cluster. Each data point is assigned to a cluster based on maximum probability of membership, while all points with probabilities less than P_{cut} remain unassigned (shown in gray in Figure 2C).

Interpretation of the SE Clusters

The classification of the SEs is obtained in an abstract latent space. While clustering and labeling are useful for understanding the statistics of SEs, equally important is the characterization of the different environments which is achieved by inspecting the CDFs, the Li^+ binding energy distribution and the VMR. CDFs are obtained by integrating the RDFs, and they carry information about the molecular composition of the solvation shells. The Li^+ binding energy distributions are obtained by selectively binning the recorded Li^+ energy over each specific SE. Finally, the VMRs are constructed by aligning multiple coordinate snapshots that belong to a given SE with respect to the atoms involved in the Li^+ solvation (a detailed description of the algorithm is given in the SI). These representations provide an useful illustration of the solvation shells and help with the molecular interpretation of the SEs.

Results and Discussions

In this section, we first discuss the results of the SEC analysis in neat ProDOT-2Hex and then in swollen ProDOT-2Hex. In the neat case, four independent, well characterized SEs (Figure 2C) are identified along the polymer backbone. The location and binding energy distributions of the solvation sites suggest a possible intra-chain mechanism for Li^+ conduction.
The SEC analysis is extended to swollen polymer at various concentrations of PC, where the SEs become increasingly more diverse. For each simulation condition, SEC predicts an independent set of SEs which can then be related to each other and relabeled consistently across simulations based on average CDFs, and VMRs. Tracking the distribution of SEs upon PC loading allows us to asses the effects of swelling on the efficiency of Li^+ ion solvation. The results presented here focus on the crystalline phase, however a similar analysis can be performed for the amorphous phase (the raw clustering for the amorphous phase is provided in the SI, while the relabeling could be achieved in a similar way as for the crystalline phase – based on CDFs and VMRs).

Neat Polymer Results

In neat ProDOT-2Hex, SEC predicts four well delimited clusters (Figure 2C), each of which corresponds to a distinct SE and which together comprise 93% of the sampled configurations.

SEs (2) and (4) are more prevalent with 44% and 32% of the solvation sites falling within these categories. SEs (1) and (3) are less common, comprising of 11% and 7% of all solvation sites.

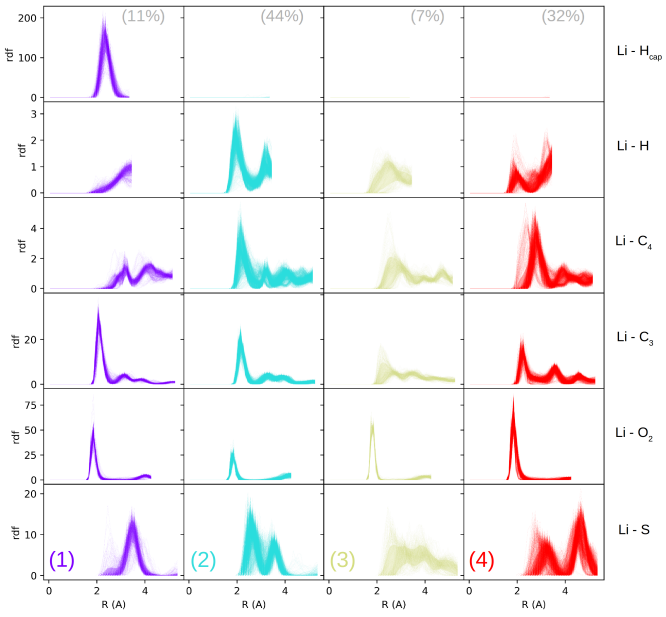

Figure 3: RDF clustering for neat ProDOT-2Hex. Each column (presented in one color) corresponds to a different SE as labeled by SEC, while each row corresponds to a different atom type, labeled on the right. The number on top is the frequency of finding each SE. Every colored dot in Figure 2C has an equivalent RDF plotted here. The faint shaded areas in the back represent the $\pm\sigma$ along the mean RDF.

Figure 3 shows the RDF feature vectors grouped by SE and arranged according to atom type. For each SE, there is an ensemble of RDF curves for every atom-type pair, where each RDF curve is produced by a different window of the MD sampling trajectory. The observed similarity among the RDFs that correspond to a give SE indicates that tight clustering in the latent space leads to the identification of structurally well-defined chemical environments. The Li-O_2 RDFs are particularly similar within each SE, confirming that this is an important feature of the classification. This conclusion is further supported by the average Li-O_2 CDFs shown in Figure 4A, which demonstrate that SEs (1), (2), (3) and (4) exhibit precise numbers of oxygen atoms (two, one, two and three, respectively). It is known that oxygen atoms play a crucial role in Li^+ solvation,[13,41] however the SEC approach identifies the oxygen atoms as the main feature of classification without this prior knowledge, which suggests that our approach may be used for cases where the important solvation interactions are not known *a priori*. Figure 3 also shows that SEs (1) and (3), both of which have two oxygen atoms in the immediate solvation shell, are distinguished from each-other based on the Li-H_{cap} pair; SE(1) comprises two H_{cap} atoms while SE(2) contains none. This observation indicates that other atom types apart from oxygen also contribute to classification.

The VMRs of each SE, shown in Figure 4B are complementary to RDFs/CDFs and help us understand the molecular interactions involved. SE(1) sites are located in the regions between polymer chains, where each chain contributes one dioxepane oxygen and a number of carbon atoms to the solvation shell. SE(2) is manifested near the polymer backbone, between two monomers which contribute one dioxepane oxygen and one thiophene sulfur, respectively. The solvation cage is completed by an hexyl sidechain from a neighbouring polymer. SE(3) is composed of two oxygen atoms and a few carbon atoms from two separate polymer chains and is located between SE(2) and SE(4). Finally, SE(4) site are located directly above a dioxepane ring, where both oxygen atoms in the ring contribute to solvation and where a neighbor polymer contributes a third oxygen to complete the solvation shell. Based on these observations, it is apparent that all solvation sites align close to the polymer backbone, establishing a periodic and uniform 1D network of sites (Figure 4D). This arrangement of sites suggests a possible Li^+ conduction mechanism as a hopping motion along the polymer backbone.

Figure 4C shows the compilation of all Li^+ ion trajectories carried out in the neat polymer, collected in a single density map and plotted against the initial polymer configuration. Each region of the isodentisy represents a solvation site, and it is colored according to the SE membership of the underlying trajectory. This partial distribution of solvation sites as sampled

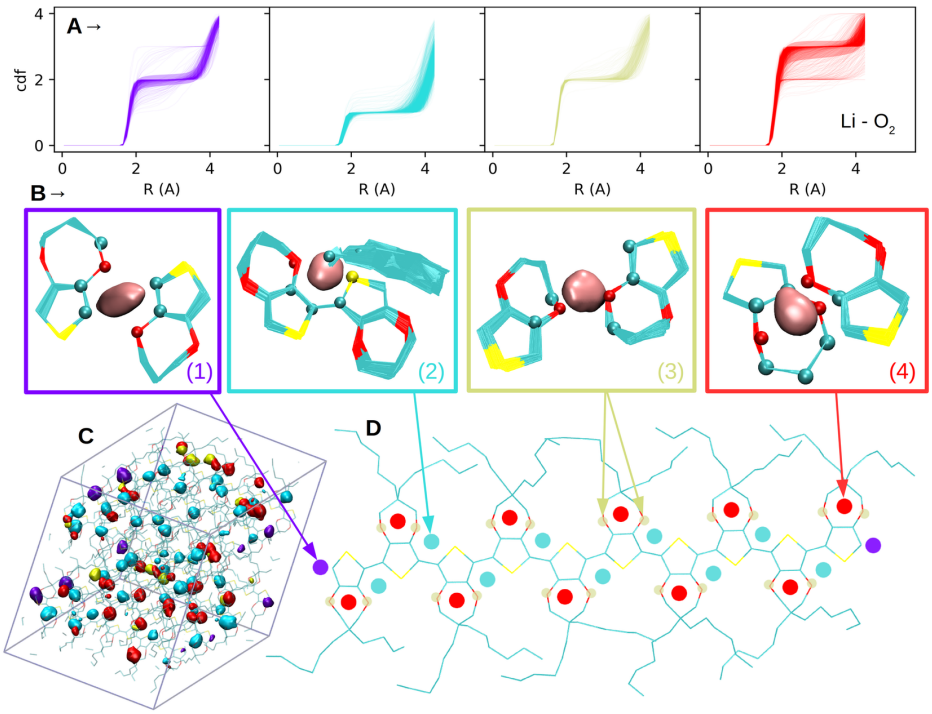

Figure 4: VMRs and spatial distribution of the four SEs found in neat crystalline ProDOT-2Hex. Panel A: Li-O_2 CDF for each SE. Panel B: the VMR of each SE obtained by aligning multiple trajectory windows. Panel C: partial distribution of Li^+ solvation sites as sampled by MD simulations, each SE labeled by a different color. Panel D: schematic representation of the inferred spatial distribution of solvation sites along the polymer backbone.

by the MD demonstrates that: first, each SE is found at multiple sites, and second, each solvation site corresponds to a single SE. These observations indicate that the SEC approach learns general solvation interactions based on specific local molecular environments and this generality can be used to infer the missing distribution of solvation sites (Figure 4D).

The characterization of SEs in neat ProDOT-2Hex is further improved by analyzing the distribution of Li^+ ion binding energies cumulated across all trajectories. Figure 5A shows the separation of this binding energy distribution into individual contributions from each environment. The binding energy is not used as a feature in learning and therefore the prediction of uniform single-peaked distributions for each environment demonstrates that the SEs determined through SEC are energetically well-defined chemical environments. Additionally, the average binding energy of each SE is correlated to the number of oxygen atoms in its solvation shell: SE(4) with three oxygen atoms, is the most energetically favourable binding site, whereas SE(2) with just one oxygen is the least favourable.

Figure 5 also shows the time evolution of the Li^+ binding energy for a few example trajectories (the rest of which can be found in the SI). Each of the 50 windows in a trajectory is colored according to the SE membership. During learning, each window is considered separately and no information about time ordering is passed to the classifier. Therefore, the color continuity of the trajectories for segments of constant energy, and the change of color upon hopping events demonstrates that SEs are robustly characterized and well distinguished from one-another.

Swollen Polymer Results

In this section, the SEC analysis was extended to swollen polymer (which is the realistic chemical environment in a battery) where PC was used as a model solvent. Owing to its highly

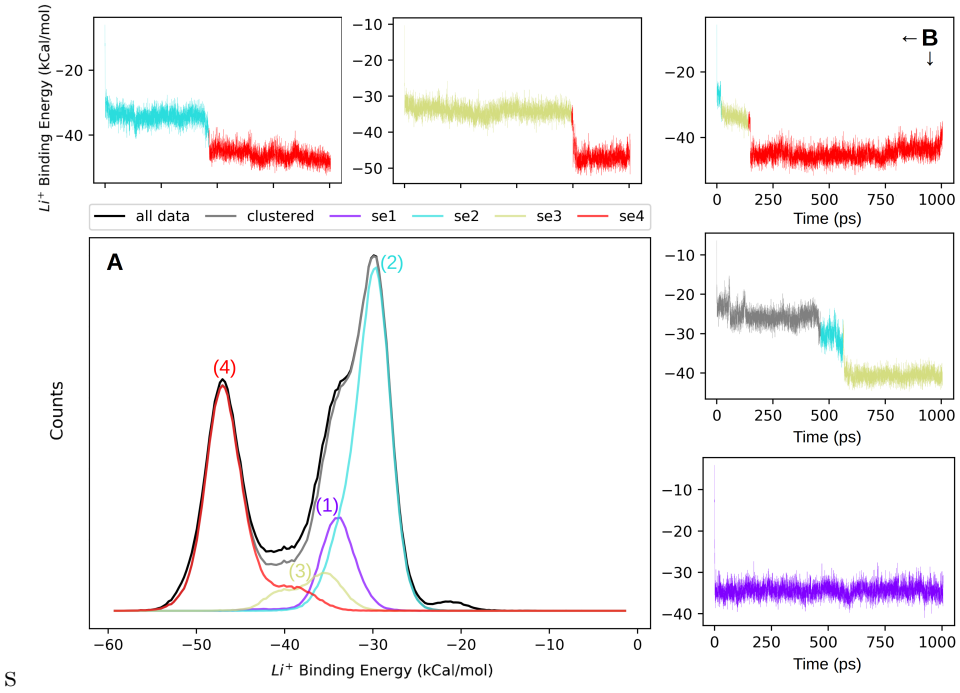

Figure 5: Li^+ binding energy in neat crystalline ProDOT-2Hex for each SE. Panel A shows the distribution of binding energies per SE (color), overall (black) and for all successfully clustered data (gray). Panels B shows the binding energy as function of time for a few example trajectories. Each trajectory window is assigned to an SE based on SEC. The rest of the trajectories can be found in the SI.

polar carbonyl group , PC can form a tight solvation shell around Li^+, strongly competing with the polymer for ion solvation. Figure 6 shows the SEC clustering obtained for swollen ProDOT-2Hex At different degrees of swelling (Table 1). Upon increasing PC concentration, the number of clusters in latent space increases substantially, indicating the polar solvent generates a rapid diversification of SEs. While only four SE clusters can be found in the neat phase, by 17% swelling concentration, there are six well defined clusters and by 33%, there are nine clusters corresponding to an equal number of SEs. A single SE type is found in the liquid phase of pure PC.

At first, SEC attributes a unique label to each cluster within a given simulation set (this raw clustering is provided in the SI). Based on average CDFs and VMRs the clusters can be compared to one-another across different swelling conditions and relabeled consistently. Clusters with primed and double-primed labels have molecular structures similar to the corresponding unprimed clusters, except comprising additional PC molecules (Figure 7). For instance, SE(2') is analogous to SE(2), except the hexyl side chain present in SE(2) is replaced by a PC molecule in SE(2'); similarly, in SE(2") the side chain is replaced by two PC molecules. SEs (1') and (1") are localized at the ends of the polymer backbones just like SE(1), but in addition their solvation shell contains one, and two PC molecules, respectively (Figure 7). SE(4) is the most stable environment in the neat polymer, and it does not hybridize in the swollen cases. Finally, clusters (5), (5') and (5") correspond to pure PC solvation, with four, five and six molecules respectively.

In pure PC liquid, there is a unique SE comprising six highly polar carbonyl oxygen atoms which belong to six independent molecules. This is clearly demonstrated by the presence of a single cluster in the latent space (SE(5")) and a unique peak in the binding energy distribution (Figure 7A). This environment is not supported by the swollen polymer, instead, other pure solvent variations with fewer molecules can be found (SEs (5') and (5)) in this case.

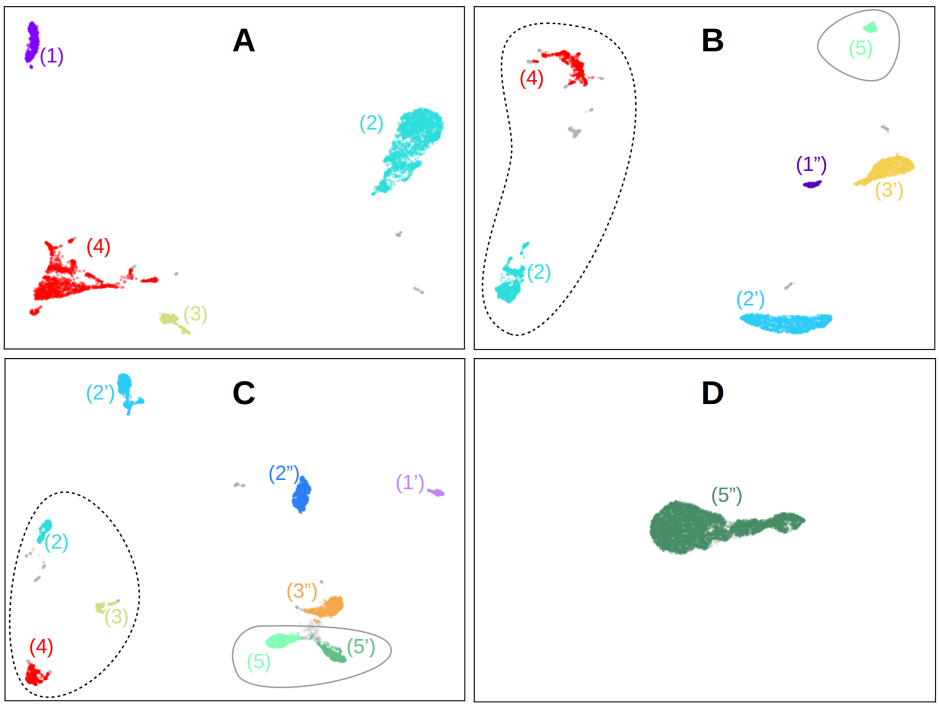

Figure 6: Latent space classification of SEs for the neat and swollen crystalline polymer. Panel A: neat ProDOT-2Hex, Panel B: swollen ProDOT-2Hex with 17% PC, Panel C: swollen ProDOT-2Hex with 33% PC, and Panel D: pure PC liquid solvent. Additional simulation details can be found in Table 1. Colors and labels are consistent throughout, such that SEs with the same color and label, have similar VMRs and characteristic CDFs. SEs in swollen polymers, with similar but not identical colors (prime labels), are closely related to each other. In Panels B and C, encircled with dashes line are the pure ProDOT-2Hex SEs and with solid line, the pure PC SEs.

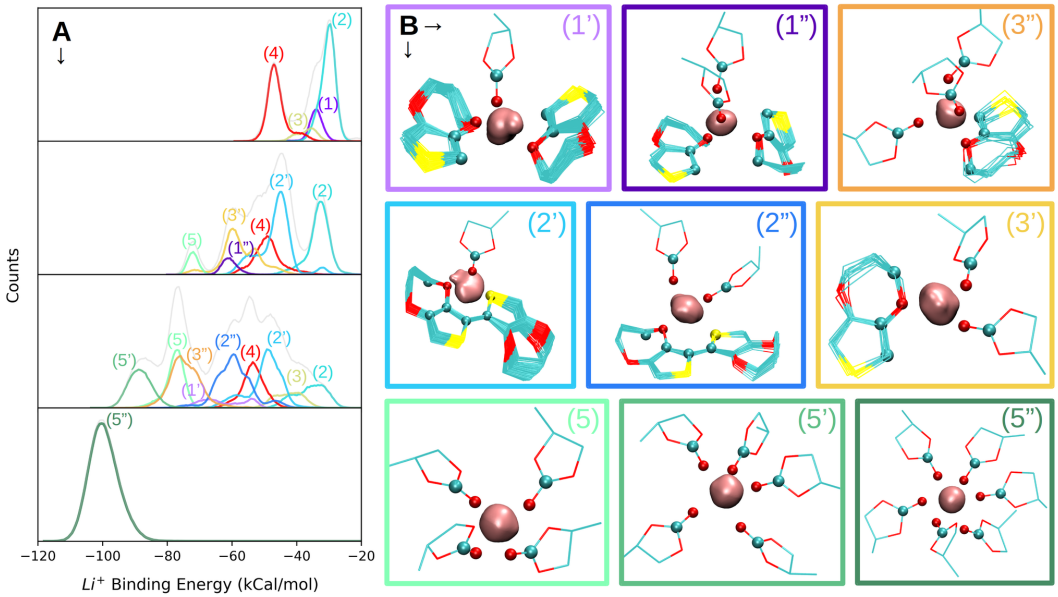

Figure 7: Li^+ binding energy and molecular representation of SEs in the swollen crystalline polymers. Panel A, top to bottom, shows the binding energy distribution for simulations 1-4 in Table 1 broken into SEs. Panel B shows the corresponding VMRs of the SEs. Labels and colors are consistent with Figure 6.

ProDOT-2Hex specific SEs, mainly (2) and (4), do survive in the swollen polymer (encircled with dashed line in Figure 6) but become gradually less frequent with increasing PC concentration. Surprisingly, their binding energy remains largely unchanged, independent of the degree of swelling. In fact, as shown in Figure 7A this is the case for all SEs, as the mean position of each energy peak remains centered around the same value upon swelling. The increasing complexity of the energy distribution with swelling concentration is due to the emergence of new energy peaks corresponding to hybrid environments.

Akin to the case of neat polymer, the number of oxygen atoms in the solvation shell is the main determinant of the average binding energy, as demonstrated for example, by the ordering of peaks (5), (5') and (5") in Figure 7A. The carbonyl oxygens, however, are more binding than dioxepane oxygens. For instance SEs (2") and (3') both contain two carbonyl oxygens and one dioxepane oxygen and have roughly the same binding energy, while SE(4) includes three dioxepane oxygens and is about 10 Kcal/mol less binding. This demonstrates that PC is significantly more effective at solvating Li^+ and explains why swelling dramatically lowers the binding energy distribution.

The work shown here gives a mechanistic understanding of solvation. Furthermore, SEC constitutes a stepping stone towards quantitative statistical methods that could access the long time scales required to study ion diffusion. On its own, SEC can reveal the solvation mechanism of various ions and counter-ions and explore diverse polymer chemistries and architectures, the method being entirely general.

Conclusions

Revealing the molecular interactions underpinning solvation can help understand the structure-function relationships in conductive polymers and guide the design of better materials for applications in photovoltaics and energy storage devices. Numerous studies have investigated the solvation mechanisms specific to canonical polymers such as PEO, but fewer tools are available for efficiently and automatically addressing this issue across the broad swath of other polymer systems.

Here we introduce the solvation environment classification (SEC) approach to identify and characterize solvation in general molecular environments, which is based on machine learning and all-atom MD simulations. The learning is based on RDF feature vectors computed from short MD trajectories. These high-dimensionality features are first embeded in a low-dimensionality latent space and then, based on similarity, they are grouped and classified into specific solvation environments (SEs).

SEC is demonstrated for the crystalline ProDOT-2Hex polymer, both dry and swollen phases, where PC was used as a model solvent. Four well characterized SEs were found in the dry phase, all localized near the backbone of the polymer. Their placement and binding energy distribution suggest a possible ion transport mechanism along the polymer chain. In the swollen polymer, the presence of polar solvent significantly increases the variety of SEs. These new SEs are characterized by lower binding energies owing to an increased number of oxygens coordinating the ion.

The SEC approach presented here can be similarly applied to other polymer materials – in both crystalline and amorphous phases – and can help explore the ion-solvation mechanisms in a wide variety of ionic charge carriers. To facilitate these applications, a fully open-source version of the SEC software is provided at `https://github.com/imagdau/SEC`.

Acknowledgement This work was supported as part of the Center for Synthetic Control Across Length-Scales for Advancing Rechargeables (SCALAR), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award No. DE-SC0019381. The authors thank Philip Shushkov, Stephen Munoz, Kim Jeongmin and the SCALAR collaboration for helpful discussions.

References

(1) Chen, H.; Ling, M.; Hencz, L.; Ling, H. Y.; Li, G.; Lin, Z.; Liu, G.; Zhang, S. Exploring chemical, mechanical, and electrical functionalities of binders for advanced energy-storage devices. *Chemical reviews* **2018**, *118*, 8936–8982.

(2) Wu, M.; Xiao, X.; Vukmirovic, N.; Xun, S.; Das, P. K.; Song, X.; Olalde-Velasco, P.; Wang, D.; Weber, A. Z.; Wang, L.-W.; Battaglia, V. S.; Yang, W.; Liu, G. Toward an ideal polymer binder design for high-capacity battery anodes. *Journal of the American Chemical Society* **2013**, *135*, 12048–12056.

(3) Zhao, H.; Wang, Z.; Lu, P.; Jiang, M.; Shi, F.; Song, X.; Zheng, Z.; Zhou, X.; Fu, Y.; Abdelbast, G.; Xiao, X.; Liu, Z.; Battaglia, V. S.; Zaghib, K.; Liu, G. Toward practical application of functional conductive polymer binder for a high-energy lithium-ion battery design. *Nano letters* **2014**, *14*, 6704–6710.

(4) Wu, H.; Yu, G.; Pan, L.; Liu, N.; McDowell, M. T.; Bao, Z.; Cui, Y. Stable Li-ion battery anodes by in-situ polymerization of conducting hydrogel to conformally coat silicon nanoparticles. *Nature communications* **2013**, *4*, 1–6.

(5) Das, P. R.; Komsiyska, L.; Osters, O.; Wittstock, G. PEDOT: PSS as a functional binder for cathodes in lithium ion batteries. *Journal of The Electrochemical Society* **2015**, *162*, A674.

(6) Higgins, T. M.; Park, S.-H.; King, P. J.; Zhang, C.; McEvoy, N.; Berner, N. C.; Daly, D.; Shmeliov, A.; Khan, U.; Duesberg, G.; Nicolosi, V.; Coleman, J. N. A commercial conducting polymer as both binder and conductive additive for silicon nanoparticle-based lithium-ion battery negative electrodes. *Acs Nano* **2016**, *10*, 3702–3713.

(7) Kaufman, J.; Chung, T.-C.; Heeger, A.; Wudl, F. Poly (thiophene): a stable polymer cathode material. *Journal of the Electrochemical Society* **1984**, *131*, 2092.

(8) Shi, Y.; Zhou, X.; Yu, G. Material and structural design of novel binder systems for high-energy, high-power lithium-ion batteries. *Accounts of chemical research* **2017**, *50*, 2642–2652.

(9) Salem, N.; Lavrisa, M.; Abu-Lebdeh, Y. Ionically-functionalized poly (thiophene) conductive polymers as binders for silicon and graphite anodes for li-ion batteries. *Energy Technology* **2016**, *4*, 331–340.

(10) Welsh, D. M.; Kloeppner, L. J.; Madrigal, L.; Pinto, M. R.; Thompson, B. C.; Schanze, K. S.; Abboud, K. A.; Powell, D.; Reynolds, J. R. Regiosymmetric dibutyl-substituted poly (3, 4-propylenedioxythiophene) s as highly electron-rich electroactive and luminescent polymers. *Macromolecules* **2002**, *35*, 6517–6525.

(11) Zeng, W.; Wang, L.; Peng, X.; Liu, T.; Jiang, Y.; Qin, F.; Hu, L.; Chu, P. K.; Huo, K.; Zhou, Y. Enhanced ion conductivity in conducting polymer binder for high-performance silicon anodes in advanced lithium-ion batteries. *Advanced Energy Materials* **2018**, *8*, 1702314.

(12) Dong, B. X.; Nowak, C.; Onorato, J. W.; Strzalka, J.; Escobedo, F. A.; Luscombe, C. K.; Nealey, P. F.; Patel, S. N. Influence of side-chain chemistry on structure and ionic conduction characteristics of polythiophene derivatives: a computational and experimental study. *Chemistry of Materials* **2019**, *31*, 1418–1429.

(13) Pipertzis, A.; Mühlinghaus, M.; Mezger, M.; Scherf, U.; Floudas, G. Polymerized ionic liquids with polythiophene backbones: self-assembly, thermal properties, and ion conduction. *Macromolecules* **2018**, *51*, 6440–6450.

(14) Kumar, A.; Welsh, D. M.; Morvant, M. C.; Piroux, F.; Abboud, K. A.; Reynolds, J. R. Conducting poly (3, 4-alkylenedioxythiophene) derivatives as fast electrochromics with high-contrast ratios. *Chemistry of Materials* **1998**, *10*, 896–902.

(15) Welsh, D. M.; Kumar, A.; Meijer, E.; Reynolds, J. R. Enhanced contrast ratios and rapid switching in electrochromics based on poly (3, 4-propylenedioxythiophene) derivatives. *Advanced Materials* **1999**, *11*, 1379–1382.

(16) Mogurampelly, S.; Borodin, O.; Ganesan, V. Computer simulations of ion transport in polymer electrolyte membranes. *Annual review of chemical and biomolecular engineering* **2016**, *7*, 349–371.

(17) Yue, L.; Ma, J.; Zhang, J.; Zhao, J.; Dong, S.; Liu, Z.; Cui, G.; Chen, L. All solid-state polymer electrolytes for high-performance lithium ion batteries. *Energy Storage Materials* **2016**, *5*, 139–164.

(18) Miller III, T. F.; Wang, Z.-G.; Coates, G. W.; Balsara, N. P. Designing polymer electrolytes for safe and high capacity rechargeable lithium batteries. *Accounts of chemical research* **2017**, *50*, 590–593.

(19) Florjańczyk, Z.; Krawiec, W.; Wieczorek, W.; Przyłuski, J. Polymer solid electrolytes based on ethylene oxide copolymers. *Die Angewandte Makromolekulare Chemie: Applied Macromolecular Chemistry and Physics* **1991**, *187*, 19–32.

(20) Bouridah, A.; Dalard, F.; Deroo, D.; Cheradame, H.; Le Nest, J. Poly (dimethylsiloxane)-poly (ethylene oxide) based polyurethane networks used as electrolytes in lithium electrochemical solid state batteries. *Solid State Ionics* **1985**, *15*, 233–240.

(21) Quartarone, E.; Mustarelli, P.; Magistris, A. PEO-based composite polymer electrolytes. *Solid State Ionics* **1998**, *110*, 1–14.

(22) Barteau, K. P.; Wolffs, M.; Lynd, N. A.; Fredrickson, G. H.; Kramer, E. J.; Hawker, C. J. Allyl glycidyl ether-based polymer electrolytes for room temperature lithium batteries. *Macromolecules* **2013**, *46*, 8988–8994.

(23) Khurana, R.; Schaefer, J. L.; Archer, L. A.; Coates, G. W. Suppression of lithium dendrite growth using cross-linked polyethylene/poly (ethylene oxide) electrolytes: a new approach for practical lithium-metal polymer batteries. *Journal of the American Chemical Society* **2014**, *136*, 7395–7402.

(24) Borodin, O.; Smith, G. D. Mechanism of ion transport in amorphous poly (ethylene oxide)/LiTFSI from molecular dynamics simulations. *Macromolecules* **2006**, *39*, 1620–1629.

(25) Diddens, D.; Heuer, A.; Borodin, O. Understanding the lithium transport within a rouse-based model for a PEO/LiTFSI polymer electrolyte. *Macromolecules* **2010**, *43*, 2028–2036.

(26) Chattoraj, J.; Diddens, D.; Heuer, A. Effects of ionic liquids on cation dynamics in amorphous polyethylene oxide electrolytes. *The Journal of chemical physics* **2014**, *140*, 024906.

(27) Müller-Plathe, F.; van Gunsteren, W. F. Computer simulation of a polymer electrolyte: Lithium iodide in amorphous poly (ethylene oxide). *The Journal of chemical physics* **1995**, *103*, 4745–4756.

(28) Neyertz, S.; Brown, D. Local structure and mobility of ions in polymer electrolytes: a molecular dynamics simulation study of the amorphous PEO x NaI system. *The Journal of chemical physics* **1996**, *104*, 3797–3809.

(29) Siqueira, L. J.; Ribeiro, M. C. Molecular dynamics simulation of the polymer electrolyte poly (ethylene oxide)/ Li Cl O 4. I. Structural properties. *The Journal of chemical physics* **2005**, *122*, 194911.

(30) Siqueira, L. J.; Ribeiro, M. C. Molecular dynamics simulation of the polymer electrolyte poly (ethylene oxide)/Li Cl O 4. II. Dynamical properties. *The Journal of chemical physics* **2006**, *125*, 214903.

(31) Duan, Y.; Halley, J.; Curtiss, L.; Redfern, P. Mechanisms of lithium transport in amorphous polyethylene oxide. *The Journal of chemical physics* **2005**, *122*, 054702.

(32) Karo, J.; Brandell, D. A Molecular Dynamics study of the influence of side-chain length and spacing on lithium mobility in non-crystalline LiPF6· PEOx; x= 10 and 30. *Solid State Ionics* **2009**, *180*, 1272–1284.

(33) Brandell, D.; Priimägi, P.; Kasemägi, H.; Aabloo, A. Branched polyethylene/poly (ethylene oxide) as a host matrix for Li-ion battery electrolytes: A molecular dynamics study. *Electrochimica acta* **2011**, *57*, 228–236.

(34) Borodin, O.; Douglas, R.; Smith, G. D.; Trouw, F.; Petrucci, S. MD simulations and experimental study of structure, dynamics, and thermodynamics of poly (ethylene oxide) and its oligomers. *The Journal of Physical Chemistry B* **2003**, *107*, 6813–6823.

(35) Diddens, D.; Heuer, A. Lithium ion transport mechanism in ternary polymer electrolyte-ionic liquid mixtures: A molecular dynamics simulation study. *ACS Macro Letters* **2013**, *2*, 322–326.

(36) Webb, M. A.; Jung, Y.; Pesko, D. M.; Savoie, B. M.; Yamamoto, U.; Coates, G. W.; Balsara, N. P.; Wang, Z.-G.; Miller III, T. F. Systematic computational and experimental investigation of lithium-ion transport mechanisms in polyester-based polymer electrolytes. *ACS central science* **2015**, *1*, 198–205.

(37) Druger, S. D.; Nitzan, A.; Ratner, M. A. Dynamic bond percolation theory: A microscopic model for diffusion in dynamically disordered systems. I. Definition and one-dimensional case. *The Journal of chemical physics* **1983**, *79*, 3133–3142.

(38) Druger, S. D.; Ratner, M. A.; Nitzan, A. Generalized hopping model for frequency-dependent transport in a dynamically disordered medium, with applications to polymer solid electrolytes. *Physical Review B* **1985**, *31*, 3939.

(39) Druger, S. D.; Ratner, M. A.; Nitzan, A. Applications of dynamic bond percolation theory to the dielectric response of polymer electrolytes. *Solid State Ionics* **1986**, *18*, 106–111.

(40) Nitzan, A.; Ratner, M. A. Conduction in polymers: dynamic disorder transport. *The Journal of Physical Chemistry* **1994**, *98*, 1765–1775.

(41) Webb, M. A.; Savoie, B. M.; Wang, Z.-G.; Miller III, T. F. Chemically specific dynamic bond percolation model for ion transport in polymer electrolytes. *Macromolecules* **2015**, *48*, 7346–7358.

(42) Neyertz, S.; Brown, D. A trajectory-extending kinetic Monte Carlo (TEKMC) method for estimating penetrant diffusion coefficients in molecular dynamics simulations of glassy polymers. *Macromolecules* **2010**, *43*, 9210–9214.

(43) Hanson, B.; Pryamitsyn, V.; Ganesan, V. Mechanisms underlying ionic mobilities in nanocomposite polymer electrolytes. *ACS Macro Letters* **2013**, *2*, 1001–1005.

(44) Mogurampelly, S.; Ganesan, V. Effect of nanoparticles on ion transport in polymer electrolytes. *Macromolecules* **2015**, *48*, 2773–2786.

(45) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Physical Review B* **2013**, *87*, 184115.

(46) Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K.-R.; Gross, E. K. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Physical Review B* **2014**, *89*, 205118.

(47) Willatt, M. J.; Musil, F.; Ceriotti, M. Atom-density representations for machine learning. *The Journal of chemical physics* **2019**, *150*, 154110.

(48) Christensen, A. S.; Bratholm, L. A.; Faber, F. A.; Anatole von Lilienfeld, O. FCHL revisited: Faster and more accurate quantum machine learning. *The Journal of Chemical Physics* **2020**, *152*, 044107.

(49) Welborn, M.; Cheng, L.; Miller III, T. F. Transferability in machine learning for electronic structure via the molecular orbital basis. *Journal of chemical theory and computation* **2018**, *14*, 4772–4779.

(50) Cheng, L.; Kovachki, N. B.; Welborn, M.; Miller III, T. F. Regression clustering for improved accuracy and training costs with molecular-orbital-based machine learning. *Journal of Chemical Theory and Computation* **2019**, *15*, 6668–6677.

(51) Cheng, L.; Welborn, M.; Christensen, A. S.; Miller III, T. F. A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules. *The Journal of chemical physics* **2019**, *150*, 131103.

(52) Hutchinson, S. T.; Kobayashi, R. Solvent-specific featurization for predicting free energies of solvation through machine learning. *Journal of chemical information and modeling* **2019**, *59*, 1338–1346.

(53) Jaquis, B. J.; Li, A.; Monnier, N. D.; Sisk, R. G.; Acree, W. E.; Lang, A. S. Using machine learning to predict enthalpy of solvation. *Journal of Solution Chemistry* **2019**, *48*, 564–573.

(54) Moorthy, N. H. N.; Martins, S. A.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A. Classification study of solvation free energies of organic molecules using machine learning techniques. *RSC advances* **2014**, *4*, 61624–61630.

(55) Lim, H.; Jung, Y. Delfos: deep learning model for prediction of solvation free energies in generic organic solvents. *Chemical science* **2019**, *10*, 8306–8315.

(56) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics* **2016**, *18*, 13754–13769.

(57) Basdogan, Y.; Groenenboom, M. C.; Henderson, E.; De, S.; Rempe, S. B.; Keith, J. A. Machine learning-guided approach for studying solvation environments. *Journal of Chemical Theory and Computation* **2019**, *16*, 633–642.

(58) Plimpton, S. *Fast parallel algorithms for short-range molecular dynamics*; 1993.

(59) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences* **2005**, *102*, 6665–6670.

(60) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14* CM1A-LBCC: localized bond-charge corrected CM1A charges for condensed-phase simulations. *The Journal of Physical Chemistry B* **2017**, *121*, 3864–3870.

(61) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic acids research* **2017**, *45*, W331–W336.

(62) McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* **2018**,

(63) Campello, R. J.; Moulavi, D.; Sander, J. Density-based clustering based on hierarchical density estimates. Pacific-Asia conference on knowledge discovery and data mining. 2013; pp 160–172.